

What is Theme and how to translate it

Von der Philosophischen Fakultät der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Philosophie genehmigte Dissertation

vorgelegt von

Jonas Freiwald

Berichter: Univ.-Prof. Dr. phil. Stella Neumann
Prof. Dr. phil. Erich Steiner

Tag der mündlichen Prüfung: 23.07.2021

„Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.“

To my mother, Ellen Freiwald

Acknowledgements

This Ph.D. project on Theme and its effect on translations has been an exciting but also effortful endeavor. Fortunately, I did not have to go on this journey all by myself. I would like to thank the following people who either directly or indirectly contributed to the creation of this thesis:

- Merle and Markus, who assisted me with some of the alignments and identifiability annotations in the context of a student apprenticeship
- Catalina Vossenkaul and Eva Schiffbauer, who reviewed some of my Theme annotations in the context of an E-Lili project
- Jessica Di Napoli, who helped me get rid of all the *peculiarities* in the thesis
- Arndt Heilmann, who was a tremendous help with the statistics and who also provided moral support whenever frustration built up in the office
- Laura Herberigs, who assisted me with the translation glosses and the bibliography
- My father, Karl Bäcker, who has always supported me and allowed me to follow my dream to become a teacher and a researcher
- Stella Neumann, who gave me the opportunity to work alongside her as a student and a research assistant, who has been a mentor to me and whose supervision and suggestions have always been incredibly helpful
- Zoë Takvorian, who took it upon herself to proofread the entire thesis during her Referendariat, who has given me the moral and emotional support I needed and who has been my steady rock during these stressful times
- Zoë Miljanovic, my right-hand woman, who has gone above and beyond in helping me and without whom this project would have been exponentially more difficult
- Paula Niemietz, who has been a great inspiration from the first day I set foot in her Begleitkurs, who has been a tremendous colleague and friend and who could not have been more helpful with her amazing proofreading
- And lastly, my mother, Ellen Freiwald, to whom this dissertation is dedicated; my constant companion, my biggest supporter, and my best friend

Contents

Acknowledgements	iii
Contents	iv
Spelling conventions and examples	viii
List of tables and figures	ix
List of abbreviations	xi
1 Introduction	1
2 Translation studies	8
2.1 Translation equivalence, procedures, and shifts.....	8
2.2 Translation features and translationese.....	11
2.3 Corpus-based translation studies.....	13
2.4 Translation and register	15
3 English and German clause structure	20
3.1 English clause structure and comparison to German.....	20
3.2 The basic word order in German	23
3.3 Forefield	29
3.4 Midfield.....	36
3.5 Animacy, agency and sentience	40
4 Systemic Functional Linguistics	48
4.1 General concepts of SFL.....	48
4.2 Metafunctions	50
4.3 TRANSITIVITY	51
4.4 MOOD	58
4.5 THEME and INFORMATION.....	62
5 Theme	65
5.1 Theme and related concepts.....	65
5.1.1 Theme and Given and New	65
5.1.2 Theme and Topic.....	68
5.1.3 Theme and Subject	71
5.1.4 Theme and communicative dynamism.....	73

5.2	The meaning of Theme.....	75
5.2.1	Theme as the point of departure of the message.....	75
5.2.2	Theme and method of development	78
5.2.3	Theme as a wave	84
5.2.4	Other attempts at defining Theme.....	85
5.3	Rheme.....	87
5.4	Theme in English	88
5.5	Multiple Themes	93
5.6	Marked Themes.....	95
5.7	Theme in other languages.....	102
5.8	Theme in German	107
5.9	English-German comparison of Theme.....	120
5.10	Theme in translations.....	124
5.11	Hypotheses.....	128
6	Methodology.....	136
6.1	The corpus.....	136
6.2	Annotations tools	138
6.3	Annotation decisions	139
6.4	Statistical testing.....	152
7	Theme in German originals	157
7.1	Results.....	157
7.1.1	Multiple Themes and Theme Markedness.....	158
7.1.1.1	Forefield hypothesis.....	158
7.1.1.2	First experiential element hypothesis	161
7.1.1.3	Subject hypothesis.....	163
7.1.2	Participant Themes and process types	165
7.1.3	Circumstances	170
7.1.4	Subject animacy and sentience.....	179
7.1.5	Subject identifiability	182
7.2	Discussion.....	186
8	Theme in English originals.....	211
8.1	Results.....	211
8.1.1	Multiple Themes and Theme Markedness.....	212
8.1.1.1	First element hypothesis.....	212

8.1.1.2	First experiential element hypothesis	214
8.1.1.3	Subject hypothesis	216
8.1.2	Participant Themes and process types	217
8.1.3	Circumstances	221
8.1.4	Subject animacy and sentience.....	229
8.2	Discussion.....	233
9	Contrastive analysis of Theme in English and German.....	254
9.1	Results.....	254
9.1.1	Multiple Themes and Theme Markedness.....	255
9.1.1.1	Forefield hypothesis vs. first element hypothesis	255
9.1.1.2	Forefield hypothesis vs. first experiential element hypothesis	257
9.1.1.3	First experiential element hypothesis in both languages.....	259
9.1.1.4	Subject hypothesis in both languages.....	261
9.1.2	Participant Themes and process types	264
9.1.3	Circumstances	267
9.1.4	Subject animacy and sentience.....	270
9.1.5	Subject identifiability	272
9.2	Discussion.....	274
10	Analysis of Theme in English/German originals and translations.....	Fehler! Textmarke nicht definiert.
10.1	Results.....	291
10.1.1	Multiple Themes and Theme Markedness.....	292
10.1.1.1	Forefield hypothesis vs. first experiential element hypothesis	292
10.1.1.2	First experiential element hypothesis in both languages.....	294
10.1.1.3	Subject hypothesis in both languages.....	296
10.1.2	Participant Themes and process types	297
10.1.3	Circumstances	300
10.1.4	Subject animacy and sentience.....	303
10.1.5	Subject identifiability	305
10.2	Discussion.....	307
11	Regression analysis of Theme in translations between English and German.....	Fehler! Textmarke nicht definiert.
11.1	Results: English to German translations.....	327
11.2	Discussion: English to German translations.....	333

11.3 Results: German to English translations.....	357
11.4 Discussion: German to English translations.....	362
12 Summary of the results	387
13 Conclusion.....	405
References	412
Appendix	433
Chapter 7: Theme in German Originals.....	433
Chapter 8: Theme in English Originals.....	441
Chapter 9: Contrastive Analysis of Theme in English and German.....	448
Chapter 11: Inferential Analysis of Theme in English-German translations.....	457

Spelling conventions and examples

This thesis follows the spelling conventions used in Halliday and Matthiessen (2014). Names of systems, such as *THEME* and *MOOD*, are spelled with small capitals and names of structural functions, such as *Subject* and *Actor*, as well as names of variables, such as *Change* and *Register*, are spelled with an initial capital letter.

Most examples that are referenced in the text are taken from the CroCo Corpus (Hansen-Schirra, Neumann, and Steiner 2012). These examples are always followed by their reference to the corpus. The references include the name of the subcorpus, the name of the register and the text number. To reduce length, examples can include ellipses if the first independent clause is followed by other paratactic clauses or if groups that are part of the Rheme of the first independent clause include excessively long post-modifiers. German examples are always followed by a translation gloss in English. Translation gloss refers to word-for-word translations that follow the word order of the source language rather than the target language. Since Theme is primarily realized through word order, translation glosses are most appropriate here because many examples are meant to highlight word order differences between the two languages. If an example includes highlighted constituents, they will be specified individually in the running text. If an example includes tildes, they are always used to separate Theme elements from each other. To avoid redundancy, their function will not be specified again in the running text.

List of tables and figures

Tables

1 Linguistic signals of identifiability and non-identifiability in English and German	150
2 Length categories in EO and GO based on characters	152
3 Basic Theme distribution in GO (Forefield)	158
4 Basic Theme distribution in GO (1 st exp. element).....	161
5 Basic Theme distribution in GO (Subject).....	163
6 Participants and process types in GO.....	165
7 Circumstances and thematic potential in GO.....	171
8 Subject Theme animacy and sentience in GO	179
9 Subject Theme identifiability in GO	183
10 Basic Theme distribution in EO (1 st element).....	212
11 Basic Theme distribution in EO (1 st exp. element)	214
12 Basic Theme distribution in EO (Subject)	216
13 Participants and process types in EO	217
14 Circumstances and thematic potential in EO	222
15 Subject Theme animacy and sentience in EO.....	230
16 Basic Theme distribution in GO (Forefield) and EO (1 st element)	255
17 Basic Theme distribution in GO (Forefield) and EO (1 st exp. element).....	257
18 Basic Theme distribution in GO (1 st exp. element) and EO (1 st exp. element)	259
19 Basic Theme distribution in GO (Subject) and EO (Subject).....	261
20 Subject Theme identifiability in GO and EO.....	272
21 Basic Theme distribution in GO/GT (Forefield) and EO/ET (1 st exp. element).....	292
22 Basic Theme distribution in GO/GT (1 st exp. element) and EO/ET (1 st exp. element)	294
23 Basic Theme distribution in GO/GT (Subject) and EO/ET (Subject)	296
24 Circumstances and thematic potential in EO, ET, GO and GT.....	301
25 Subject Theme identifiability in EO, ET, GO and GT	305
26 Summary of regression analyses 1-6	330
27 Summary of regression analyses 7-12.....	359

Figures

1 Process type distribution in GO.....	166
2 Participant roles and Theme types in GO.....	168
3 Circumstance Theme distribution in GO.....	172
4 Thematic potential of circumstances in GO overall.....	174
5 Thematic potential of circumstances in GO overall and per register.....	176
6 Subject Themes and verb types in GO.....	180
7 Process type distribution in EO.....	218
8 Participant roles and Theme types in EO.....	219
9 Circumstance Theme distribution in EO.....	223
10 Thematic potential of circumstances in EO overall.....	225
11 Thematic potential of circumstances in EO overall and per register.....	227
12 Subject Themes and verb types in EO.....	231
13 Process type distribution in GO and EO.....	264
14 Circumstance Theme distribution in GO and EO.....	267
15 Thematic potential of circumstances in GO and EO.....	268
16 Subject Themes as part of non-sentient constructions in EO and GO.....	270
17 Process type distribution in EO, ET, GO and GT.....	297
18 Participant roles and Theme types in EO, ET, GO and GT.....	298
19 Circumstance Theme distribution in EO, ET, GO and GT.....	300
20 Subject Themes as part of non-sentient constructions in EO, ET, GO and GT.....	303

List of abbreviations

acc	Accusative
CD	Communicative dynamism
dat	Dative
EO	English originals
ET	English translations
FICTION	Fiction register
FSP	Functional sentence perspective
GO	German originals
GT	German translations
INSTR	Instruction manual register
px lex. verb	Prefix of the lexical verb
refl-1/2/3sg	First/second/third person singular reflexive pronoun
SFL	Systemic Functional Linguistics
SPEECH	Political speeches register
TP	Thematic progression
TPot	Thematic potential
TOU	Tourism leaflets register

1 Introduction

Theme. The starting point, the point of departure, the grounding, the setting, the jumping-off point, the peg on which the message is hung. Throughout the decades, Theme has had many different descriptions. Despite that, or possibly because of it, its precise meaning has remained somewhat of a mystery. Given its abstract, metaphorical nature, Theme is one of the most elusive concepts in the linguistic framework of Systemic Functional Linguistics; and the focal point of this thesis.

One of the central assumptions in Systemic Functional Linguistics is that language is a probabilistic system of choice. Whenever a speaker wants to express something, they are faced with a choice of how they utilize the language system to attain their communicative goals. Almost instantaneously, they advance through a multitude of linguistic subsystems, in which they choose between paradigmatic options until they ultimately arrive at the actual utterance. These subsystems are probabilistic in nature in that some choices are more probable than others, largely determined by the situational context. One such subsystem is *THEME*, which is part of the textual metafunction, and which divides the message into Theme and Rheme.

In 1967, Michael Halliday defined Theme as the point of departure of the message, which is still the definition that is used most often in contemporary systemic functional grammars (see Halliday and Matthiessen 2014: 89). Theme is the portion of a clause which opens up the message, which leads over to the rest of the message, the Rheme, and which serves as a point of reference on the basis of which the hearer can interpret the message accordingly. In English, thematic meaning is expressed through positioning, with thematic meaning preceding rhematic meaning. In other words, what comes early in the clause acts as the point of departure of the message, the Theme. The fact that most English clauses can be arranged in different ways shows that clause sequencing is a matter of choice and as such has meaning.

The assumption that the choice of clause structure is meaningful is in my eyes indisputable. The same clause can often be expressed in multiple different orders; and depending on the context, some orders are more likely and more appropriate than others. If this choice is meaningless, why do we prefer different structures of the same underlying clause in different situations?

That being said, it is a considerable leap from calling fronted elements meaningful to calling them the point of departure of the message. The functional descriptions of Theme are often criticized for being too vague and metaphorical (see for example Hudson 1986: 798; Fawcett 2007: 137) and the different kinds of elements that can be part of the point of departure, for instance conjunctive Adjuncts, Vocatives, and Subjects, function in such different ways that their union under one functional umbrella is difficult to imagine. Additionally, the concrete boundary between the Theme and the Rheme of an English clause has continuously been a matter of debate.

These issues only intensify when Theme is applied to other languages. Systemic Functional Linguistics is designed to describe more than just the language system of English and thus it is plausible that other languages also divide their message into a Theme and a Rheme. But even if Theme exists in other languages as well, its formal realization may differ substantially from that of English, to the point that it is questionable whether they still have the same function. Many systemic functional linguists have risen to this challenge and have proposed thematic structures for various languages. In the case of German, it was Steiner and Ramm (1995) and Steiner and Teich (2004) who were the first to describe German in systemic functional terms, including a detailed description of the German Theme system. They equate Theme in German with the topological field called Forefield, which is essentially the position before the finite verb. This formal description in German differs quite substantially from that of English since the German Theme is mostly restricted to a single element and does not necessarily include an element that carries experiential meaning.

Contrastive differences represent a challenge for many purposes of language use, but perhaps uniquely so for translations. Catford (1965: 20) defines translation as “the replacement of textual material in one language (SL) by equivalent textual material in another language (TL)”. However, the goal of equivalent textual materials is only an ideal, as any transfer from one language system to another is accompanied by change, be it formally, functionally, or both. As such, translation is a special linguistic activity because the choices that the translator makes in the target language are affected by previous choices that were likely made by a different author in the source language. This challenge of balancing language systems is only intensified if linguistic phenomena are contrastively different.

It is unlikely that most translators are aware of the linguistic concept of Theme in SFL. Nevertheless, if a language does linguistically mark a unit as the point of departure of the message, translators may be consciously or subconsciously aware of this linguistic resource and balance contrastive differences between languages. SFL-based translation studies is a relatively young field of research and even within this field, the analysis of Theme in translation has often been neglected (Kim and Matthiessen 2016: 336). Studies on English-German translations using systemic functional concepts have been on the rise in recent years (see for example Neumann 2003; Teich 2003; Neumann 2014; Freiwald 2016; Niemietz, Neumann, and Freiwald 2017) and important work has already been done in analyzing thematic differences in English-German translation. That being said, most of these studies do not solely focus on thematic aspects and only account for more general aspects of Theme. The state of the art is still lacking a detailed, empirical analysis of the many facets of Theme in English and German and its intricate effects on translation.

I first started working with Theme in 2014, as a student research assistant alongside Prof. Dr. Stella Neumann and Dr. Paula Niemietz. I was asked to analyze the thematic structure of English originals and German translations in the popular scientific register of the CroCo Corpus (Hansen-Schirra, Neumann, and Steiner 2012) to be presented at the European Systemic Functional Linguistics Conference in Paris. I had relatively little experience with systemic functional analyses, but I knew how to analyze Theme in English, and I assumed that Theme in German was analyzed the same way as in English, going up to the first experiential element in the clause.

Halliday and Matthiessen (2014: 111-112) argue that any Theme in English has to include an experiential element without which the thematic grounding of the message cannot be concluded. An experiential element is any unit of the clause whose meaning describes something we know from experience, like a person, an object, a process, or a period of time for example. The German Theme, however, is said to be restricted to the Forefield and the Forefield does not necessarily include an experiential element. If Theme form were in fact so fundamentally different between the two languages, German translators would frequently have to either abandon the experiential Theme of the source clause or eliminate any non-experiential Theme that took up the Forefield position. Conversely, translators translating into English would need to add experiential elements to the Theme which were not part of the German source clause. In any event, Theme would

have to be a frequent source of change in translations and thus warrants a detailed analysis.

Based on the shortcomings of the state on the art and my own research interests in the field of Theme and translation studies, I arrive at two main objectives that I pursue in this Ph.D. project: First, I want to gain a thorough understanding of Theme regarding its functional definition and its formal realization in both English and German. To do so, I explore different plausible Theme hypotheses on how far the thematic space extends and on the basis of this, work out all the contrastive differences of Theme between English and German. And second, I aim to investigate the effects of Theme on English-to-German and German-to-English translations. In this context, I am particularly interested in Theme aspects that are systematically changed in translations and the most common translation procedures that are associated with them. For these purposes, I will focus only on declaratives in both languages since thematic structure and mood are tightly linked and multiple separate accounts of Theme would go beyond the scope of this project.

To work out contrastive differences and the effects on translations, I annotated authentic German and English original texts and their matching translations in the translation corpus CroCo (Hansen-Schirra, Neumann, and Steiner 2012). As such, the project has an empirical, quantitative research design since any drawn conclusions will be corroborated by authentic translation data that was produced independently of this study. The results will be tested statistically to allow claims about the general population. The study is quantitative as opposed to qualitative due to the fact that it takes note of a large number of texts and data points instead of only analyzing exemplary texts.

While the thesis is in essence empirical, it is not designed to rely on the state of the art solely to inform the empirical analysis. When I first started working with Theme, I searched for a comprehensive and detailed discussion of Theme, touching on a multitude of thematic aspects and different formal realizations in different languages. The state of the art offers a number of great introductions to the central concepts of Systemic Functional Linguistics, one of them being Theme. But to my knowledge, there is no reference book that discusses Theme in a comprehensive way by contrasting different approaches, related concepts, and realizations in different languages. Beyond the detailed empirical results on Theme, I aim for this thesis to be the reference for Theme in the systemic functional research community.

While the thesis is quantitative, the results go beyond summary tables, global averages, and significance tests. The vast majority of annotations were carried out manually, which offers a more detail-oriented perspective than an automatic annotation of large amounts of data would. The thesis also includes close to 300 examples that were predominantly taken from the CroCo corpus to explain and substantiate the quantitative results. Thus, this thesis will also be relevant for more qualitative-oriented scholars.

This project is part of the research field of empirical translation studies informed by the linguistic theory of Systemic Functional Linguistics given its strong emphasis on translations and the focus on Theme as the linguistic phenomenon. However, the beauty of a corpuslinguistic analysis of translation is that through the annotations of the original subcorpora, intralingual and contrastive insights inevitably come to light. The results of this thesis can therefore also be of relevance to scholars who are only interested in contrastive linguistics or register studies. And while Theme is located in the systemic functional approach, it is a syntactic concept in practice given its formal realization of early position in English and German clauses. The results are therefore meaningful for any linguist who is interested in clause structure in English and/or German, irrespective of whether they are theory-driven or theory-neutral.

The structure of the thesis is as follows: Chapter 2, following this introduction, outlines the relevant concepts and terminologies in the field of translation studies. By no means does this chapter represent a thorough account of the different approaches and insights of translation studies. The chapter is only meant to set the foundation and introduce some of the terminology that are relevant for the purposes of this thesis.

Chapter 3 discusses the basic clause structures in English and German from a theory-neutral perspective. This is necessary because the formal descriptions of Theme in German, in particular, rely on a thorough understanding of the basic German word order, including the concept of the clause frame and the topological fields in German. Chapter 3 discusses all the characteristics of the German clause structure that are relevant for Theme and contrasts them with the word order in English.

In Chapter 4, the linguistic theory of Systemic Functional Linguistics is introduced and some of the central concepts are presented. The focus of this chapter is on the three metafunctions of a clause, which describe the different kinds of meanings that an utterance expresses: the experiential, the interpersonal, and the textual metafunction. As the main

system of the textual metafunction, *THEME* is introduced in this chapter as well but will only be discussed briefly in terms of its relation to other systems.

Chapter 5 comprises the central part of the theoretical discussion of this thesis as it revolves around all aspects regarding Theme. Theme is first distinguished from related concepts and subsequently described in functional terms. Following this, the formal realization of Theme in English is presented, including a description of multiple Themes and Theme markedness. This account of Theme in English is then contrasted with other languages and subsequently, Theme in German is discussed. The chapter concludes with a summary of previous findings in translation research on Theme and with the presentation of the hypotheses.

Chapter 6 outlines the methodology. It includes a presentation of the CroCo corpus, the four registers that were annotated and the annotation tools that were used. All relevant annotation decisions are also discussed, and the chapter ends with a description of statistical tests.

Chapter 7 and 8 are the first two results chapters. These two chapters present the intralingual analyses of the Theme structures in the German and English original subcorpora respectively. The focus here is on differences between the four registers and differences regarding various Theme hypotheses within each language.

Chapter 9 presents the contrastive Theme analysis. The results from the previous two chapters are contrasted to work out all the thematic differences between English and German. Given the importance of contrastive issues for translation studies, some of the insights gained in this chapter are also used to specify the translation hypotheses.

Chapter 10 is concerned with the thematic differences between original and translated clauses. This chapter illustrates these differences descriptively by presenting the Theme distributions of all four subcorpora, namely English and German originals as well as their matching translations. Both translation directions are presented side-by-side and discrepancies between the four groups are tested statistically.

Chapter 11 also deals with thematic differences in translations but focuses more on the effects that individual Theme aspects have on translation. It includes elaborate logistic regression models which feature a variety of Theme categories and which calculate the effects of each of these categories on change in translations. Unlike the previous chapter, Chapter 11 separates the two translation directions and discusses the effects of Theme on change and potential reasons surrounding these effects individually.

Chapter 12 summarizes the results and interpretations of the five results chapters. Chapter 13 is the conclusion chapter, in which the thesis in general and the methodology in particular is critically assessed, and potential future steps are outlined.

2 Translation studies

In this chapter, some of the basic concepts and terminologies in translations studies that are relevant for the purposes of the thesis are discussed. The chapter starts off with a discussion of the nature of translation generally and the significance of translation equivalence. Different translation strategies and procedures are presented followed by an outline of common translation features, also often referred to as translation universals. Given the orientation of this research project, special focus will be put on the area of corpus-based translation studies. This chapter concludes with a discussion of register and its relationship to translation.

2.1 Translation equivalence, procedures, and shifts

There have been numerous attempts to define translation, many of which share principal features but have different foci. Oettinger (1960: 110) describes the process of translation as follows: "Interlingual translation can be defined as the replacement of elements of one language, the domain of translation, by *equivalent* elements of another language, the range" (emphasis in original). Catford (1965: 20) largely shares Oettinger's understanding of translation, when he defines translation as "the replacement of textual material in one language (SL) by equivalent textual material in another language (TL)". Nida and Taber (1969: 12) add to these descriptions by defining the different linguistic domains that the translator must consider when transferring text from one language to another, such as semantic and stylistic considerations. And still, in all these descriptions of translation, the concepts of equivalence and adequacy are described as essential parts of translation. Koller (2011: 77-78) even considers equivalence to be the distinguishing factor between translation and other text processing activities.

Translated texts are inevitably different from texts produced in the source language as the linguistic materials originate from different sign systems. Thus, one of the central struggles of translating is adhering to the language rules of the target text while at the same time doing justice to the meaning of the original text. Different grades and subtypes of equivalence have been proposed (see for example Nida 1964; Catford 1965; Kade 1968). The degree to which equivalence is achievable or even desirable in a translation is largely determined by the language pair and the specific utterances. And yet in most

cases, translators have the choice between being more faithful to the source language by choosing formally equivalent (or near equivalent) formal structures or being more faithful to the target language by opting for natural, authentic language. This choice is of course not a question of either-or but rather more-or-less.

Bell (1991: 6) rightly points out that equivalence can be achieved at different levels of representation and at different linguistic ranks, which sometimes complement each other but can also be in contrast to each other. For example, a translation may be formally equivalent on the level of semantics but, as a consequence, less equivalent on the grammatical level. Correspondingly, a translation might have a high formal equivalence at the phrase rank but lack the same equivalence at the clause rank. According to Bell (1991: 6), some level of formal equivalence is always lost in the process of translation. Perfect equivalence between source and target text therefore remains an unattainable goal.

One of the more prevalent questions in translation studies is the relationship between source text and target text. Originally, a greater focus was placed on the source text and the requirement of a translation to do justice to the original. In recent decades, however, this focus has shifted more towards the target text (Baker 1996: 176). As a consequence, translation studies today is less concerned with comparing a translation to its original text and more engaged in comparing translated language with original texts in the target language to illuminate the similarities and differences between original language and translated language in general (Delaere 2015: 19). Of further interest are the effects of the source language on the translation product. It is generally accepted that features of the source language have an influence on the language use in the target text, known as shining through (Teich 2003: 145). The strength of this effect can also vary depending on differences in dominance or prestige of the languages involved (see Evert and Neumann 2017).

When translating a text, the translator has a seemingly infinite number of possibilities as to how a source text can be translated. And yet, translators often choose similar strategies to translate a source text. Newmark (1988) proposes different categories of translation strategies and procedures to categorize the different general approaches towards the translation of a text. For example, translators can choose to stay as close to the source text as possible and use a literal translation or even a word-for-word translation strategy. Alternatively, they can also depart from the source and only try to preserve themes, characters and plots while deviating from most structural conditions of the source text, which

Newmark (1988: 46) calls adaptation. Between these two ends of the spectrum, there are multiple translation strategies in between that vary in regards to their orientation towards the source and the target language. Some of the translation strategies are of little relevance for this thesis since they deal with the aesthetic values of highly artistic texts like poems or with the borrowing and adaptations of cultural references.

Linked to translation strategies are translation procedures (Catford 1965; Newmark 1988) which represent the approach of the translator to deal with a single translation problem. Translation problems can arise from multiple different sources but are often due to contrastive difference between source and target language. In case of such a translation problem, the translator is usually forced to abandon the most basic translation procedure, the literal translation (Newmark 1988: 70), and make changes to the source text. Such deviations are referred to as translation shifts and are a commonly discussed topic in translation studies (see for example Catford 1965; Taylor 1993; Munday 2001; Toury 2012).

There are different reasons why a translator may opt for a change in the translation. The most obvious reason is that the most appropriate equivalent of a source language unit corresponds to a different grammatical category in the target language. Also, in some cases the target language system requires a formal shift of the original because the language systems are contrastively different. Additionally, a translator may be forced to deviate from the source text on one linguistic level to preserve the source text structure on a different linguistic level. This is particularly relevant for English-German differences in thematic structure, as the order of clause elements and the order of semantic meaning can often not be preserved simultaneously (for example in non-sentient constructions; see Section 3.5).

However, translators can also deviate from the source text deliberately for several different reasons. Even if the language system of the target language does not impose a hard rule on the translator, frequency distribution and levels of naturalness and authenticity can still be different between source and target language. Therefore, the translator may deviate from the original text to make the translated text sound more natural in the target language. Also, the situational context can differ in the source and target language culture and with that the requirements for the translation can change. Register-related differences between the source and target culture as well as workflow-related factors can influence translations. Also, what may look like a shift in translation can also be the result

of editorial intervention (Kruger 2017; Bisiada 2018). Lastly, stylistic differences between author and translator can be responsible for shifts, which is especially relevant for the translation of literary texts.

Translation scholars like Catford (1965) and Vinay and Darbelnet (1995) have identified a variety of different translation procedures, which will not all be discussed in detail here. As will be shown in Chapter 5, Theme is primarily realized through word order in both English and German. Changes to the word order and changes to the grammatical structure of source clause elements are therefore the most likely types of shifts, which Catford (1965: 77) refers to as structure-shifts. However, Catford (1965) only considers formal correspondences and largely ignores functional shifts and while formal changes are likely prevalent, functional meaning differences in Theme translations are also inevitable. As an alternative, Vinay and Darbelnet (1995: 36-39) propose a different set of shift categories, including modulation, which can be applied to account for functional variation between source and target text. For the purposes of this project, a translation shift will be defined as any deviation from form or function of the Theme between the source and target clause.

2.2 Translation features and translationese

One of the most popular research areas in early corpus-based translation studies included so-called translation universals. However, the idea that translating was a unique form of text production and that translations had certain aspects that distinguished them systematically from original writing did not just arise with the rise of corpora. Well before corpus-based translation studies, scholars and translators noticed that translated texts were longer, simpler, more explicit, and more standardized (see for example Vinay and Darbelnet 1995; Blum-Kulka 1986). It was assumed that translations differed from original writing in the same language and that this difference does not represent a flaw or inadequacy of translated language, but that translation constitutes a third language (Duff 1981), third code (Frawley 1984) or hybrid language (Trosborg 1997). Instead of being reduced to the patterns or distinctive features of the source or the target language, translation can be seen as its own form of language, also sometimes referred to as translationese (for example Granger 1996: 49).

With the accessibility of translation corpora, these unique characteristics of translations could be tested. Baker (1993: 242) postulates that translations exhibit linguistic patterns, "patterns which are not the result of interference from the source or target language", referred to as translation universals. She defines translation universals as "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker 1993: 243). This idea of translation universals has also been met with a fair amount of criticism (see for example Bernardini and Zanettin 2004; Mauranen 2007; House 2008; Becher 2010; Kruger and van Rooy 2012). Universality implies that these features are present in any instance of translation, as a kind of prerequisite or axiom of translations. However, Baker's original definition already shows that she refers rather to features that are simply more common in translated language than in original language and that is a question of probability rather than of universality. Mauranen (2007: 35) also argues that a study on language universals should not restrict itself to absolute universals but also consider general tendencies shared by the majority of languages (or in this case translations). And while features of translations may not be unique or universal, translated texts can be distinguished from original texts with a high level of accuracy (Volansky, Ordan, and Wintner 2015; Evert and Neumann 2017).

Originally, Baker (1993: 243-245) postulated the following typical features of translation: explicitation, simplification, conventionality, avoidance of repetition, exaggeration of target language features, and a specific type of feature distributions in translations, which represents the status of translations as a third code. Explicitation refers to the tendency of translators to make a text more explicit in the sense that extra information is added to the text. Simplification means that translated language is generally simpler than original language, though Baker is not particularly explicit in what that means linguistically. According to Mauranen (2007: 39), simplification as a typical feature of translation has also not been generally accepted by the translation studies community. Conventionality or conservatism describes the tendency of translators to follow target language norms and to avoid marginal linguistic features. This feature is often conflated with the exaggeration of target language features and referred to as normalization (Baker 1996: 176-177). Textual conventionality and target language exaggeration were demonstrated by Vanderauwera (1985) for translations and Shlesinger (1991) for interpreting.

Baker (1996) adds an additional typical feature of translation called *levelling out*. Levelling out describes the tendency of translations to gravitate towards the center of any continuum. A comparison of individual translations is therefore less likely to exhibit variation than a comparison of original texts, regardless of source and target language (Baker 1996: 177). For instance, translated texts can be assumed to be more similar than original texts in terms of average sentence length, type-token ratio, and lexical density (Hansen and Teich 1999: 313).¹

Baker (1993) defines translation universals as not being the result of source language interference, which is why none of the features described so far are tied to the influence of the source language. However, many have suggested that this interference from the source language is in fact a unique characteristic of translated language (for example Eskola 2004; Laviosa-Braithwaite 1996; Mauranen 2004). Toury (2012) even assumes there to be the law of interference in translations. This tendency of translations to correspond to the source language in terms of pattern frequencies and system probabilities is also referred to as shining through (Teich 2003: 145). One common effect in translations, especially in the case of a translation problem caused by contrastive differences, is that the translation shows both signs of shining through of the source language and of normalization toward the target language. In terms of frequencies, the translation is often in between source and target language norms, which gives translationese the status of a hybrid language (Neumann 2012: 191).

2.3 Corpus-based translation studies

The introduction of language corpora in the 1980s has marked a methodological breakthrough in linguistic studies, which has allowed linguists to analyze language and to test linguistic hypotheses in a completely new way. By 1990, monolingual corpora for 16 European languages had already been designed (Leech 1991: 21) and since then the popularity of corpus-based approaches has only increased. Corpora can include a large number of texts, which are typically produced by different authors and possibly belong to different registers. Texts in a corpus represent authentic language use in the sense that they

¹ Additional typical features of translations have been proposed in translation studies, for example under-representation of unique target-language items, untypical collocations, and source-language interference (Mauranen 2007: 38). However, these will not be described in detail here.

were not produced specifically to be included in a corpus. In most cases, corpora are designed following a specific set of criteria which matches the research area for which the corpus is used. Common criteria are geographical or social commonalities of the authors or comparable contexts of use of the texts.

The advantages of corpus-based studies are obvious. Up until the breakthrough of corpus linguistics, most linguistic theorizing had to rely on introspection with little evidence to confirm assumptions about language. With electronic corpora, linguists can study actual, authentic language use from a variety of different registers. Corpora also allow the study of spoken language, which, for the longest time, had not been considered in linguistic studies. The large number of texts, sentences and words provide evidence for frequency distributions of linguistic features which are likely to represent the general population of texts. The digital nature of corpora allows (semi-)automatic annotations of large amounts of data, which are not only much faster but also more reliable and accurate than human annotations. Infrequent or subtle linguistic features and patterns can be detected by means of statistical methods. Lastly, electronic corpora can be shared easily, which allows linguists to not only replicate previous studies but to also base new studies on the same data (Biber 1995: 32).

While the use of corpora in monolingual studies became more and more popular, corpus-based approaches received very little attention in translation studies at first. Most general introductions to corpus linguistics like Biber, Conrad, and Reppen (1998) and McEnery and Wilson (2001) covered a variety of different areas of linguistic application but did not include translations at all. This lack may have been due to the limited exchange between linguistics and translation studies or to the unfavorable perception of translated texts as not representing authentic language use (Olohan 2002: 419). This changed when Baker (1993, 1996) strongly argued for the use of corpora to study the product and process of translation. Like monolingual corpora, corpora of translation can be designed in different ways, allowing comparisons between different types of texts. Some translation corpora include original and translated texts of one language, which allows the translation scholar to work out the linguistic differences between these two different kinds of text productions. A comparison between original writing in the source language and translated writing in the target language sheds light on the influence of the source language during the translation process. A corpus of translations in various languages helps

to uncover features and tendencies that result from the process of translating itself and that are not based on particular language pairs (Johansson 1997: 283).

A bidirectional translation corpus is a corpus that includes two or more language pairs with original and matching translated texts in all of the languages included. Ideally originals and translations are taken from a variety of registers. Aside from the areas of applications outlined above, such a bidirectional translation corpus can be used for contrastive and typological research designs as well as register studies. The fact that the translations match the original texts allows for an accurate assessment of the interference of the source text (Neumann and Hansen-Schirra 2012: 25).

Early corpus-based studies on translations followed Baker's (1995: 235) suggestion to study "patterns which are either restricted to translated text or which occur with a significantly higher or lower frequency in translated text", also referred to as translation universals (see Section 2.2). These studies covered various language pairs and dealt with linguistic patterns that are characteristic of translated texts (see for example Vanderauwera 1985; Shlesinger 1991; Baker 1992; Hansen 2003; Toury 2012). Since then, the landscape of corpus-based translation studies has expanded and now covers a variety of languages, topics, and areas of application, including for example lexical choices and creativity in translations (Kenny 2001), cohesion (Klein 2007; Hansen-Schirra, Neumann, and Steiner 2007) editorial interventions (Kruger 2017; Bisiada 2018), and register effects (Neumann 2014), to name only a few.

2.4 Translation and register

While translation studies have put a great focus on the relationship between source and target text, typological differences between particular language pairs, and, more recently, translation-inherent processes, one factor that has been largely overlooked is register. The effects of registers in translations are briefly addressed in Baker (1992) as well as in Hatim and Mason (1997), who do include some aspects of register analysis in their text analysis model. However, systematic, and detailed analyses on the influence of register variation and contrastive register differences in translations have been undertaken only recently.

From a systemic functional perspective, register analysis refers to the analysis of the context of use or context of situation of language use (Halliday 1978). A very similar definition of register is used by Biber (1988, 1995) and Biber and Conrad (2009), describing register as “any variety associated with particular situational contexts or purposes” (Biber 1995: 1).² The underlying assumption of register theory is that the language use context has an influence on language exchange (Halliday and Matthiessen 2014: 29). Any instance of language use comes with a set of situational parameters that have an influence on what language features are more or less likely to be chosen by the speakers. Language usages in situations that are similar are thus more likely to feature similar feature distributions. These similar contexts of situations can then be grouped into registers, which in most cultures receive their own labels like newspaper articles, telephone conversations or political speeches. Different registers can vary in their degree of specificity regarding situational contexts, which also affects the level of similarity between the different texts of a register.³

In Hallidayan register theory, the context of situation can be divided into three broad domains: field, tenor, and mode of discourse (Halliday and Hasan 1985). The field of discourse refers to the nature of the social interaction, which is largely determined by the subject matter but also by “the activities that make the subject matter relevant – activities such as exploration, learning, and imagining” (Matthiessen 1995a: 33). The tenor of discourse describes the relationship between the participants involved in the social interactions, which are mainly speaker and hearer but possibly also other parties. Essential to the analysis of tenor of discourse are the speech roles that speaker and hearer engage in (see Section 4.4) and the communicative functions that are being pursued. Lastly, mode of discourse refers to the way that the exchange is transmitted through language and “the particular status that is assigned to the text within the situation; its function in relation to the social action and the role structure, including the channel or medium, and the rhetorical mode” (Halliday 1978: 143). These three situational parameters also correspond to the three metafunctions in the systemic functional framework (see Section 4.2).

² The state of the art makes a difference between register, genre, and text type. However, for the purposes of this thesis, a distinction between these concepts is not necessary and instead only the term register will be used, which may correspond to what other studies refer to as genre or text type. For a comprehensive summary of the differences between these three concepts, see Biber (1995: 7-10).

³ For example, Biber (1995: 1) uses the example of methodology sections in psychology articles, which is a highly specific situational context and is therefore also heavily conventionalized regarding frequent and infrequent linguistic patterns.

While the systemic functional understanding of register is very similar to that of Biber (1988, 1995) and Biber and Conrad (2009) in that both define register based on context, Biber's approach is more inductive, focusing on linguistic features to explain the context of situation. The systemic functional approach is deductive in the sense that different registers are identified first based on differences regarding the three situational parameters and the resulting linguistic variation is analyzed thereafter (Neumann 2014: 37). In the context of this Ph.D. project, all analyses are based on the registers included in the translation corpus CroCo (Hansen-Schirra, Neumann, and Steiner 2012), and the texts that are contained in each of the registers are assumed to have been produced in similar contexts of use. In this sense, the systemic functional approach to register analysis is followed in this thesis since a difference in context of use is assumed at the outset and linguistic variation is analyzed based on this assumption.

The situational context influences linguistic choices. However, it is rare that different registers can be distinguished on the basis of unique features which occur exclusively in one register, known as register markers (Biber and Conrad 2009: 53). Differences in linguistic features between different registers are, for the most part, a question of frequency so that certain features and patterns are characteristic but not unique in a register. A register filters the general language system and alters the likelihood of certain linguistic choices to be made (Matthiessen 1993: 258). A linguistic analysis of register thus includes identifying pervasive and frequent linguistic features and contrasting these sets of features with other registers to assess the impact of the situational context on language use. Such analyses require a quantitative and ideally also a comparative approach, where the linguistic make-ups of multiple, independent texts of a register are studied and then contrasted with texts of other registers. A quantitative approach also has the advantage that very subtle but revealing differences in linguistic patterns become visible, which would have remained unnoticed in a study of only a small number of texts (Biber and Conrad 2009: 56).

Considering register characteristics is crucial for analyses of any speech event and has long been recognized as a significant variable in general linguistic studies. However, in translation studies, register has been looked at as a methodological necessity rather than being the focus of translation analyses (Neumann 2014: 28-29). Steiner (1997, 2001a, 2004) has repeatedly stressed the significance of register for the evaluation of text and even considers register to be one of the three sources, and explanations, for properties of

translated text, besides typological factors, and properties of the translation process (Steiner 2001b: 5).

One possible explanation of why register has seen only little consideration in translation studies may be that for most translations, the register remains the same for both source and target text and that presumably register differences are not relevant in translations. However, as Johansson (2002: 51) notes, even if source and target text can be sorted into roughly the same register, the contextual parameters may still differ between cultures and, as a consequence, be responsible for changes in the translation.

Most corpus-based studies on translation that consider register effects focus on one register in particular and compare originals of different languages as well as source and target text features in translations. For example, Neumann (2003) focuses on the register of tourism leaflets in English and German originals and German translations. Following a systemic functional approach, she analyzes a multitude of linguistic features from all three metafunctions. Regarding Theme,⁴ she found an increase in circumstantial Themes of Place in both English and German original tourism leaflets compared to a reference corpus, which she attributes to the importance of spatial references in the register (Neumann 2003: 163). This increase is more distinctive in German originals, which can be explained by the freer word order in German (Hawkins 1986: 37). Theme distributions in the German translations are largely in between English and German originals but closer to the German originals regarding Predicator Themes and spatial circumstance Themes.

In her study on language typology, contrastive linguistics, and translation studies of English and German, Teich (2003) analyzes English and German originals and translations from the register of popular scientific writing. Teich (2003), too, follows a systemic functional approach and investigates different linguistic features, such as Theme markedness, both from a contrastive and translation studies perspective.⁵ Regarding translations, she focuses specifically on the influences of the source text and the target language system by investigating shining through and normalization effects (see Section 2.2). In her analysis of Theme markedness, she found a significant difference between English and German originals, with English featuring more cases of marked Themes (Teich 2003:

⁴ It is important to note that in both Neumann (2003) and Neumann (2014), Theme was analyzed as the very first element of the clause for both English and German. As will be shown in Sections 5.4 and 5.8, Theme is identified differently here, which is why her results are only partially comparable with the results of this thesis.

⁵ Teich (2003) also analyzes agency in terms of effective and middle constructions, which overlaps with the analyses on agency in this paper to some extent. However, her methodology contrasts so heavily from the Subject agency analysis in this thesis that a comparison does not seem plausible.

185-186). Comparing English originals with English translations, she reports a higher frequency of marked Themes in the translations, which, however, is not statistically significant. Similarly, German translations use fewer marked Themes in comparison to German originals, which also do not result in statistical significance. Nevertheless, Teich (2003: 194) attributes these effects to shining through and normalization. Teich's popular scientific texts were later included in the CroCo Corpus (Hansen-Schirra, Neumann, and Steiner 2012), which is the same register used in Freiwald's (2016) study on Theme in translations between English and German.

Also working with the CroCo corpus, Neumann (2014) contrasts two registers with each other, namely fictional writing, and letters to company shareholders. With this comparative approach, she is able to not only control for register effects, but also directly identify the influence the registers have on original texts as well as on translations. Regarding Themes, for instance, Neumann (2014: 254) demonstrates that German generally features more marked Themes than English, especially in terms of Complement Themes. In both languages, however, the Theme distribution is heavily dependent on register, with Complement Themes being generally more common in fictional writing than in letters to shareholders. Thematic structures in the English and German translations are situated in between English and German originals, which is a sign of shining through and normalization. The translations in the fictional register corresponded more closely to the source texts, while translations of letters to shareholders followed the target language distributions more (Neumann 2014: 304). This shows that registers not only affect the structure of source texts but also influence the extent to which translation-inherent processes are at play.

3 English and German clause structure

This chapter represents an excursion into the basic clause structures of English and German. Given the way in which Theme is formally realized in these two languages, a thorough understanding of word order is essential. These grammatical descriptions are not based on systemic functional concepts but are presented from a theory-neutral perspective. The chapter leads out with the common clause structure of English declaratives and continues with the basic word order of German declaratives. Central concepts of German syntax are introduced here, including the German clause frame and the three resulting topological fields, of which two are discussed in more detail. The chapter concludes with a discussion of syntax and semantics in the two languages with a focus on the relationship between Subject meaning and Verb requirements.

3.1 English clause structure and comparison to German

One of the most common syntactic descriptions of English is that it is a fixed word order language. The most dominant word order in English declaratives is S-V-O with very little room for alternation (Biber et al. 1999: 123). If a constituent other than the Subject is placed in initial position, the Subject-Verb sequence still stays intact.⁶ Present-day English has lost most of its inflectional forms, particularly case marking, which is one grammatical resource to express relations between constituents. In fact, the position of the Subject before the Verb is one of the criteria on how to identify the Subject function in English.

English is what Thompson (1978) calls a grammatically determined language, where grammatical considerations determine the sequence of clause constituents, as opposed to a pragmatically determined language, where word order is predominantly determined by pragmatic consideration and information structure. Despite this rigidity of word order in English, it is of course not accurate to assume that no variation in the sequencing of clause elements can be found. Both the Subject and the Verb are generally inflexible in

⁶ There are a few exceptions to this rule where Subject and Verb order is inverted, for example in some relational processes where the Attribute is fronted: *Central to this approach is the concept of choice*. These word order variations are likely traces of older stages of English where the inflectional system was more complex and word order more flexible (Hawkins 1986: 12).

their position relative to each other but Adjuncts and to some extent also Objects are generally movable to adhere to considerations other than grammatical ones. Adjuncts in particular are rather flexible in their positioning and can occupy both pre- and post-verbal positions in English (Thompson 2014: 149). Even Objects have the potential to be the very first constituents in English declaratives, albeit only in rare, marked cases (see example (1); Object in bold). In those cases where English does demonstrate a certain degree of word order flexibility, other sequencing principles such as old information before new information (Hawkins 1986: 43-44) and syntactically lighter constituents before heavier constituents (König and Gast 2009: 162) can be observed in English.

(1) '**Hard Hand Stan** they called him,' she said every time I went to see her.

[E2G_FICTION_008]

To make up for this lack in word order flexibility, English makes use of an array of syntactic structures to maintain the grammatically governed S-V-O sequence and still vary different types of meaning in different positions in the clause. One common strategy of alternating the order of experiential meaning (see Section 4.3) is the passive, which allows the undergoer of an action to take initial position in the clause. Of course, other languages have this grammatical resource as well, but it is arguably more productive in grammatically determined languages like English (Firbas 1964a). One other syntactic strategy to break up the static Subject-Verb sequence are focus constructions, in particular cleft and pseudo-cleft constructions (Erdmann 1990b: 69-70).

Finally, because of its more rigid word order, English has developed a flexibility regarding the mapping of semantic meaning onto grammatical functions. That is to say, grammatical functions in English are less restricted in what kind of meaning they can refer to (Kast 2012: 148). This flexibility allows non-agentive or non-sentient constructions in English (see Section 3.5). They, too, are one resource to ensure variation in information structure.

Despite the fact English and German are both Germanic languages, there are clear differences between their basic clause structures. German still has a rich inflectional system which expresses grammatical relations, such as the distinction between four cases: nominative, genitive, dative, and accusative. For this reason, German is less dependent on a fixed word order and can therefore vary word order more freely than English can. That

being said, word order permutations in German are not as free as in some of the prototypical pragmatically determined languages like Russian and Czech. Thus, German assumes only an intermediate position in the range of grammatically and pragmatically determined languages (Hawkins 1986: 44). Nevertheless, word order variations are still more common in German than in English. Fronting non-Subject constituents is generally possible in both languages, but from a frequency perspective, Complements and Adjuncts in early position are overall more common in German (Freiwald 2016). There are also some non-Subject options available for the German Forefield which do not exist in English, for example the Experiencer dative and accusative (Steiner 2001b: 3).

Subject-Verb positioning is also one major contrast between English and German as English has a strict order of Subject followed by Verb whereas German inverts Subject and finite Verb positions if a different constituent occupies first position (see example (2); Subject in bold, finite Verb underlined). In this respect, English is the odd-one-out since the Finite-second constraint is productive among other Germanic languages (Kim and Matthiessen 2015: 341). As a consequence, the finite verb often comes in third, sometimes fourth or fifth position in English declaratives, which is virtually impossible in German.

(2)

- a. *And today **Japan** has sent its SelfDefense Forces to Iraq [...].*
- b. *Und nun hat **Japan** seine Selbstverteidigungstruppen in den Irak entsandt.
'and now has **Japan** its SelfDefense.forces to Iraq sent.'*

[E2G_SPEECH_013]

Most of the grammatical resources that are available to English to change the dominant sentence structure like passives and focus constructions are also available in German. However, Doherty (1996) shows that while the options for passive and passive-like structures are similar between the two languages, they are utilized less often in German. Erdmann (1990b) shows the same for focus constructions: All types of cleft and pseudo-cleft constructions are generally present in both language systems, and yet they are 50.3% more frequent in English (299 clefts in English to 199 clefts in German). The reason for this preference of alternative constructions in English is clear: Since German has more clause structure flexibility, there is also less need for constructions, such as clefts, that

break up the dominant word order. For the same reason, there are also fewer cases of non-sentient constructions in German (see Section 3.5).

Yet, it is also important to note that despite all of these grammatical differences between English and German clause structure, the two languages are also very similar in many respects. German may be less restricted in the kinds of non-Subject elements to be placed in initial position; however, the Subject is still the most common first constituent in German as well (see Section 7.1.1). Also, while frequency distributions may vary, most syntactic options are still available in both languages. So, in spite of the contrastive differences discussed here, it remains evident that English and German do descend from the same proto language.

3.2 The basic word order in German

As was shown in the previous section, English allows some flexibility in the way that clause elements are ordered and yet there is generally a dominant way that a clause is structured according to the mood type. In German, it is debatable whether there is such a basic or dominant word order.⁷ Arguably, the most central element in German clauses is the finite verb. Engel (2004: 87) even considers it the most important element of the German sentence by far because its position signals the relationship between the proposition of the clause and reality. What that means is that the position of the finite verb is responsible for expressing the mood in German and thus specifies whether a clause constitutes an assertion about reality in the form of a declarative clause or an inquiry about reality in the form of a polar interrogative, for example.

Generally, three different positions of the finite verb are possible in German: finite-first, finite-second, and finite-last clauses. The finite verb is in first position in polar interrogatives and imperatives as well as conditional and wishing clauses (Mode 1987: 3). Finite-last constructions are mostly subordinate or embedded clauses such as relative

⁷ The following descriptions of the German word order are theory-neutral and focus on the most basic positional rules in German clause structure. The only SFL-based descriptions of clause structure in German that I am aware of are Steiner and Ramm (1995) and Steiner and Teich (2004), which do not focus on the sentence frame or constituent sequences in particular. Other frameworks have studied German clause structure using approaches like GPSG (Uszkoreit 1987) or scrambling (Rosengren 1993), which will not be described in detail.

clauses and most hypotactically dependent clauses (Mode 1987: 4). The finite verb occupies the second position amongst others in independent declarative clauses and *wh*-clauses, which makes finite-second constructions the most commonly found clause structure in German (Mode 1987: 4).

The position of the finite verb alone is not sufficient to determine the mood type, as different kinds of moods feature the same finite position. Other aspects such as intonation, orthography, and the content of topological fields contribute to mood in German (Altmann 1981: 145). However, unlike English, the Subject in German does not have "a high functional load for expressing Mood" (Steiner and Teich 2004: 180) as its position in the clause is generally not significant for the mood type. While the distinction between different clause structures on the basis of the finite verb position is helpful, it is also important to note that the position of the finite verb is not absolutely fixed. For example, it is possible to add a (peripheral) element to the beginning of a finite-first construction without turning it into a finite-second construction. In example (3), the imperative, which is a finite-first construction, is opened up by a Vocative (in bold).

(3) **Thomas**, *mach mal das Fenster auf.*
'**Thomas**, open sometime the window.'

In independent declarative clauses, the finite verb typically occupies the second position; nevertheless, no consensus has been reached concerning the question of whether German is an SVO or SOV language (for example Ross (1970), Esau (1973), and Huber and Kummer (1974) argue for an SVO order, while Bach (1962) and Haider (2000) are in favor of German being an SOV language). The reason for this dispute stems from the relationship between the position of the finite verb and the rest of the verb phrase. If the finite is conflated with the lexical verb, the verb phrase comes second. However, if the German verb phrase is more complex, only the finite portion of the verb phrase is positioned second and the remainder of the verb phrase is moved to the back of the clause.

The question of whether the dominant order in German is SVO or SOV thus rests entirely on the question which part of the verb phrase is considered central, the finite verb or the lexical verb (plus additional auxiliaries). Voyles (1978: 14) makes a valid argument for the latter, saying that children learning German as their native language often use SOV structures in contexts where the verb should have been in second position. He takes this

as an indication that the deep structure of German clauses is in fact an SOV order. Personally, I consider the finite portion of the verb to be more relevant for German clause order given its responsibility for mood. I would thus rather side with the SVO classification. This is supported by the fact that the positional criteria of the finite override the positional criteria of the lexical verb if they are conflated. Yet, it is important to note that in such cases the verb occupies second position due to its role as the finite and not as the lexical verb (which will be relevant when discussing the role of the finite in the German Theme, see Section 5.8).

The question of the dominant word order will not be answered definitively here. One alternative would be to argue that there is no dominant order (Etzensperger 1979: 38) or that German is a mixture of an SOV and SVO language (Steiner and Teich 2004: 143). In any event, it would be more accurate to call German a $SV_{\text{fin}}OV_{\text{non-fin}}$ language but then it would not fit into Greenberg's (1990) classification of universal word orders (which shows that such a classification is overly simplistic).

Up to this point, it has been assumed that the Subject occupies the first position in the German declaratives or at least a position prior to the finite verb. And, in fact, this order of Subject and finite verb is the most common word order in German. However, it is not uncommon to find a different element than the Subject occupy the very first position. This positional flexibility is due to German's case-marking, which signals grammatical functions like Subject and Object based on inflection (Steiner and Teich 2004: 152). Given that the position of the finite verb is rather fixed, the Subject-Finite order is inverted if a different element is moved to the beginning of the clause. Almost any clause element, whether of experiential or non-experiential nature (see Section 4.3), can be positioned first in German declaratives. Example (4) shows five different clause structures of the same underlying clause, each featuring a different clause element in the Forefield: a. Subject, b. Circumstantial Adjunct, c. Object, d. conjunctive Adjunct, e. modal Adjunct. If a clause includes a coordinating conjunction, this conjunction does not exhaust the first position in the declarative and is therefore not directly followed by the finite verb (see example (5)).

(4)

- a. *Mein Vater hat gestern ein neues Auto gekauft.*
'my father has yesterday a new car bought.'
- b. *Gestern hat mein Vater ein neues Auto gekauft.*
'yesterday has my father a new car bought.'
- c. *Ein neues Auto hat mein Vater gestern gekauft.*
'a new car has my father yesterday bought.'
- d. *Außerdem hat mein Vater gestern ein neues Auto gekauft.*
'moreover has my father yesterday a new car bought.'
- e. *Glücklicherweise hat mein Vater gestern ein neues Auto gekauft.*
'fortunately has my father yesterday a new car bought.'

(5) *Und mein Vater hat ein neues Auto gekauft.*
'and my father has a new car bought.'

Despite this positional flexibility in German, the word order is by no means completely free. Especially in the position after the finite verb, known as the Midfield, elements cannot be just randomly ordered (see Section 3.4). Heidolph, Flämig, and Motsch (1981: 703) argue that German generally follows a communicative-pragmatic word order, which is to say that in general, given information precedes new information. However, this sequence is only a tendency and can be disregarded if other criteria like rhythmic or semantic criteria override it (Admoni 1970: 295-296).

To accurately model the word order of declarative clauses, Höhle (2019b: 83) ultimately arrives at the following formula:

(KOORD)—(KL*)—K—fV>(KM*)>(iV*)⁸

(6) *Aber (KOORD) dem Günther (KL*), dem (K) kann (fV) man nicht (KM*) trauen (iV*).*
'but (KOORD) the Günther (Kl*) him (K) can (fV) you not (KM*) trust (iV*).'

⁸ (KOORD) = coordination, (KL*) = possible fronting of K, K = one preverbal element, fV = finite verb, (KM*) = element(s) following the finite verb, (iV*) = infinite verb(s)

As detailed as this formula is, it does not account for the possibility of something following the infinite verb and should therefore be changed to:

(KOORD) — (KL*) — K — fV > (KM*) > (iV*) > (KN*)⁹

(7) *Aber dem Günther, dem hab ich damals gesagt, dass er das Haus kaufen soll.*
'but the Günther him have I then told that he the house buy should.'

This revised formula marks the general word order in German declaratives. The finite and infinite parts of the verb phrase are tied to particular positions in the clause, whereas all other fields can be filled or not filled by a variety of different elements. This structure of the verbal positions and the fields before, in between, and after the verbs is called *Satzklammer* or *Satzrahmen* (*sentence frame* or, more accurately, *clause frame*), which is considered the underlying clause structure in German (Admoni 1962: 166).

The fact that German clause structure resembles that of a frame had been observed as early as 1691 (Etzensperger 1979: 44). Drach (1963) and Boost (1964) were among the first to recognize the significance of the two verb positions and the resulting topological fields for the word order in German. Originally, the German clause was only divided into two fields, the *Vorfeld* (Forefield), which was the field before the finite verb, and the *Nachfeld* (Postfield), which was the field following the finite verb (Drach 1963: 17). However, in later descriptions of German clause structure, the Postfield was divided further into *Mittelfeld* (Midfield), which is the field in between finite and infinite verbs, and the new Postfield, which then only referred to the field following the infinite parts of the verb phrase (see for example Engel 1970; Mode 1987; Dürscheid 1989; Zifonun, Hoffmann, and Strecker 1997). This division into different fields is not arbitrary, as all fields have unique rules of which elements can occur, must occur, and are likely to occur there. Example (8) represents a German declarative clause, in which all three fields are realized.

⁹ (KN*) = element(s) following the finite verb; Höhle (2019b) is well aware that the infinite part of the verb phrase is not necessarily the last element in German declaratives. The reason why he did not account for the possibility of a post-verbal element in his formula is that he was planning on continuing this paper on topological fields in German, but never did (see Müller, Reis, and Richter 2019: 4).

(8)

Forefield	Finite	Midfield	Lex. verb	Postfield
<i>Wir</i>	<i>können</i>	<i>deswegen gerne die Hoffnung</i>	<i>aussprechen</i>	<i>dass sich die deutschen Län- der ähnlich verhalten.</i>
<i>'we</i>	<i>can</i>	<i>hence gladly the hope</i>	<i>express</i>	<i>that [refl-3sg] the German states similarly act.'</i>

[G2E_SPEECH_012]

As was pointed out earlier, the finite verb can conflate with the lexical verb if it has simple aspect. However, this does not change the division between the three fields. In other words, just because the infinite part of the verb is missing in a given clause does not mean that the Postfield merges with the Midfield. The sentence frame is virtual and underlies every finite clause structure in German, which can be demonstrated by a transformation from simple to perfect aspect (Hoberg 1981: 26; see example (9): In a. finite and lexical verb are conflated and in b. they are separated).

(9)

a.

Forefield	Finite	Midfield	Lex. verb	Postfield
<i>Ich</i>	<i>schrieb</i>	<i>meinem Bruder</i>		<i>dass ich am Montag in Köln bin.</i>
<i>'I</i>	<i>wrote</i>	<i>my brother</i>		<i>that I on Monday in Cologne am.'</i>

b.

Forefield	Finite	Midfield	Lex. verb	Postfield
<i>Ich</i>	<i>haben</i>	<i>meinem Bruder</i>	<i>geschrieben</i>	<i>dass ich am Montag in Köln bin.</i>
<i>'I</i>	<i>have</i>	<i>my brother</i>	<i>written</i>	<i>that I on Monday in Cologne am.'</i>

The sentence frame is not a unique structure of German but also underlies other Germanic languages that have the finite in second position in declaratives, for example Norwegian (Hasselgård 2000: 16).

3.3 Forefield

The Forefield is one of the three topological fields in German, which is defined positionally as the zone before the finite verb. Forefields are usually associated with statements, but any kind of finite-second construction in German has a Forefield. Finite-first constructions like polar interrogatives and imperatives cannot have a Forefield by definition.¹⁰

A variety of different clause elements can be positioned in the Forefield. There are two elements that, if present in the clause, must occupy the Forefield position, namely the expletive *es*-Subject and the *wh*-element in *wh*-interrogatives (Hoberg 1981: 158). These expletive *es*-Subjects are semantically empty, non-referential Subjects that only act as a placeholder in German if a Subject is missing (Engel 2004: 164). Unlike other Subjects, the expletive *es* does not switch positions with the finite verb if a different element is positioned in the Forefield but simply disappears, resulting in a subject-less clause (see example (10); Subject in bold).

(10)

- a. **Es** wurde auf der Party viel getanzt.
'**it** was at the party much danced.'
- b. Auf der Party wurde viel getanzt.
'at the party was much danced.'

There are some syntactic structures in German that cannot be in the Forefield, for example referential *es*-Objects and reflexive pronouns (Hoberg 1981: 159-161). Other than these, any kind of clause element can be positioned before the finite verb including but not limited to the Subject, Objects, conjunctive, modal and circumstantial Adjuncts, the lexical verb and even the entire Predicate (Dürscheid 1989: 3). Of course, not all elements are equally likely to be positioned in the Forefield and these probabilities are also heavily dependent on mood.

There are different reasons why an element is positioned first in German declaratives. Given that German clause structures are not as syntactically restricted as in English for example, it is commonly assumed that German word order is primarily based on pragmatic-communicative principles (Beneš 1971: 174). For this reason, the Forefield is said

¹⁰ There is no consensus on whether hypotactic dependent clauses have a Forefield. As finite-last constructions, dependent clauses would entirely consist of the Forefield (see for example Hoberg (1981) in favor of Forefields in dependent clauses and Götze and Hess-Lüttich (2002) against).

to primarily contain elements that represent given information (Hoberg 1981: 167).¹¹ If given information is positioned first, this element is usually unaccented, and the stress of the information unit falls on the very last element of the clause. It connects the clause at hand with its context, thus increasing cohesiveness (Brinkmann 1971: 505), and it allows the hearer to interpret the rest of the clause on the basis of something familiar. Even if there is more than one piece of given information in a clause, the element that receives the positional prominence of coming first plays an important role for the interpretation of the clause (Beneš 1973: 46).¹²

However, Forefields are not restricted to given information, but can also feature information that has not been mentioned before (Beneš 1973: 46). In fact, there is debate over whether new information is just as likely or even more likely to appear in the Forefield. Originally, it was assumed that new information occupies the Forefield position primarily in spoken, colloquial language (Beneš 1971: 170), but Engel (1974) shows that new information in the Forefield is much more likely to be found in written language, namely in 57% of cases in written compared to 21% in spoken language. Regardless of frequencies, it can be argued that new information in the Forefield constitutes marked word order in German because the default intonation pattern changes with the accent or focus of the information falling on the Forefield instead of the final element in the clause (Götze and Hess-Lüttich 2002: 482). Oftentimes, such a shift in intonation leads to two information units, meaning two foci, in one clause with the first focus falling on the Forefield element and the second focus on another one, usually the last element in the Midfield (Heidolph, Flämig, and Motsch 1981: 755-756). Additionally, the sequence of new information followed by given information needs to be contextually motivated, whereas given information followed by new information is largely context independent (Höhle 2019a: 186).

In her detailed quantitative study of indefinite clause elements in German and Norwegian Forefields, Fabricius-Hansen (2016) found that the majority of Forefield constituents do not refer to new information but take up a referent that was either mentioned in the context or at least implied, irrespective of its function in the clause. Nevertheless, she also found that the Forefield in German could be occupied by indefinite, new information

¹¹ Given information is usually called *Thema* following the *Thema-Rhema* distinction of the Prague School. This thesis follows Halliday and Matthiessen's (2014) terminology and uses the label *Given* to refer to information that is either known or inferable from the context. For a detailed discussion on the relationship between *Theme*, *Given*, and also communicative dynamism, see Sections 5.1.1 and 5.1.4.

¹² Beneš (1973) calls this function of the Forefield the *Basis (base)* or *Ausgangspunkt (point of departure)*, which is remarkably similar to how *Theme* is defined in SFL.

if this choice is well-motivated. So, while there seems to be a tendency for the Forefield to contain given information, new information can be easily moved before the finite verb if the context requires it.

Related to the relationship between Forefield, given and new information, and intonation is the question of which clause element most commonly occupies the Forefield position. Regarding intonation, only the Subject can be considered an unmarked Forefield choice since only the Subject in first position leads to an unmarked stress pattern (Bierwisch 1973: 111). According to Dürscheid (1989: 37), all other clause elements in Forefield position are more marked since a marked intonation also restricts the number of contexts in which a clause can be used. However, Dürscheid also argues that while the Subject-Forefield conflation represents the only unmarked clause structure, other kinds of clause elements like temporal Adjuncts can also be used in a variety of different contexts and therefore receive a high acceptability rating by German natives (Dürscheid 1989: 37).

This high acceptability of non-Subject Forefields also matches their frequency in texts. Engel (1974) shows that only slightly more than 50% of all Forefields contain a Subject, while the rest are primarily Adjuncts or to a lesser extent Complements.¹³ Heidolph, Flämig, and Motsch (1981: 715) argue that temporal Adjuncts especially can frequently be found in the Forefield because they are often given or inferable from the context (see example (11); temporal Adjunct in bold, Subject underlined).

- (11) ***Nach der Dekomprimierung** finden Sie im gewählten Verzeichnis die Datei **BE-
UNEA**.*
*'**after the decompression** find you in.the chosen register the file **BEUNEA**.'*
[G2E_INSTR_002]

Based on these frequency distributions between Subject and non-Subject elements, Dürscheid (1989: 18) questions whether the Forefield position can actually be considered the default Subject position. Similarly, Hoberg (1981: 163) challenges the distinction between marked and unmarked word order in German given the similar distribution of clause elements.

¹³ Freiwald (2016) reports an even more surprising distribution in popular scientific writing, where circumstantial Adjuncts are more likely to be positioned in the Forefield than Subjects. In their corpus analysis, Winter (1961) found that 65% of Forefields contain a Subject; however, they also analyzed some constituents as Subjects which in this thesis would be considered Complements.

There are different reasons to have a non-Subject element in Forefield position. Non-Subject elements most likely constitute new information but may also be given. If they are given, Adjuncts and Objects can also link the clause with its context or serve as an appropriate point of departure from which the rest of the clause can develop, just like given Subjects can. Alternatively, a non-Subject element can also be fronted for highlighting purposes (Admoni 1970: 301). One common reason for highlighting an element in Forefield position is contrastiveness, where the expectations of the hearer are subverted by either restricting or rebutting previously mentioned information (Dalmas 2008: 91). Similarly, if a non-Subject element is particularly emotionally charged, the speaker may also decide to highlight this element by moving it to the early position (Engel 1970: 90). In both of these cases, the Forefield receives the phonological focus of the information unit, which constitutes a marked intonation pattern (Fabricius-Hansen 2016: 90-91).

To summarize, the Subject is the least marked clause element to appear in the first position in German declaratives, both in terms of frequency and in terms of intonation. Nevertheless, non-Subject elements can be made the Forefield very easily if the context motivates this fronting. This explains the particularly high number of Adjunct Forefields in German.

Brinkmann (1971: 503) suggests that the causes for these changes to the unmarked word orders do not necessarily have to do with the non-Subject element. Instead, he argues, the Subject can also be moved out of the Forefield position deliberately, thus leaving a void in the Forefield that needs to be filled by another element. He suspects that the position in the Midfield right after the finite verb represents a stronger relationship to the verb phrase. Unfortunately, Brinkmann does not elaborate on why the connection to the verb is stronger following the finite verb instead of preceding it, which is why I do not find this argument very compelling. I do, however, believe he is right that in some instances it is not the non-Subject that is chosen as the Forefield but rather the Subject that is chosen as not being the Forefield.

The Forefield is usually described as consisting of only a single constituent (for example Drach 1963; Paul 1968). In fact, one of the fundamental characteristics of German clause structure is that the finite comes in second position in declaratives, which makes German one of the V2 languages. However, it is also quite clear that there can easily be additional independent elements, for example conjunctions and Vocatives, without violating grammaticality. The possibility and acceptability of multiple Forefield elements in

German has been and continues to be a popular topic among German grammarians (see for example Hoberg 1981; Müller 2005; Bassola and Schwinn 2016; Uzonyi and Dabóczy 2016).

One of the most common arguments against multiple Forefields is that the majority of the elements that can be added to the position before the finite verb are "outside of the Subject-Predicate relationship"¹⁴ (Brinkmann 1971: 475), and thus do not actually occupy the Forefield but rather stand in a peripheral field known as linkes Außenfeld (left outfield).¹⁵ Constituents that are considered part of the left outfield are conjunctions (12a), Vocatives (12b), left dislocations (12c), certain conjunctive Adjuncts (12d), certain modal Adjuncts (12e), and continuatives (12f) (Brinkmann 1971: 475-476; Bassola and Schwinn 2016: 233).

(12)

- a. *Und das ist erst der Anfang.*
'and that is just the beginning.'
- b. *Georg, das Geschenk ist nicht für dich.*
'Georg, the present is not for you.'
- c. *Dem Uwe, dem würde ich nicht trauen.*
'the Uwe, him would I not trust.'
- d. *Jedoch meine Mutter hätte das nicht erlaubt.*
'yet my mother would have that not allowed.'
- e. *Immerhin du bist auf meiner Seite.*
'at least you are on my side.'¹⁶
- f. *Nun ja, das würde ich so nicht sagen.*
'well, that would I so not say.'

The state of the art is not terminologically consistent when referring to these peripheral fields; for example, Engel (1970) uses the term *Erstglied* (first constituent) to describe what others consider Forefield and the term Forefield to refer to everything before the finite verb including the left outfield. Linguists such as Engel (1970), Beneš (1971), and van de Velde (1978) question the usefulness of a distinction between Forefield and left outfield.

¹⁴ Original quote: "außerhalb der Subjekt-Prädikat-Beziehung"

¹⁵ There is a second peripheral field called *rechtes Außenfeld* (right outfield) that follows the Postfield (Bassola and Schwinn 2016: 229).

¹⁶ Conjunctive and modal Adjuncts can generally also be Forefields, meaning that they can constitute the only constituent before the finite verb. So an alternative, possibly less marked word order of (12d) and (12e) would be *Jedoch hätte meine Mutter das nicht erlaubt* and *Immerhin bist du auf meiner Seite*. When these Adjuncts do stand in the left outfield and are followed by another element in the Forefield, one would expect a slight pause between the Adjunct and the Forefield element in oral discourse, which signals their peripheral status.

Regardless of whether a left outfield exists, there are other examples of multiple elements before the finite verb which cannot be argued to be peripheral. It is possible to include multiple constituents in pre-verbal position which contribute to the experiential meaning of the clause. A frequent combination of multiple Forefield elements is for instance two or more circumstantial Adjuncts which set the stage for the rest of the clause (see example (13); first Adjunct in bold, second Adjunct underlined).

- (13) **Am Sonnabendmorgen** im frühesten Zug war er unausgeschlafen genug, das ganze Unternehmen zu verfluchen
'**on Saturday.morning** in.the earliest train was he unrested enough, the whole company to curse.'

(taken from Van de Velde 1978: 134-135)

Some argue that even though these circumstantial Adjuncts can be moved to different positions in the clause, they merge into a single constituent if positioned together before the finite verb (see for example Brinkmann 1971; Faucher 1976). Van de Velde (1978) also believes that two Adjuncts that are both positioned before the finite verb have a stronger relationship to each other than if they were positioned in the Midfield. Van de Velde (1978: 134-135) questions, however, whether this is a good enough argument to consider them a single constituent.

It is somewhat plausible to merge multiple circumstantial Adjuncts in the Forefield, at least from a functional perspective, because you can argue that they serve a single function as one frame setting element. However, circumstantial Adjuncts are not the only elements that can be combined in the Forefield. It is possible to combine Objects, lexical verbs, and Subjects with each other without disrupting them with the finite verb in second position (see Müller 2005 and Bassola and Schwinn 2016 for detailed lists of different possible combinations of Forefield elements). And the same general sentence can be separated into different possible Forefields with different numbers of Forefield elements, which makes it even less plausible to speak of a single constituent in Forefield position in these cases. Example (14) opens up with a circumstantial Adjunct, two Objects and the process before the Forefield is concluded by the finite verb (finite verb in bold).

- (14) *Mit ruhiger Stimme seiner Tochter ein Märchen erzählen* **kann** er.
'with a.calm voice to.his daughter a fairytale tell **can** he.'

(taken from Heidolph, Flämig, and Motsch 1981: 720)

Such constructions that include the lexical verb and other experiential elements in the Forefield are only used in 0.13% of declarative clauses (Uzonyi and Dabóczy 2016: 56), which arguably makes them a fringe phenomenon. Bassola and Schwinn (2016: 257) suspect that text types highly influence the frequency of multiple Forefields, claiming that especially text types that are in between written and oral communication like online discourse are very likely to feature multiple Forefields. Regardless, there really is no debate whether it is possible for the Finite to occupy the third or even later positions in declarative clauses.

Höhle (2019b) maintains that conjunctions are also part of multiple Forefields and does not locate them in the left outfield. His first argument is that conjunctions like *und* (*and*) and conjunctive Adjuncts like *obendrein* (*additionally*) have the same clause linking function and should therefore also have the same potential of occupying the same field if positioned before the finite verb (Höhle 2019b: 79). However, he does not explain what causes the difference in word order between conjunctions and conjunctive Adjuncts if their functions are the same. His second argument is based on word order, where he shows that conjunctions like *aber* (*but*) can also be moved to the Midfield (see example (15); conjunction in bold), which proves that they are not just peripheral elements (Höhle 2019b: 80). This is a strong argument but as far as I can tell, *aber* is the only conjunction where this movement to the Midfield is possible, which rather suggests that *aber* is a special case, which should not be used as a representative for conjunctions as a whole.

(15)

- a. *Karl füttert den Hund **aber** Maria füttert die Katze.*
'Karl feeds the dog **but** Maria feeds the cat.'
- b. *Karl will den Hund füttern, Maria will ihn **aber** auch füttern.*
'Karl wants.to the dog feed, Maria wants.to him **but** also feed.'

(taken from Höhle 2019b: 80)

Engel (1970: 77) even argues that Forefields which include appositions and relative clauses are multiple. He justifies this claim the same way as for all other instances of multiple Forefields, which is that appositions and relative clauses can be moved individually to a different position in the clause. In example (16a), the relative clause follows the head in Forefield position, while in (16b), the relative clause is moved to the Postfield (relative clauses in bold). While it is true that post-modifications like these can in fact be moved in German, I still find this argument unconvincing. It is quite clear that they belong to the

nominal group in the Forefield both semantically and syntactically, even though they can be separated from their heads in some cases.

(16)

- a. *Der Mann, **der Birnen verkauft**, ist wiedergekommen.*
*'the man, **who pears sells**, has come.back.'*
- b. *Der Mann ist wiedergekommen, **der Birnen verkauft**.*
*'the man has come.back, **who pears sells**.'*

In conclusion, the Forefield is a very versatile topological field in German as it can include a variety of different clause elements. Subjects are the most common clause element in the Forefield, but Adjuncts, especially, can be positioned relatively freely both in pre- and postverbal position (Dürscheid 1989: 39). Consequently, if there is any difference in markedness between the different clause elements in first position, this difference in markedness is very small. Similarly, Forefields represent given information in the majority of cases, but new information may just as easily be moved to the Forefield position if the context motivates it. Therefore, German speakers are not heavily restricted in their choice of Forefield units. What is restricted is the number of elements in the Forefield, which is typically only one, especially if one distinguishes between Forefield and left out-field. While it is factually incorrect to assume that the Forefield can only consist of one element, Forefields with multiple elements from the transitivity system are very rare.

3.4 Midfield

The Midfield is the typological field in German that is located in between the finite verb and the rest of the verb phrase. The Midfield is optional in finite-second constructions like declaratives. However, the Midfield is usually realized and is not restricted in the number of elements it can contain, unlike the Forefield and the Postfield. If an element is not used in the Forefield, it is moved to the Midfield, which is why Subject and Finite invert if the Subject does not conflate with the Forefield.

The constituent order inside the Midfield is a highly debated issue and no consensus has been reached regarding which factors contribute to the word order to what extent

(Eisenberg 1994: 417). In general, a variety of aspects can influence the order in the Midfield, including givenness/identifiability/definiteness¹⁷, case, syntactic function, part-of-speech, and others (see for example Lenerz 1977; Hoberg 1981; König and Gast 2009). Definiteness, case, and part-of-speech seem to be the main contributing factors to determine the unmarked word order in the Midfield. In this context, Lötscher (1983: 187) distinguishes between hard and soft sequencing rules, where soft rules can only make a distinction in word order if an element's position has not yet been determined by one of the hard rules.¹⁸ Generally, there are three recognized rules that govern word order in German Midfield: noun phrases follow pronouns, indefinite noun phrases follow definite noun phrases, and dative and accusative case follow the nominative case (König and Gast 2009: 167-168).¹⁹

Hence, in the case where there are multiple pronouns, the Subject, which is in nominative case, precedes the accusative Object and the dative Object in this order (see example (17); Subject in bold, Objects underlined). The sequence of dative and accusative Objects is not strictly fixed so that they can usually be used in any order, especially if they are both pronouns (Altmann 1981: 5).

(17) *Außerdem habe **ich** ihn ihr gegeben.*
 'moreover have **I** it[acc] her[dat] given.'

Müller (1999: 11) argues that the order between Subject and Object pronouns is fixed, meaning that nominative pronouns will always be the very first element in the Midfield. Surprisingly, he uses the pronoun *irgendetwas* (*anything*) as an example, where it is clearly not true that Object pronouns cannot precede it (see example (18); Subject in bold, Object underlined). However, at least for nominative personal pronouns, it is true that they cannot occupy any other Midfield position but the first if they are unaccented (Engel 2004: 170).

¹⁷ These three concepts are not entirely synonymous, but they are closely related. If a nominal group is definite, meaning that it is introduced by a definite article or demonstrative determiner, its referent must be given or at least identifiable from the context (Adam and Delettres 2016: 120). For the purposes of this thesis, the concepts identifiability, definiteness and givenness will be used interchangeably. See Kunz (2010) for a very detailed discussion of identifiability and definiteness in German and English.

¹⁸ Lötscher actually considers thematicity, which in this case is comparable to givenness, a soft rule which is dominated by hard rules such as grammatical constraints.

¹⁹ Uszkoreit (1987: 24) proposes a fourth rule arguing that long elements follow short elements, which may be a side effect of the other three rules.

- (18) *Außerdem hat ihm **irgendetwas** nicht gepasst.*
'moreover has him **something** not suited.'

A similar sequence of Subject and Objects holds true for noun phrases, only that the dative Object precedes the accusative Object (König and Gast 2009: 168). However, unlike with pronouns, it is easier to change the order so that Objects can also be positioned in front of the Subject, albeit in a marked word order pattern (see example (19); Subject in bold, dative Object underlined).

(19)

- a. *Außerdem hat **der Mann** dem Kind den Brief gegeben.*
'moreover has **the man** the child the letter given.'
- b. *Außerdem hat dem Kind **der Mann** den Brief gegeben.*
'moreover has the child **the man** the letter given.'

If all clause elements have the same syntactic form, their unmarked order is easily identifiable. However, if syntactic forms differ, it is not clear whether, for example, case or part-of-speech is the more dominant contributing factor. For instance, in example (20), the dative and accusative Objects are pronouns, and the nominative Subject is a definite noun phrase. Pronouns precede noun phrases, but the Subject precedes the Objects, which means that these two sequencing rules compete with each other. Haftka (1978: 163) argues that pronouns always precede definite noun phrases,²⁰ regardless of their case, which is clearly not accurate. If there is a dominant word order for such cases, it seems to be easily changeable.

(20)

- a. *Außerdem hat **der Mann** ihn ihr gegeben.*
'moreover has **the man** it[acc] her[dat] given.'
- b. *Außerdem hat ihn **der Mann** ihr gegeben.*
'moreover has it[acc] **the man** her[dat] given.'
- c. *Außerdem hat ihr **der Mann** ihn gegeben.*
'moreover has her[dat] **the man** it[acc] given.'
- d. *Außerdem hat ihn ihr **der Mann** gegeben.*
'moreover has it[acc] her[dat] **the man** given.'

²⁰ "Es ist daraus ersichtlich, daß die Pronomina stets vor den substantivischen determinierten Konstituenten der Prädikatsgruppe erscheinen und diese wiederum vor allen neuen Einheiten." (Haftka 1978: 163)

In general, the order in the Midfield follows the principle of topologischer Verbferne (topological verbal distance), which means that constituents are more distant from the finite verb the more syntactic-semantically related they are to the process and the more newsworthy they are from an information structure standpoint (Götze and Hess-Lüttich 2002: 485). This explains why pronouns and definite noun phrases precede indefinite noun phrases, since their references are either given or identifiable from context, which means they carry little communicative dynamism (see Section 5.1.4 for a discussion of communicative dynamism). This also means that the Subject can occupy a very late position in the Midfield, especially if it is indefinite and thus newsworthy. Etzensperger (1979: 63) even argues that the order in the Midfield is entirely determined by communicative dynamism,²¹ which is, however, clearly not accurate.

Another aspect of word order which is not discussed frequently in the state of the art is animacy. Götze and Hess-Lüttich (2002: 489) contemplate whether the sequencing of case in German is a product of animacy differences with the dative Object typically being animate and the accusative Object more often inanimate. However, as was shown above, the unmarked order of dative and accusative can change depending on syntactic form. Animacy is, in Lötscher's (1983) terms, a soft sequencing rule, which means that it can only make a difference in word order if the other, harder sequencing rules compete with each other. For example, Lenerz (1977: 105-106) observed that the sequence of Object followed by Subject in the Midfield was sometimes acceptable and at other times not. In all cases where the Object was animate and the Subject was inanimate, a switch of Subject and Object positions was judged grammatical. In contrast, if the Object was inanimate and the Subject animate, this order was not appropriate. This would suggest that animate constituents are more likely to precede inanimate ones. Lenerz (1977: 107) eventually argues that animacy is not the decisive factor in these sentences and tries to prove so with an example where an inanimate Object precedes an animate Subject. However, the Object he uses is definite and the Subject indefinite, so that the givenness constraint overrides the animacy constraint (see example (21); Subject in bold, Object underlined). If Subject and Object are both definite (22a) or both indefinite (22b), the results are more questionable.

²¹ "Zwischen Personalform und Paraverb richtet sich die Gliedfolge nur nach dem Mitteilungswert." (Etzensperger 1979: 63) He later relativizes this strong claim and also considers other factors that play a role in constituent order.

- (21) *Ich glaube, dass diesen Baum **ein Förster** gefällt hat.*
*'I believe that this tree **a ranger** felled has.'*
(taken from Lenerz 1977: 107)

- (22)
a. *?Ich glaube, dass ein Baum **ein Förster** gefällt hat.*
*'I believe that a tree **a ranger** felled has.'*
b. *?Ich glaube, dass diesen Baum **dieser Förster** gefällt hat.*
*'I believe that this tree **this ranger** felled has.'*

Circumstantial Adjuncts can generally appear anywhere in the Midfield but are more likely to follow Subjects and Objects even if the Adjuncts constitute given information (see example (23); Midfield Adjunct in bold). There is also an unmarked sequence order between different kinds of circumstantial Adjuncts, which will not be discussed in detail at this point (see Götze and Hess-Lüttich 2002 for an elaborate discussion).

- (23) *Auf dieser Tafel erscheint ein zusätzliches Zeitfenster **in der Titelzeile**.*
*'on this table appears an additional time.frame **in the title.row**.'*
[G2E_INSTR_003]

While the unmarked Midfield order is noticeably influenced by identifiability, case, and part-of-speech, the exact word order is always influenced by multiple factors. With the exception of nominative personal pronouns, which can only be either in the Forefield or first Midfield position, most other constituents can be moved around in the Midfield to some extent if it serves a cohesive or communicative purpose. As was shown above, given or identifiable information is more likely to precede new information but can also follow if other constraints weigh more heavily. Such word order patterns are more marked, as evidenced by the associated change in intonation, but for the most part acceptable in German. This will be particularly relevant when discussing Theme in German.

3.5 Animacy, agency and sentience

In Section 3.1, word order flexibility was argued to be one of the key syntactical differences between English and German. One of the ways in which English can make up for this lack of syntactic flexibility is by allowing a variety of semantic mappings onto clause elements. English Subjects, for example, can express a variety of different meanings,

sometimes regardless of the semantic requirements of the verb. One such construction involves the combination of an inanimate Subject and a Verb that requires agency or sentience.

The questions of what constitutes an action, and an agent are of central importance not only in the field of linguistics but also in fields like philosophy and sociology. Ahearn (2001: 112) tentatively proposes the following definition: "Agency refers to the socioculturally mediated capacity to act." This capacity to act has often been related to different similarly abstract concepts such as free will, volition, and sentience, but to my knowledge, there is no generally agreed upon definition which unifies all of these related concepts.

In the state of the art, different criteria have been put forward to determine whether an entity qualifies as an agent, namely volition, causation, sentience, and independent movement (Dowty 1991), but also other factors such as dynamic self-sufficiency and affective potency (Cruse 1973), control (DeLancey 1984), and others. Many of these criteria are interrelated, such as animacy and sentience, volition and sentience, and causation and control. An entity undoubtedly represents an agent if it unifies all of these criteria, that is, if it is a living, sentient being that willfully caused an action through independent movement. However, the state of the art has identified different cases where some but not all of these conditions are fulfilled, which makes the binary distinction between agent and non-agent problematic.

In this context, there are two central questions regarding agency which are discussed frequently in the current relevant literature: Can non-sentient entities be agents, and can sentient beings ever be non-agents even if they cause an event? Closely related to these questions is the aspect of volition or intentionality. If an entity intentionally carries out an action, consciously or subconsciously, it can be considered the agent of that event (Davidson 2001: 44). This distinguishes most actions carried out by people from events involving inanimate entities. Some actions such as lying (as opposed to not telling the truth) can only be carried out intentionally, which means that the participant carrying out the action must necessarily be an agent.

The problem is that beings that are capable of volition do not always carry out actions intentionally. For example, if a person accidentally knocks over a table, it is not an intended action, which suggests that the participant is not an agent. However, Davidson questions whether that person then has the same status of non-agentivity as a stone that knocks over a table (by rolling down the hill for example): "[...] for although intention implies

agency, the converse does not hold" (Davidson 2001: 45). He concludes that any event is an action, and every person is an agent if that event could be done intentionally, irrespective of whether this is the case in a given situation (Davidson 2001: 46). If this is the decisive distinction between agency and non-agency, however, volition as a criterion would be unnecessary since only sentient beings can have volition. Sentience could then be used as the criterion instead.

The relationship between inanimate objects and cause is another, related issue in the discussion about agency. If an event is initiated by a sentient being, that person is the agent of the event even if they used an instrument. Although it may have been the hammer that caused the nail to penetrate the wood, it was the person using that hammer who is ultimately responsible for the event (Davidson 2001: 53). This relationship between person and instrument becomes less clear when considering machines. If a person uses their computer to carry out a calculation, are they really the agent of that calculation even if they did not calculate anything and possibly do not even understand how such a calculation is carried out? Or is the programmer that 'taught' the computer how to calculate the agent of that action even though they were not even present during the event? Natural forces represent another difficult issue with regard to the question of agency. While obviously non-sentient and incapable of volition, natural forces clearly cause change in the world without being used as an instrument by sentient beings (DeLancey 1984: 181).

The main problem in defining criteria for agency is that the different criteria mentioned above do not necessarily entail one another and can apply or be lacking independently of each other. One solution to this problem is to not treat agency as a binary distinction between agents and non-agents but as a scale with decreasing levels of agency depending on how many criteria are met. Different types of scales or hierarchies have been proposed, for example Silverstein's (1976) animacy hierarchy, Hopper and Thompson's (1980) index of agency, and the animacy rankings by Garretson (2004) and Zaenen et al. (2004).

A sentient being is capable of being an agent of an action, but not every process constitutes an action. The fact that different kinds of processes require different kinds of participants is of central importance in Systemic Functional Linguistics but is also recognized in other frameworks as well (for example Fillmore 1968; Dowty 1991; Croft 1993). Relational processes (see Section 4.3) for example do not represent intended actions and have

fewer restrictions on the nature of their participants and can therefore occur with agentive and non-agentive participants alike. For this reason, Gruber (1967) uses the term agentive verbs rather than agentive Subjects to talk about processes that require an agent Subject. This combination of an agent Subject and a verb that requires an agentive first participant²² will from now on be referred to as an agentive construction as opposed to the combination of an agentive verb and an inanimate, non-agentive Subject, known as a non-agentive construction (Lasch 2016).

In more recent studies on agency, the criterion of animacy was abandoned in favor of sentience (for example Ackerman and Moore 2001; De Swart 2014; and García García, Primus, and Himmelmann 2018), since sentience entails not only animacy but also other agency related criteria. Following SFL terminology, most material processes represent intended actions and therefore require a sentient agent. Sentience as the sole requirement for an agentive construction is not sufficient because other processes such as mental ones also require sentience without describing intended action.²³ Therefore, other criteria, such as volition, have to be considered as well to distinguish between sentient verbs in general and agentive verbs in particular (Serbina 2015: 125).

The reason why agentivity and sentience are of interest for a study on Theme and word order shifts in English-German translation is that German and English show contrastive differences concerning acceptability and frequency of such non-agentive constructions. Agentive constructions are common in both languages, and so are constructions including agentive or non-agentive Subjects and processes which do not require agency, such as relational processes (see example (24)).

- (24) *The exact origin of the state name is unknown.*
Die genaue Herkunft des Namens ist unbekannt.
'the exact origin of.the name is unknown.'
[E2G_TOU_003]

While non-agentive constructions are generally not ungrammatical in either language, they are less marked in English than in German. In English, the semantic mapping onto the Subject is generally more diverse, allowing English to express a variety of Subject-

²² Participant roles will be discussed further in Section 4.3. First participant refers to the participant role of each of the six process types that conflates with the Subject in active, declarative clauses.

²³ While the Sayer in mental processes and the Behaver in behavioral processes are conscious beings (Halliday and Matthiessen 2014: 301), the sentience of the Actor in material processes and the sentience of the Sayer in verbal processes depend on the semantic meaning of the verb.

Verb combinations. Such non-agentive constructions exist in German as well, but they are generally less frequent and more stylistically questionable as judged by native speakers (Hawkins 1986: 57-58). Also, some Subject-Verb combinations, which are acceptable in English, are outright ungrammatical in German (see example (25)).

- (25) *A poll asked people which party they supported.*
**Eine Meinungssumfrage hat Menschen gefragt, welche Partei sie unterstützen.*
'a poll has people asked which party they supported.'
(taken from Königs 2011: 117)

Even if a German inanimate, non-agentive Subject is paired with an agentive verb, it is perceived as clearly metaphorical or metonymical, whereas in English, this metaphoricity is not as noticeable (Teich 2003: 18). The fact that English and German are contrastively different regarding non-agentive constructions is generally accepted in the state of the art, and yet both König and Gast (2009) and Königs (2011) claim that this tendency for anthropomorphism is steadily increasing in German, which they attribute to English influence on the German language.

The reason for this difference can be attributed to the morpho-syntactic differences between English and German, already discussed in Section 3.1, namely case-marking and word order freedom. German has a relatively elaborate case system, which allows the hearer to identify syntactic function based on the inflectional form of the constituent, while English has to rely on a stricter word order to express grammatical relations. The flexibility of semantic mapping in English thus compensates for the inflexibility of the word order (Hawkins 1986: 67).

Non-agentive constructions are of particular interest in translations from English to German. Given that non-agentive constructions are not strictly ungrammatical in German, the translator, when faced with a non-agentive construction in the English original, has to make a choice between keeping the original clause structure intact or changing it to make the translation more authentic in the target language. If they decide against the stylistically marked literal translation, there are different translation procedures available to resolve the non-agentive construction. In many cases, a non-agentive Subject can be replaced by a prepositional phrase in German (Königs 2011: 120). The advantage of such a translation is that the sequence of lexical meaning stays largely intact and only the grammatical functions are changed (Serbina 2015: 114-115). If possible, the translator

can also change the semantic content of the verb to turn an agentive verb into a non-agentive verb (Königs 2011: 120). In this case, the sequence of semantic and grammatical units can largely be preserved, but not every non-agentive construction lends itself to such a change. Lastly, the Subject can also be replaced by either an agentive Subject or a semantically empty Subject such as *man (one)* or *es (it)*.

This contrastive difference of non-agentive constructions is not only relevant for the language pair English-German but for other languages as well. Other Germanic languages such as Norwegian (Lødrup 1993) and Dutch (Doms, Clerck, and Vandepitte 2016) have very comparable semantic restrictions on clause elements given their similar clause structure and word order flexibility. Other languages with a similar degree of markedness of non-agentive constructions include Hare, a Canadian indigenous language, and Newari, a language spoken in Nepal (DeLancey 1984: 203).

Kast (2012), Serbina (2015) and Freiwald (2016) all worked on variations of Subject translations between English and German with some focus on non-agentive constructions. All these studies worked with the same translation corpus CroCo (Hansen-Schirra, Neumann, and Steiner 2012), but used different queries and annotation guidelines. Kast (2012) searched for all cases in which the semantic content of the original Subject was mapped onto a different grammatical function in the translation. In translations into German, original Subjects were often mapped onto different clause elements if they had less prototypical semantic roles like Temporals, for example.

Serbina (2015) searched for specific verbs in the corpus and annotated the Subjects in terms of their animacy. She focuses on sets of verbs that described volitional and non-volitional processes. Not only did she find that the combination of $S_{inanimate}$ and $V_{volitional}$ were more common in English, but also that 43% of all such constructions underwent a shift in German translations. The majority of these constructions in the popular scientific register involve a research noun like *studies* and the verb *show*.

Freiwald (2016) analyzed non-agentive constructions in popular scientific texts by relying on a manual process type analysis as well as manual annotations of each verb in terms of whether they require a sentient Subject. Like Serbina (2015), he annotated Subject animacy as well. In translations from English to German, Freiwald (2016: 92-93) found that roughly half of all non-agentive constructions are changed and that all of the above-mentioned translation procedures are utilized. The two most common translation procedures involve changing the Subject to an Adjunct (see example (26); Subject in bold,

Adjunct underlined) and rephrasing the verb to eliminate its sentience requirements (see example (27)). Johansson (2004: 273) found that the German *man* (*one*) is used much more frequently in both original and translated texts, which he attributed, among other things, to the restrictions of semantic mappings in German.

(26)

EO: **Cookies** can also store personally identifiable information.

GT: In Cookies können ausserdem **persönliche Identifikationsdaten** gespeichert sein.
'in cookies can moreover **personal identification.data** saved be.'

[E2G_INSTR_009]

(27)

EO: Chilham Castle opens its magnificent gardens to the public [...].

GT: Die prächtigen Gärten von Chilham Castle sind der Öffentlichkeit zugänglich [...].
'the magnificent gardens of Chilham Castle are to.the public open [...].'

[E2G_TOU_005]

Doms, Clerck, and Vandepitte (2016) studied non-agentive constructions in translations between English and Dutch. They used a similar approach to Serbina (2015) where they searched for the verbs *give* and *show* and divided the co-occurring Subjects into human and nonhuman. Doms, Clerck, and Vandepitte (2016) report that 43% of all constructions involving the verbs *give* or *show* and an inhuman agentive Subject are changed in the Dutch translations, which incidentally is also the exact same frequency that Serbina (2015) reports for German translations. The translation procedures for Dutch are also comparable to those in German.

Heilmann et al. (2020) studied the effects of non-agentive constructions in English-German translations, focusing on the translation process rather than the product. Following the results of Serbina (2015) and Freiwald (2016), they designed popular-scientific texts that included either an animate or an inanimate Subject in combination with a verb like *show* or *indicate* and invited professional translators to their translation laboratory to translate the texts into German. They recorded the behavioral measures during the translations with the help of eye-tracking and keystroke logging. Heilmann et al. (2020) found that constructions containing an inanimate Subject are changed more often in accordance with the corpus findings. However, the translation procedures do not mirror

those from the corpus data and the behavioral measures do not differ significantly between the translation of animate and inanimate Subject. They attribute these unexpected results to automatization.

Up until now, the state of the art has mostly focused on processes that are actions and their combinability with animate and inanimate Subjects. However, agentive verbs are not the only kind of process that make demands on the Subject which inanimate entities cannot satisfy. As was pointed out above, an alternative analysis to agency is sentience, where a similar distinction can be made between sentient and non-sentient constructions. A non-sentient construction is a construction which combines a verb that requires a sentient first participant with an inanimate, non-sentient Subject. Given that sentience is a central criterion for agency, non-sentient constructions would include all non-agentive constructions plus those processes that require a sentient first participant but are not actions, which in SFL terminology would be all mental processes and some behavioral processes (see Section 4.3).²⁴

This distinction between non-agentive and non-sentient constructions is arguably irrelevant, at least from a contrastive perspective. There is little reason to assume that the restrictions on the semantic mappings of the Subjects in German are only restricted to agentive processes. The combination of an inanimate Subject and a mental verb should be just as if not more stylistically questionable. Conversely, English may show the same semantic flexibility of its Subject mappings in mentally oriented processes.

²⁴ To avoid confusion, I will only refer to non-agentive verbs and non-agentive constructions if I refer to agentivity specifically. In all other cases, the non-sentient constructions are used as an umbrella term to refer to any case where a verb that requires either agency or sentience is paired with an inanimate Subject.

4 Systemic Functional Linguistics

In this chapter, some of the central systemic functional concepts and terminologies are introduced. Similar to the chapter on translation studies, this fourth chapter does not aim to thoroughly describe the theory of Systemic Functional Linguistics but only serves as an introduction to the concepts that are relevant for the purposes of this thesis. One of the central assumptions in SFL is that a clause expresses multiple strands of meaning simultaneously with the main strands being referred to as metafunctions. These metafunctions and their core systems TRANSITIVITY, MOOD, THEME, and INFORMATION are discussed in detail here. For a more comprehensive overview of theoretical and descriptive concepts in SFL, see Matthiessen, Teruya, and Lam (2010). For a detailed account of the theoretical underpinnings of SFL, see Halliday and Matthiessen (2014), but also Halliday (1967a/b, 1968, 1970, 1973, 1978).

4.1 General concepts of SFL

One of the central assumptions in SFL is that language is made up of systems. A system is a set of options to choose from, where one choice excludes all other choices that are on the same level in the system. Systems are thus paradigmatic in nature as they represent a list of alternative meanings that are mutually exclusive. Halliday and Matthiessen (2014: 22) use the example of POLARITY, which sets up a system of two paradigmatic choices of positive and negative polarity. The choice of one of these options in the system excludes the choice of its alternative.

Every system has an entry condition, which means that certain conditions have to be met first in order for this system to be available to the speaker. If these entry conditions are not met, the system can also not be entered (Sampson 1980: 227). For example, one of the entry conditions of MOOD is that the unit in question is a free rather than a bound clause (Halliday and Matthiessen 2014: 170).

Related to this is the concept of rank. In SFL, units are considered to be on different ranks, for example the ranks of words, groups and phrases, clauses, and clause complexes. The relationship between different ranks is hierarchical as clause complexes are realized

as clauses, clauses are realized as groups and phrases, and groups and phrases are realized as words (Neumann 2003: 52-53). A system always has a point of origin at a certain rank (Halliday and Matthiessen 2014: 49), so that, similarly to entry condition, a system can only be entered if a unit is of the appropriate rank. The system of POLARITY, for instance, is only available at the rank of the clause (Halliday and Matthiessen 2014: 23). An element is considered rank-shifted if it functions as a constituent of a rank that is of equivalent or lower rank. That is to say, rank-shifted units can enter systems that have their point of origin on a lower rank than their own. Embedded clauses that act as qualifiers inside of nominal groups are an example of rank shift (Halliday and Matthiessen 2014: 382).

A system usually consists of multiple levels. That means if a system is entered, a speaker first has to choose between paradigmatic options on the first level of the system, which then opens up an array of new options to select from. In POLARITY, after the speaker has made their first decision between positive and negative, they have to continue making choices to ultimately arrive at a polar clause. The further the speaker progresses in the system, the more delicate the system becomes (Halliday and Matthiessen 2014: 23).

Central to the idea of system networks and paradigmatic relationships is choice. The occurrence of a certain feature or language use is only meaningful if this is the result of a choice on the part of the speaker. Choice does not necessarily entail conscious or intentional decisions, as most language choices are made in a split second on a subconscious level (Matthiessen and Bateman 1991). Choice also does not entail that each of the options available in a system are equally or similarly likely. Some choices in certain system networks are so probable that there hardly seems to be any decision-making involved at all. Furthermore, the freedom of your decisions may be heavily restricted by choices made in other systems or by the rules of the language system in general (Berry 1975: 54). And yet, only through choice does language use become meaningful.

Consequently, systems are inherently probabilistic (Halliday 1991: 31). One level in a system may feature multiple different options, but not all of these options are equally likely to be chosen by speakers. For example, the vast majority of clauses have positive polarity, so that POLARITY can be considered a skewed system (Halliday 1991), meaning that the probabilities between choices inside the system are imbalanced.

When we use language, we do not only enter a single system but a large network of systems simultaneously. Choices made in one system may then impact entry conditions

or probabilities of another system. The sum of all of these systems represents the underlying meaning-making potential of a language (Halliday and Matthiessen 2014: 27). By making choices in the system network, this potential of meaning is then gradually instantiated, and once the speaker arrives at the most delicate choices in each system they entered, they produce text. System and text are two poles of the cline between potential and instance (Matthiessen, Teruya, and Lam 2010: 121).

4.2 Metafunctions

Another central assumption in SFL is that any kind of language use carries not just one single strand of meaning but rather that several meanings are always expressed simultaneously. While there are numerous types of meaning in language, the three main strands are called metafunctions (Fawcett 2008: 45-46). Halliday and Matthiessen (2014) distinguish between three metafunctions, namely the ideational, interpersonal, and textual metafunctions,²⁵ with the ideational metafunction being further divided into the experiential and the logical metafunction.²⁶

The experiential metafunction describes the function of language as a resource to construe human experience. Among other things, language is used to make sense of the world and share experience with others. Such experiences can refer to real-world actions and happenings, but also to states and relationships as well as cognitive and emotional events. When experience is shared with other people, language use inevitably also becomes a social event. We carry out social relationships through language, which is described by the interpersonal metafunction. Any interaction between speaker and hearer is determined by certain social aspects which are expressed through language. The textual metafunction highlights the fact that language use is always embedded in a textual context and that the construal of experience and the enactment of social relationships has to be

²⁵ While English and other languages clearly consist of more than just these three strands of meaning, the majority of the SFL community agrees on these three major metafunctions. For an opposing view, see Fawcett (1980), who argues for additional metafunctions, for example negativity and attitude.

²⁶ The logical metafunction describes the logico-semantic relationships that are set up between construed events, typically through clauses inside of clause complexes. While clause complexes (or to be more precise, T-units) are used as the unit of analysis in this thesis (see Section 6.3), the logico-semantic relationships between single clauses are not analyzed. For this reason, the logical metafunction will not be discussed further. For a detailed account of the logical metafunction, as well as all other metafunctions, see Halliday and Matthiessen (2014).

organized in such a way that it structures the discourse and creates a cohesive text (Halliday and Matthiessen 2014: 30-31).

Metafunctional meaning is not restricted to any rank in particular; however, in English, there are three systems at the clause rank which directly correspond to the three main strands of meaning: TRANSITIVITY, MOOD, and THEME. The clause is thus a multifunctional unit, which serves these three major functions at the same time. These different strands of meanings are ultimately realized as a "single 'rope' of words" (Fawcett 2008: 46) on the level of form. A complete metafunctional analysis of the clause thus always includes a separate annotation for each of the three systems.

In the following sections, each system will be discussed, including an introduction of all of the terms relevant for the purposes of this thesis and a comment on the interrelations between these systems, with particular attention to the textual metafunction. Although these metafunctions are not tied to any language in particular, the descriptions of the systems and their relationship to Theme are based on English.

4.3 TRANSITIVITY

The experiential metafunction describes the resources of a language to construe experience. The system at the clause rank which instantiates this construal of experience primarily is TRANSITIVITY. Human experience is a flow of events that merge into one another. Speakers chunk these continuous events into quanta of change, and TRANSITIVITY describes the lexico-grammatical resources to model them as figures (Halliday and Matthiessen 1999). These figures can describe different kinds of events like doings and happenings, feelings, sayings etc.

A figure consists of up to three components: a process, participants involved in the process, and circumstances surrounding the process. At the center of a figure is the process, which, at the clause rank, is typically realized as a verbal group. As speakers, we can construe an almost infinite number of different events using different verbal groups, and yet we can group these different events both semantically and grammatically into a finite number of process types. Each of these process types construes one domain of human experience and comes with a particular set of participants (Halliday and Matthiessen 2014: 213). Such processes do not only cover actions, but also cognitive events and states.

Every process has participants or participant roles which are expected to be involved in the process. For instance, the process of thinking usually involves two participants; an entity that does the thinking as well as the respective thought or idea. Participant roles do not merely describe the number of arguments a process requires but come with their own set of semantic criteria (for example, the first participant in the process of thinking has to be sentient). Apart from very few exceptions, every process has at least one but often more than one participant. Participants map onto Subjects and Complements in MOOD (Halliday and Matthiessen 2014: 311).

Unlike participants, circumstances are not inherent parts of the process and rather provide additional information about the process. Halliday and Matthiessen (2014: 221) describe participants as close to the center of the process, while circumstances are only peripheral. Common kinds of circumstantial information are information about place, time, and manner. Circumstances function as circumstantial Adjuncts in the MOOD system (Halliday and Matthiessen 2014: 311). Halliday and Matthiessen (2014) argue that in total 22 different types of circumstances can be distinguished in English. This list appears, by their own admission, somewhat arbitrary (Halliday and Matthiessen 2014: 312) and could be extended by even more circumstances. While most processes have a set of participants that are specific to them, the different types of circumstances are not tied to any process type in particular.²⁷

The distinction between participants and circumstances and their ties to Mood elements is not always clear-cut. Circumstantial relational processes (see below) have a circumstantial Attribute as one of their participants, which is analyzed as a Complement in MOOD (Halliday and Matthiessen 2014: 294), presumably because it is part of the argument structure of the verb (see example (28); circumstantial Attribute in bold). For no apparent reason, circumstances that are part of the argument structure of other process types, like material processes (see example (29); circumstance in bold), are not considered participants or Complements (Halliday and Matthiessen 2014: 226 – Footnote).

²⁷ There are some circumstances that are more or less likely to occur with specific process types, for example Matter with verbal and relational processes (Halliday and Matthiessen 2014: 310). Nevertheless, no circumstance type is exclusively used for one process.

(28) *San Francisco is **on the coast of California about 400 miles north of Los Angeles.***
[E2G_TOU_011]

(29) *You can place the monitor directly **on top of the computer.***
[E2G_INSTR_002]

For English, six different process types can be distinguished, of which three are considered principal and the other three subsidiary types of processes (Halliday and Matthiessen 2014: 300).²⁸ The first principal process type is the material process. Material processes are processes of doings and happenings. They are defined as construing “a quantum of change in the flow of events as taking place through some input of energy” (Halliday and Matthiessen 2014: 224). The primary participant involved in the material process is the Actor, which is the one that performs the action and brings about the quantum of change. Apart from a few exceptions, every material process includes an Actor participant. Depending on the semantics of the verb, the Actor can be an inanimate entity or an animate conscious being. Nonetheless, consciousness or sentience is not a general requirement for the participant role Actor.²⁹

²⁸ This distinction between principal and subsidiary process types seems rather arbitrary. Halliday and Matthiessen (2014: 300) argue that the principal process types present three distinct kinds of structural configurations and that they are the most common process types in actual language use. Also, the subsidiary process types can be considered in between the principal process types, so that behavioral processes are in between material and mental processes, existential between material and relational processes and verbal between relational and mental processes.

For one, it is not clear in what way the structural configurations of principal processes are distinct from each other since one of the defining grammatical characteristics of both material and relational processes is that they cannot project, and both relational and mental processes prefer simple aspect over the present-in-present. In this sense, verbal processes can also be considered unique in their structural configuration as they can project and can take the present-in-present, which is not common for any of the other process types.

Regarding frequency, Halliday and Matthiessen (2014: 300) admit themselves that it is highly dependent on register. As will be shown in Sections 7.1.2 and 8.1.2, verbal processes are almost as common as mental processes in both English and German.

Lastly, the arrangement of subsidiary process types in between principal process types appears to be arbitrary also. Verbal processes are considered in the middle of mental and relational processes with no explanation for that whatsoever. Grammatically, verbal processes are more like material processes than relational processes given that they can take the present-in-present, they pass the what-do test, and they allow a third participant which, in traditional grammar, would be regarded an indirect Object. Besides, verbs like *show* and *offer*, which, when used in their literal sense, are typically material processes, can also be used metaphorically as verbal processes. To my mind, if process types have to be grouped into principal and subsidiary types of processes, material, mental, relational, and verbal processes should be considered principal and behavioral and existential subsidiary given their low frequency.

²⁹ Halliday and Matthiessen’s (2014: 250) claim that “the distinction between conscious and non-conscious beings simply plays no part [in material processes]” is not convincing. It is accurate that many material processes allow both conscious and non-conscious beings as Actors and even some verbs which would seem to require a sentient Actor can be paired with inanimate Actors in English (known as a non-agentive or non-sentient construction; see Section 3.5). Nevertheless, there are clearly certain kinds of material verbs like

Some material verbs only require the Actor participant (see example (30); Actor in bold), but most also include a second participant, which is usually the Goal (see example (31); Goal in bold). The Goal describes the participant that undergoes the action, that is, the participant that is affected by the quantum of change brought about by the Actor. In active voice, the Actor maps onto the Subject and the Goal onto the Complement. This configuration is inverted in passive voice. Alternatively, the second participant can also be the Scope (see example (32); Scope in bold) which, unlike the Goal, is not affected by the process, but is either part of the process itself or is the domain over which the process takes place (Halliday and Matthiessen 2014: 239).

(30) **Two of the women** smoked.

[E2G_FICTION_003]

(31) The aircraft enthusiasts can visit **superb collections of historic aircraft** [...].

[E2G_TOU_006]

(32) We are making **progress**.

[E2G_SPEECH_003]

Certain material verbs also allow a third participant called the Beneficiary. The Beneficiary is the participant that 'benefits' from the action. It can be further divided into the roles Recipient, to whom something is given or presented, and Client, for whom something is being done.

The second principal process type is the mental process, which covers all events that are part of a person's consciousness. The mental process is a process of sensing and feeling. The participant that senses is fittingly called the Senser. Unlike the Actor in material processes, the Senser is necessarily human or human-like and requires sentience (Halliday and Matthiessen 2014: 249). The second participant in the mental process is the Phenomenon which is for example the feeling that is felt or the idea that is thought (see example (33); Senser in bold, Phenomenon underlined). Mental processes are further divided into 'perceptive', 'cognitive', 'desiderative', and 'emotive' subtypes, which are grammatically distinct subtypes (Halliday and Matthiessen 2014: 256). The division between Senser and Phenomenon is the same for all subtypes of mental processes.

hire or *lend* that most definitely require an animate, conscious being and that, even in English, can only be paired with inanimate Actors in very marked, metaphorical or metonymical constructions.

(33) ***These groups*** *do not want a democratic Iraq.*

[E2G_SPEECH_004]

The final principal type of process is the relational process. Relational processes describe states of being and having and have the function of characterizing or identifying their participants. In the vast majority of cases, relational processes include exactly two participants, the participant that is being characterized or identified and the participant which serves as the characterization or identity. Furthermore, relational processes divide into two modes, the attributive relational process, which construes a relationship of characterization, and the identifying relational process, which construes a relationship of identity. Both the attributive and the identifying mode have three main types of relation that they can express, namely intensive, possessive, and circumstantial relations (Halliday and Matthiessen 2014: 263). The by far most common verb realizing relational processes is *be* in English and *sein* in German.

The two participants in an attributive relational process are Carrier and Attribute. The Attribute serves as a characteristic of the Carrier which describes but does not uniquely identify them. In English, such an attributive relationship is typically realized using an indefinite article or no article at all (see example (34); Carrier in bold, Attribute underlined).

(34) ***These concerns*** *are well founded.*

[E2G_SPEECH_004]

In identifying relational processes, as the name suggests one participant is ascribed an identity, which allows the hearer to uniquely identify the referent. Identifying relational processes are often realized with the help of definite articles or possessive and demonstrative determiners. The two participants in identifying relational processes can be classified in two different ways using two different sets of labels: The participant roles Identified and Identifier describe the relationship between one participant that is identified and the identity that is ascribed to them. These two participants can come in any order and their categorization depends entirely on context and, in spoken language, intonation since the intonation focus is usually placed on new information, which is typically the Identifier.

The other two participant roles in identifying relational processes are Token and Value. The Token represents the entity that is more specific and concrete, whereas the Value is the more general and abstract category of the two (Thompson 2014: 103). Token and Value can also come in any order as examples (35) and (36) show (Token in bold, Value underlined). These roles are not alternative labels for Identifier and Identified but represent an additional layer of analysis. A complete annotation of identifying relational processes therefore requires the assignment of both of these participant roles. Any combination between Identifier/Identified and Token/Value in any order is possible.

(35) **Real ale** is also the pride of the region served with a smile in traditional pubs.
[E2G_TOU_010]

(36) A highlight for sport fans is **the Volkswagen Cup International Ladies Tennis Championship in June which attract the world's top players.**
[E2G_TOU_005]

The first subsidiary process type is the verbal process, which describes events of saying or, more accurately, of the transfer of meaning. The participant that communicates that meaning is called the Sayer and the meaning that is being communicated is the Verbiage. Verbal processes can include a third participant, the Receiver, to whom the Verbiage is communicated. The labels of the participant roles suggest that verbal processes only deal with oral discourse, but they also cover written discourse as well as more metaphorical or symbolic exchanges of meaning (Halliday and Matthiessen 2014: 303).

Behavioral processes characterize events where one participant, the Behaver, has an inner experience, which manifests in an outer behavior. Behavioral processes are similar to mental and material processes, both semantically and grammatically: Akin to material processes, they are more likely in present-in-present and they cannot project other clauses but, like mental processes, they require a sentient first participant. Behavioral processes are usually intransitive and cover, among other things, physiological processes like laughing, nodding, or yawning. It is possible to include a second participant, the Behaviour, which, like the Scope in material processes, is an extension of the process (see example (37); Behaver in bold; Behaviour underlined).³⁰

³⁰ The behavioral process is, by Halliday and Matthiessen's (2014: 301) own admission, the least well defined and distinct process type since it is grammatically so similar to material processes. The question of what qualifies as an inner experience is particularly problematic. Most material processes acted out by a sentient Actor involve an intention, and therefore require some sort of prior conscious or subconscious

(37) *I pulled a face [...].*
[E2G_FICTION_008]

The final process type is the existential process, which establishes a state of existence. Grammatically it is very similar to relational processes, with the exception that the existential process only has one participant, the Existent, and the Subject is filled with a semantically empty *there* in the vast majority of cases.³¹

Choices made in TRANSITIVITY have an effect on choices for MOOD but even more so THEME. The first participant of each process type is the most likely candidate for the Theme in declaratives, given that they usually conflate with the Subject in MOOD as well. The Subject in Theme position is considered unmarked in declaratives (see Section 5.6). Material, mental, and verbal processes can be passivized, which also allows the second participant to be thematic. In identifying relational processes, Identifier/Identified and Token/Value can occur in any order so that the Theme-participant conflation may be less strict. Attributive relational processes cannot be passivized, which means that the Attribute can only be used as Theme if the Complement is moved to the beginning of the clause. In this case, unlike most other instances of Complement Themes, the Subject does not immediately follow the Complement, but Subject and verbal group are inverted (see example (38); Subject in bold, Complement underlined). This reverse of Carrier and Attribute is quite rare (2.8%, see Section 8.1.2).

(38) On her face was **a look of sour scepticism**.
[E2G_FICTION_009]

The likelihood of a circumstance Theme is heavily dependent on whether the speaker chooses to include a circumstance in the TRANSITIVITY system. Since only 54% of clauses include a circumstance, 46% of times this Theme option will not be available (see Section

mental process. Physiological reactions like laughing or crying, which are behavioral processes, are typically not intended actions and require a mental trigger more so than verbs like *drive* or *cook*, which are material. In how far this also holds true for a verb like *dance*, which is one of the examples of a behavioral process (Halliday and Matthiessen 2014: 302), is questionable.

³¹ Beside the participant roles mentioned here, most process types also have some additional participant roles that are only used in very specific subtypes. These include Attribute and Initiator for material processes, Inducer for mental processes, Assigner and Attributor for relational processes, and Target for verbal processes. See Halliday and Matthiessen (2014), for more detail.

8.1.1). These frequencies decrease tremendously if one considers specific types of circumstance Themes, like a Guise or a Viewpoint circumstance, which have a very low frequency given that they are selected rarely in TRANSITIVITY. The choice of process type can also influence the likelihood of a non-Subject Theme in declarative clauses. For one, not all process types lend themselves equally well to have a Complement be used as the Theme of the clause. As will be shown in Section 8.1.2, the process type that is most likely to feature a fronted Complement is the attributive relational process, whereas Complement Themes in existential processes are entirely ungrammatical. Also, circumstances are used in some processes more often than in others. For example, a very common construction in the register of tourism leaflets involves existential processes in combination with circumstances of Place (see Section 8.2). Thus, depending on process type, circumstance Themes can become more or less probable.

4.4 MOOD

At the clausal rank, the system that is primarily responsible for expressing interpersonal meaning is MOOD. Through MOOD, the speaker can assign speech roles to themselves and to their audience and these speech roles determine the nature of the interactive exchange between speaker and hearer. There are only two basic types of speech roles, giving and demanding, and the speaker decides for each free clause which of these speech roles they want to assume for themselves (Halliday and Matthiessen 2014: 135). The two commodities that can be given or demanded are information and goods-and-services. These distinctions result in four primary speech functions: statements, in which the speaker assumes the role of information giver; questions, in which the speaker assumes the role of information demander; offers, where the speaker assumes the role of goods-and-services giver; and commands, where the speaker assumes the role of goods-and-services demander.³²

The main grammatical resource for realizing these speech functions is mood. In English, the Mood element consists of the Subject and the Finite, and their realization and

³² It is quite clear that this is an oversimplification of speech functions since not every exchange can be easily categorized into one of these types of interactions. Especially such interactions that involve emotions like jokes, insults, and complaints cannot be satisfactorily described using Halliday and Matthiessen's (2014) 2x2 matrix in my opinion. Such interpersonal nuances are likely realized by other systems like intonation or appraisal though; hence, for the grammatical resources stemming from MOOD, this basic model suffices.

order determine the mood of the clause, which, to a certain degree, maps onto one of the four speech functions. The first distinction is made between the indicative mood type, which includes the Subject and the Finite, and the imperative mood type, which only includes the Predicator (Halliday and Matthiessen 2014: 162). Indicative distinguishes between two subtypes: declaratives, in which the Subject precedes the Finite, and interrogatives, in which the Subject follows the Finite. Interrogatives can be further divided into *yes/no-* or polarity interrogatives, where the Finite assumes the first position of the Mood element, and *wh-*interrogatives, where the first position is occupied by a *wh-*element which represents the information that is sought.³³

The Subject usually takes the form of a nominal group but can also be a prepositional phrase or adverbial group in circumstantial relational processes (Halliday and Matthiessen 2014: 140). The Subject has modal responsibility, which is evident from the fact that it is the Subject that is used as a pronoun in a tag question. This shows that the Subject, together with the Finite, forms the basis of a proposition that can be either affirmed or denied and is thus responsible for the success or failure of the entire proposition (Halliday 1984). If the Subject is realized as a pronoun, it is in nominative case and it is positioned either before or after the Finite depending on the mood. Also, it agrees with the Finite in terms of person and number (Hasan and Fries 1995b: xxi). The semantic description of the Subject being the doer of the action is a conflation of interpersonal and experiential meaning and is thus not helpful in a systemic functional approach – and also not true for any process that is not material in active voice.³⁴ The Finite is the part of the verbal group that expresses either tense or modality, which can also be fused with the Predicator (Halliday and Matthiessen 2014: 140).

All other constituents besides Subject and Finite (and Predicator in imperatives) are irrelevant for determining mood. The Predicator is the remainder of the verbal group without the Finite. In experiential terms, the Predicator includes the actual process, also known as the lexical verb. Complements are defined as elements that have the potential of being the Subject but are not (Halliday and Matthiessen 2014: 153). This category encompasses what in traditional grammar is separated into Objects and Complements or

³³ The *wh-*interrogative can feature the order of Subject preceding the Finite only if the information that is sought is in fact the Subject of the clause realized as *who* or *what*.

³⁴ This distinction is tied to the divide between psychological, grammatical, and logical Subject. What Halliday and Matthiessen (2014: 80) describe as Subject is what in traditional grammar would be considered the grammatical Subject, while the functions of logical and psychological Subject are mapped onto the experiential and textual metafunction respectively.

Predicatives.³⁵ All further elements that do not have the potential of being the Subject and thus cannot assume modal responsibility are labelled Adjuncts. Adjuncts can be divided into three types, which determines whether they are part of the Mood, the Residue, or outside of mood altogether. Circumstantial Adjuncts map onto circumstances in TRANSITIVITY, are part of the Residue, and experiential in nature (see example (39); circumstantial Adjunct in bold). Modal Adjuncts primarily express interpersonal meaning and can be further divided into mood Adjuncts and comment Adjuncts. Due to the fact that comment Adjuncts represent a personal comment by the speaker, they do not contribute to the proposition and are outside of mood. Mood Adjuncts, on the other hand, can express meanings of modality and intensity and are not only part of mood but even part of the Mood, together with Subject and Finite (Halliday and Matthiessen 2014: 157; see example (40); comment Adjunct in bold). Lastly, the conjunctive Adjunct has primarily textual functions of conjoining clauses textually and is not part of the proposition and mood (see example (41); conjunctive Adjunct in bold).

(39) ***In mid-July***, the President announced new agreements that involve DOE.

[E2G_SPEECH_002]

(40) ***Luckily***, we still have poetry written in proto-Welsh [...].

[E2G_SPEECH_006]

(41) ***Moreover***, I want to reiterate President Bush's and Lord Robertson's pledges not to give Russia a veto over NATO operations.

[E2G_TOU_007]

³⁵ The category of Complement in SFL is very vague. The main defining property is their potential of being the Subject, yet Halliday and Matthiessen (2014: 154) acknowledge themselves that there are Complements that cannot be the Subject, namely Attributes in relational and material processes. Complements are almost always nominal groups which, according to Halliday and Matthiessen (2014), also include what in traditional grammar would be analyzed as adjective phrases. However, they do not explain why this is the case. Moreover, prepositional phrases can also be Complements if they are the Attribute in circumstantial relational processes, even though they, too, cannot be the Subject of the clause. Halliday and Matthiessen (2014) argue that a distinction between Object and Complement is not relevant for MOOD and is rather a distinction between participant roles in TRANSITIVITY, which I agree with. However, if these two clause elements are summarized in one category in MOOD, this category needs to have at least one defining characteristic in the MOOD system. The potential of being the Subject is not a universal feature of Complements and neither is their formal structure. Thus, the category as a whole remains fuzzy. Given that the distinction between Complements and Adjuncts solely rests on their ability or inability to be Subject, Adjuncts are accordingly also ill-defined. These inconsistencies have already been pointed out by Huddleston (1988) and have still not been addressed properly in my opinion.

As an alternative, Fawcett (2008: 138-139) defines Complements as any participant role that is not the Subject, which then shifts the problem away from Complements to the definition of participant roles. While this would make the category of Complements more uniform, it also raises the question why the category is even relevant for MOOD if it is defined solely on experiential grounds.

The choices the speaker makes in the MOOD system have a great impact on which element in the clause is more likely, less likely, or outright impossible to be thematized. For this reason, the mood of a clause determines which kinds of Themes are considered marked and unmarked Themes (see Section 5.6). Hence, while it is almost impossible to thematize the Predictor in declarative clauses (see example (42) as one of these very rare cases), they are by far the most common Theme type in imperative clauses (see example (43); process in bold in both examples).

(42) ***Surrounding*** the town are picturesque villages, ancient monuments and beautiful country parks.

[E2G_TOU_010]

(43) ***See*** them as a band of terrorists.

[E2G_FICTION_007]

The probabilities of the different Theme types mirror the primary purpose of the different speech functions. A demand for goods and services, often realized as an imperative, revolves around what the speaker wants the hearer to do, ergo the Predicator. The Subject typically does not need to be specified as it is obvious from the context. Nevertheless, if the speaker takes on the role of information seeker, it is natural to use the element that embodies the information that is sought as the point of departure, so either the *wh*-element in a *wh*-interrogative or the Finite, which embodies the expression of polarity, in *yes/no*-interrogatives.

Fawcett (2007: 124) argues that these choices of early elements like the finite verbal operator in *yes/no*-interrogatives only represent consideration of the MOOD system and are completely independent from choices in THEME, which is why he does not consider such interpersonal elements thematic. Halliday and Matthiessen (2014), on the other hand, consider these strong relations between THEME and MOOD evidence for the significance of early position in English. Unmarked Themes like the *wh*-Themes in *wh*-interrogatives and Predicator Themes in imperatives have become part of the system of English Theme and are used as the point of departure to signal the mood (Halliday and Matthiessen 2014: 101-102).

I believe Halliday and Matthiessen are right in this respect. I do agree with Fawcett that if the speaker decides to use a *yes/no*-interrogative, there really seems to be no choice left

in the *THEME* system whether or not to thematize the finite verbal operator. And one important assumption in a systemic approach to language is that meaning is created through choices (Halliday and Matthiessen 2014: 23-24). However, it is clearly not a coincidence that in *yes/no*-interrogatives, the first position, which is the Theme position, is filled with the element that expresses polarity. Therefore, the *THEME* system seems to have developed in such a way that some choices have gained a probability of almost 100%, just because they so perfectly capture the meaning that the speaker wants to express.

4.5 *THEME* and *INFORMATION*

In this section, I will present the textual metafunction and its associated systems *THEME* and *INFORMATION* at clause rank. Given that Chapter 5 comprises a detailed discussion of the functional meaning and the formal realization of Theme, this illustration will be brief and is intended to complete the description of the division between the three metafunctions.

The textual metafunction highlights the fact that any language use has the character of a message or quantum of information. With the experiential metafunction, speakers construe their experiences and with the interpersonal metafunction, they enact their social roles. The textual metafunction allows the speaker to organize these two strands of meaning as information that can be communicated and thus enables these meanings to be presented as contextualized text (Matthiessen and Martin 1991: 42). Because of that, the textual metafunction is often also called the enabling metafunction (Halliday 1974: 95).

At clause rank, two systems contribute to this textual development of the discourse: *THEME* and *INFORMATION*.³⁶ *THEME* gives thematic prominence to some part(s) of the clause, which acts as a local context and links the clause to its textual environment. This portion of thematic prominence is called Theme. The remainder of the message is the Rheme, which represents the direction in which the message develops (Martin, Matthiessen, and Painter 1997: 21). In English, the Theme is realized positionally, with the Theme preceding the Rheme in the clause. The exact boundary between Theme and Rheme in English

³⁶ The information unit is strictly speaking not the same as the clausal unit, but often coincides with it. See Halliday and Matthiessen (2014) and Halliday and Greaves (2008) for more detail.

is a matter of debate (see Section 5.4), but there is agreement that units of all three metafunctions can receive thematic prominence. Also, every English Theme must include an experiential element, that is, either a participant, a circumstance, or the process, so that the point of departure of a clause is always rooted in experience (Halliday and Matthiessen 2014: 111-112).

INFORMATION awards prominence to parts of the clause by highlighting them as newsworthy. Similar to the dichotomy Theme-Rheme, the clause can be divided into two parts to express the gradual difference in newsworthiness, namely Given and New. Newsworthiness in English is typically established through intonation in spoken discourse and position in written discourse. Thematic and information prominence are realized in an inversely cumulative pattern: In the unmarked case, thematic prominence decreases as the clause progresses, whereas information prominence increases (Halliday and Greaves 2008: 72).

The textual metafunction is sometimes viewed as less significant, given its second-order status as the enabling metafunction. The experiential and interpersonal metafunction first bring the linguistic resources into existence, which are then organized and structured through the textual metafunction (Matthiessen 1992: 54). However, Matthiessen (1995b: 46) rejects the idea that the textual metafunction is simply a linguistic post-production, which only starts to operate once all other experiential and interpersonal decisions have been made. He argues that textual considerations are active when making choices in TRANSITIVITY, so that the use of process type and choice of circumstances fit in well with the textual surrounding. To his mind, the metafunctions are not ordered temporarily but instead operate simultaneously (Matthiessen 1992: 74).

Nonetheless, Fawcett (2007: 42) rightly points out that "the grammar cannot make a decision about whether or not to thematize something until it know[s] that there is something that could be thematized." It thus seems evident that some choices in the different systems have to be made in a chronological order, at least on some level of delicacy. It is not possible to simultaneously decide in favor of a circumstance Theme in the THEME network but decide against the inclusion of a circumstance in the TRANSITIVITY network. This opens up the question which of the systems takes precedence during the decision-making. Personally, I believe that there is no universal answer to this question and that depending on the situation, any of the three metafunctions may be in focus when making choices. Beyond that, the decision-making process is most likely not one straight path but

rather a loop where choices in one system can make the speaker re-evaluate choices made in other systems. Nevertheless, I agree with Fawcett (2007) that generally, a speaker first makes experiential and interpersonal decisions on what is expressed through grammar before arranging them textually.

As was shown in the previous two sections, choices in *THEME* are dependent on choices in *MOOD* and *TRANSITIVITY*. Also, there is a close relationship between the point of departure and given information in English, which is why the two are often conflated (see Section 5.1.1). Thus, the Theme choices are never just made in isolation but always dependent on decisions made in other systems. That being said, most clauses can be ordered in different ways, which shows that the clause segmentation into Theme and Rheme is ultimately a choice. In the next chapter, Theme will be thoroughly discussed regarding its form, function, and relation to other linguistic concepts.

5 Theme

In this chapter, Theme, the central linguistic concept surrounding this research project, is thoroughly discussed. As a first step, Theme in SFL is distinguished from other related linguistic concepts, which in the past have been compared and sometimes confused with the concept of Theme. Once these distinctions are made clear, different approaches to a functional description of Theme in SFL are put forward, including common lines of criticism. The counterpart of Theme, called Rheme, is briefly discussed followed by a formal description of Theme in English. This description includes various English Theme hypotheses proposed by the state of the art and continues with an account of the two central thematic aspects of multiple Themes and Theme markedness. The chapter proceeds with a summary of Theme descriptions in a variety of other languages, which then leads over to a discussion of Theme in German. In this section, the most commonly used German Theme hypothesis, proposed by Steiner and Teich (2004), is introduced and then contrasted with alternative suggestions. Subsequently, Theme differences and similarities in English and German are discussed and research on Theme in translations is presented. The chapter ends with the formulation of the hypotheses.

5.1 Theme and related concepts

5.1.1 Theme and Given and New

The label Theme is used in various linguistic frameworks to describe different aspects of the English language. In particular, the relationship between Theme and given information is commonly discussed. Vilem Mathesius, one of the founders of the Prague school, was among the first modern linguists to introduce the concept of Theme (or rather *téma*) into his account of clause structure, called functional sentence perspective (FSP). According to Mathesius (1939), the Theme is "that which is known or at least obvious in the given situation, and from which the speaker proceeds in his discourse" (translated in Firbas 1964b: 268). With this definition of Theme, Mathesius essentially combined two aspects in one concept, namely known or given information and a start-off point from which the discourse develops, which is why Fries (1995a: 1) dubs Mathesius' position on Theme the

"combining approach". The rest of the clause is called Rheme and represents new information (Vasconcellos 1985: 23).

This definition of Theme rests on two assumptions about the relationship between information and clause structure: Every clause contains information that is known, and known information is tied to a particular position in the clause. However, neither assumption is necessarily correct. Other members of the Prague school like Trávníček (1962, as cited in Vasconcellos 1985, 2008) and Daneš (1964b, 1970, 1974) have demonstrated that known information is not necessarily tied to early position and argue to distinguish between start-off point and given information. Firbas (1964b) also shows that clauses do not have to contain known information at all. Instead, he defines Theme as the element with the lowest degree of communicative dynamism (Firbas 1964b: 272; see Section 5.1.4 for a discussion of communicative dynamism).³⁷ In his analysis of sentence structure and initial position in German, Beneš (1971: 164) uses Theme to refer to known information and further introduces the term Basis, which he describes as the opening element of a sentence that links up with the rest of the text and serves as the starting point from which the rest of the utterance proceeds.

Halliday (1967b) adopts the Prague terminology of Theme and Rheme to refer to the organization of clause constituents. He defines Theme as "the point of departure for the clause as a message" (Halliday 1967b: 212), which is structurally realized as coming early in the clause and which is followed by the Rheme. Halliday acknowledges that Theme and Given are related, since given information is frequently used as the point of departure of the message but makes a distinction between the system of THEME and the system of INFORMATION within the textual metafunction. The system of THEME organizes the clause as message by highlighting one part (Theme) which serves as the starting-off point from which the rest of the message (Rheme) can be interpreted. The system of INFORMATION is a parallel system to THEME, which does not map onto the clause but the information unit and highlights one or several parts of the message as New as opposed to Given (Halliday and Matthiessen 2014: 115). In the unmarked case, a clause represents one information unit, but it is possible for one information unit to extend over more than one clause or for one clause to include more than one information unit (Halliday and Matthiessen 2014: 115).

³⁷ In fact, Mathesius himself realized that known information does not necessarily come early in the clause. According to Daneš (1964b), Mathesius originally intended to distinguish between three different terms: *východiště* (*point of departure*), *téma* (*theme*), and *základ* (*basis, foundation*), without explaining the exact difference. Mathesius later decided against using the term *východiště* and conflated the labels *téma* and *základ* into one concept.

Halliday and Matthiessen (2014: 116) define information from a linguistic point of view as "the tension between what is already known or predictable and what is new or unpredictable". However, givenness and newness are not absolute but rather graded properties (Halliday and Greaves 2008: 102). Chafe (1976: 30) describes given information as any kind of information that the speaker believes to be in the consciousness of the hearer. As such, what is given information is chosen by the speaker, but is essentially listener-oriented (Halliday and Matthiessen 2014: 120). Gundel, Hedberg, and Zacharski (1993) postulate a Givenness hierarchy, in which they distinguish between six different cognitive statuses that are relevant for the speaker to assume that the hearer considers a piece of information as given. The most common way of establishing givenness by the speaker is to refer to something that has been previously mentioned and is thus 'activated' in the consciousness of the hearer. Identifiability is a closely related concept to givenness and is linguistically realized through the use of definite articles or proper names, for instance (Kunz 2010: 118).

In English, there is a close relationship between thematic structure and information structure, which is why they are often conflated. Thematic structure in English is realized sequentially with the Theme preceding the Rheme. While Given and New are not necessarily tied to any particular position in the clause, in the unmarked case, the system of INFORMATION parallels that of THEME with Given coming early in the clause and being followed by New (Fries 1997: 233). It is plausible to use given information as the point of departure of the message because that seems to be a suitable start-off point, from which the rest of the message, which is new information, can be interpreted (Thompson 2007: 672). Notwithstanding, a speaker may also decide to use new information as Theme to establish contrastiveness or put special emphasis on it (see example (44); Theme in bold).

(44) **A relative address** always starts from the directory in which the current document is located.

[E2G_INSTR_005]

If information structure is not signaled through sequence, there needs to be other linguistic resources to mark given and new information. In spoken language, information units are realized through intonation. Each information unit corresponds to one tone unit. A tone unit consists of one tonic and multiple optional pretonics (Halliday and Greaves 2008: 101). The tonic is a strong pitch movement that falls on one syllable inside of the

tone group. The element that receives the tonic is considered the information focus of the information unit. The focus is highlighted as new information and marks the end and the peak of the new information. This means that any element that precedes the tonic syllable can be either given or new, so that its information status remains indeterminate (Halliday and Greaves 2008: 99). However, it also means that any element that follows the focus in the same tone group must be given information (Halliday and Greaves 2008: 103). In the unmarked case, the tonic falls on the last new word of the tone group, which is typically a noun, adjective, adverb or lexical verb (Fawcett 2008: 119). In written language, there is no comparable system that signals information status, so even though there is no one-to-one relationship between clause structure and information structure, the reader has to rely on word order and identifiability markers to distinguish between given and new information.

With the parallel systems of *THEME* and *INFORMATION*, the message thus has two peaks of prominence. The thematic peak is sequential, with the prominence being on the Theme, which takes up an early position in the clause. Therefore, thematic prominence decreases as the message unfolds. The information peak is phonological, with the prominence being on New, which is signaled by pitch movement. In the case of unmarked tonicity, information prominence therefore increases as the message unfolds (Halliday and Matthiessen 2014: 116).

5.1.2 Theme and Topic

In Halliday's early illustrations of Theme, he described the nature of Theme on the basis of two characteristics: Theme is "the point of departure" and "what is being talked about" (1967b: 212). In the first edition of IFG, Halliday (1985) defines Theme similarly as "that with which the clause is concerned" (38) and "what the clause is going to be about" (39). Theme as the point of departure is a functional description still used today. However, defining the Theme as what the clause is about proved to be controversial. The main problem with this definition is that it makes Theme virtually indistinguishable from another related concept called Topic, which is also defined as what a sentence is about (Lambrecht 1994: 118). In fact, linguists like Gundel (1988), Huddleston (1988), and Downing (1991) equated Halliday's early description of Theme with the concept of Topic in other frameworks and thus questioned the usefulness of this alternative label. Halliday did maintain

that Theme and Topic were not the same concepts (1967b: 205; 1985: 221), but his earlier definitions lend themselves to misunderstandings.

To explore the relationship between Theme and Topic in English, a robust definition of Topic needs to be established first. However, this proves to be challenging since Topic, similarly to Theme, is defined differently in distinct frameworks. The terms Topic and Comment were first introduced by Hockett (1958: 301), who described these labels as "the most general characterization of predicative constructions [...] where the speaker announces a topic and then says something about it". Most linguists agree with the general definition of Topic as what something is said about, but there is disagreement on how Topic can be identified in a language like English.

In some languages like Japanese and Tagalog, the topic is morphologically marked. In Japanese, the marker *wa* is attached to a definite or generic phrase about which something is said. The element that carries the *wa* marker is also typically in sentence-initial position (Gundel 1988: 17). In Tagalog, the particle *ang* is placed before a definite noun phrase to give one part of the clause prominence (Schachter and Otañes 1972). In English, there is no such conventional marker, which is why Topic needs to be identified on other grounds (Chafe 1976: 50).

In English, the Topic is frequently positioned sentence-initially and it is often conflated with the Subject. In fact, there is a strong relationship between the Subject-Predicate and Topic-Comment structure in most languages as the Predicate typically says something about the Subject (Lambrecht 1994: 131). Nonetheless, Topics neither have to come first in the clause nor do they have to coincide with the Subject. There are some ways to overtly signal topicality in English, namely using prepositional phrases introduced by *as for* or *regarding* (Huddleston 1991: 100), which shows that constituents other than sentence-initial Subjects can be topicalized (see example (45); Topic as Theme in bold).

(45) ***As for Jeremy Mohonk***, the third principal player in the mortal drama about to unfold, he didn't pay rent, hadn't ever paid rent and never would.

[E2G_FICTION_004]

According to Lambrecht (1994: 121-122), one common characteristic of the Topic is that it has low pitch prominence and does not carry the tonic of the tone unit in unmarked intonation. The Topic is also usually part of the shared informational space of speaker and hearer, while the information focus constitutes information that is not shared (Dik 1980:

212). Topic is often used synonymously with Given (Vasconcellos 1985: 150-151), which is usually accurate in the case of unmarked tonicity. However, Topic cannot always be synonymous with given information because there are sentences that do not contain any given material at all (see example (46)). This would mean that such sentences are not about anything, which Gundel (1988: 35)³⁸ rightly rejects.

(46) *Someone typed a message.*
[E2G_INSTR_004]

There is, however, disagreement about what the Topic is in the case of marked tonicity. Ward and Prince (1991: 170) argue that in sentences like (47), the Topic of the sentence is *Brains*, which is new information and carries the tonic of that tone unit. Gundel (1988: 40), on the other hand, suggests that Topics never carry primary stress. This shows one of the main issues of such topic analyses. Since English is missing an overt marker for topicality, especially in written discourse, the identification of Topic in English is more reliant on the speaker's intuitions what a clause is about and less reliant on linguistic cues.³⁹

(47) *Brains you 're born with. A great body you have to work at.*
(taken from Ward and Prince 1991: 170)

Given Halliday's original description of Theme as what the clause is about, its comparison to Topic seems obvious. Yet, he never equated these two concepts himself. One reason for this is that, at least in English, the Theme can consist of more than one element, called a multiple Theme (see Section 5.5), and that they cannot all represent the Topic of the Theme. Other linguists like Gundel (1988) also noticed that the left-most element cannot generally be considered the Topic, as that would lead to quite counter-intuitive analyses of what a clause is about. Elements like *in addition* and *clearly* are called textual and interpersonal Themes if they are fronted but do not conclude the Theme. The Theme ends with

³⁸ Gundel (1988) makes the claim that these sentences must be about something, so that they must include a topical element, but also explains that non-generic, indefinite noun phrases cannot be Topics as in *A French king married his mother and later he divorced her*. She comes to this conclusion on the basis that the sentence cannot be rephrased to **Concerning a French king, he married his mother*. However, it is not immediately clear what she would then consider the Topic of the above example.

³⁹ See for instance Huddleston criticism of IFG 1, where he discusses the matter of aboutness in sentences like *Nothing will satisfy you* (1988: 158).

the first element that carries experiential meaning, which is either a participant, a circumstance, or a process, which Halliday calls the topical Theme (Halliday and Matthiessen 2014: 111-112).⁴⁰ Over the years, different systemicists have argued to disassociate Theme from what the clause is about and instead focus more on its role as the point of departure of the message (see for instance Fawcett 1980, 1981; Downing 1991). In addition, some linguists that have worked with Theme decided against the label of topical Theme and instead use experiential Theme, since the experiential part of Theme does not necessarily represent what the clause is about (for example Fawcett 1981, 2007; Berry 1995).⁴¹ Halliday kept the label topical Theme, but, in later publications on Theme, decided against the definition of what the clause is about.

However, there is an argument to be made that the Theme can represent what the clause is about but still be conceptually separate from Topic. For that, a distinction between different kinds of aboutness has to be made. When Halliday and Matthiessen (2014: 76ff.) discuss the relationship between the different metafunctions and in particular the relationship between Theme, Subject, and Actor, they show that a clause can be centered around different elements depending on the perspective. Thompson talks about “three possible kinds of 'aboutness'” (2014: 54), which in the unmarked case are actually conflated. That implies the Theme may still be what the clause is about but not in the sense that is meant when discussing the concept of Topic. The clause as a message is about the Theme from a textual point of view, as it grounds the message and acts as a point of departure for further interpretation. However, if the meaning of the word *about* is understood so loosely, one might also just refer to Theme and other concepts as different kinds of prominences to avoid ambiguity.

5.1.3 Theme and Subject

One other concept that is related to Theme is the Subject. However, to speak of the Subject is misleading as there are arguably three kinds of Subject that can be distinguished: the grammatical Subject, the logical Subject, and the psychological Subject (see for instance Seuren 1998: 120–133). Halliday and Matthiessen describe the logical Subject as the

⁴⁰ To my knowledge, Halliday first introduced this label of topical Theme in Halliday (1985) and still uses it in the latest edition of Introduction to Functional Grammar (Halliday and Matthiessen 2014).

⁴¹ I will use the label experiential Theme rather than topical Theme for the same reasons.

"doer of the action" (2014: 79), which is clearly not how the Theme is defined in SFL. The logical Subject corresponds more closely to Halliday and Matthiessen's system of TRANSITIVITY and participant roles, in particular to the participant role of "actor". While the first participant role of a process is likely to be in Theme position, passive constructions demonstrate that there is no one-to-one correspondence between participant roles and clause positions.

The grammatical Subject, on the other hand, is frequently compared to the concept of Theme and they are even used synonymously by some linguists. Lindeberg (1985: 333), for example, compares Theme "with the grammatical subject apart from dummy subjects" and Katz writes: "The notion of a discourse topic is that of the common theme of the previous sentences in the discourse, the topic carried from sentence to sentence as the subject of their predication" (1980: 26). Similarly, Chafe (1976: 44) describes the Subject as a starting-point and explains that the rest of the clause adds new information about the Subject which is relevant to the addressee. On the basis of Chafe's gloss of starting-point and Halliday's gloss of point of departure, it is difficult to see how the concepts of Theme and Subject differ from each other conceptually.

Another common reason why grammatical Subject and Theme are often equated is the fact that the Subject is often described as "what we are talking about" (Chafe 1976: 43), which is similar to 'what the clause is about'. As was shown in the previous section, Topic is often confused with Theme because they both carry the meaning of aboutness. Since the Topic in English is often realized as the grammatical Subject of the clause, it is unsurprising that this ambiguity also affects the relationship between Theme and Subject.

Most of the confusion concerning Subject and Theme arise from the vague functional definitions of both terms in SFL. However, if one considers the formal realization of the two in English, it is quite clear that Subject and Theme are frequently not the same element in the clause. In the unmarked case, the Theme is also the Subject in declarative clauses (see example (48); Subject in bold), but the left-most position may also be occupied by a circumstantial Adjunct (see example (49); circumstantial Adjunct in bold) or even a Complement (see example (50); Complement in bold).

- (48) ***The impact*** would be global.
[E2G_SPEECH_007]
- (49) ***If you load the tray***, the printer begins printing.
[E2G_INSTR_001]
- (50) ***Some*** he took back with him to Tuxedo [...].
[E2G_FICTION_001]

As was mentioned before, Halliday has not described Theme in terms of what the clause is about in his more recent works. Consequently, to my knowledge, Subject and Theme have also not been used as synonyms in the more recent research on Theme, so it seems that most of the ambiguity between the two concepts has been cleared up. Nevertheless, a number of systemicists make the claim that while Theme and Subject are not the same, Subjects are inherently thematic (see Section 5.4).

Interestingly enough, when discussing Theme in the most recent edition of IFG, Halliday and Matthiessen (2014: 89) define it primarily on the basis of it being the point of departure of the message. However, when they introduce the three types of Subjects, Halliday and Matthiessen still equate Theme with the psychological Subject, which they define as "that which is the concern of the message" (Halliday and Matthiessen 2014: 79). Unfortunately, they still do not make explicit what the concern of the message is supposed to mean and in what way this is different from what the clause is about or about what something is being predicated.

5.1.4 Theme and communicative dynamism

As was discussed in Section 5.1.1, the term Theme originated in the Prague school, referring to both the starting-point of the message and given information. Firbas continued Mathesius' work on FSP and developed his notions of Theme and Rheme further. There are three major problems that arose from Mathesius' Theme definition: New information can be positioned before given information, the first position of the clause does not have to be the element from which the speaker proceeds, and a clause does not have to contain any given information at all (Vasconcellos 1985: 24). Based on these inconsistencies, Firbas redeveloped Mathesius' concepts of Theme and Rheme and interpreted them on the

basis of what Firbas calls communicative dynamism (see for example Firbas 1964a/b, 1987, 1992). Firbas defines the degree of CD as "the extent to which the sentence element contributes to the development of the communication, to which it 'pushes the communication forward,' as it were" (1964b: 270). That means if an element has a low degree of CD, it does not advance the flow of the communication as much as other elements do. Firbas defines the Theme as the element with the lowest degree of CD, so the one element that contributes least to the development of the communication (1992: 175). The Theme then transitions to elements with higher degrees of CD, the Rheme.

The fact that Firbas speaks of degrees of CD shows that the distinction between Theme and Rheme is not binary but gradual. In fact, he specifies different kinds of Theme, like Theme proper and Theme proper oriented element, and also transition elements that pave the way between Theme and Rheme.⁴²

With this definition of Theme, Firbas is able to explain clauses that do not contain any given information, which previously would not have had any thematic element. Even if every element in a clause is new information, one of these new pieces of information still contributes to the development of the communication the least and thus is the one that carries the lowest degree of CD. In a clause like (51), *A man* would be considered the Theme of the clause because even if the identity of the man is unknown to the hearer, all other elements advance the communication in a more substantial way.

(51) *A man comes into a bar.*

Firbas also insists on the fact that the Theme, meaning the element with the lowest degree of CD, does not have to come first in the clause. In the unmarked case, the order of Theme and Rheme is linear, but if there is a good contextual reason to deviate from this sequential order, the Rheme may also precede the Theme (Firbas 1992: 172). Firbas identifies four factors that signal thematicity and rhematicity, namely sentence position, contextual factors, semantics and, in spoken language, intonation (1987: 138).

Firbas and Halliday disagree on the proper meaning of point of departure of the message. While Halliday identifies the point of departure based on position, at least in English, Firbas rejects the idea of linear arrangement as a meaningful way to identify the point of

⁴² For a detailed discussion of the different thematic and rhematic elements in Firbas' FSP, see for example Firbas 1989 (Firbas 1992: 176).

departure (Firbas 1987: 145). Other than that, Halliday and Firbas really do speak of different features of the clause as a message but, unfortunately, use the same label. To discuss which meaning of Theme is right or better is quite pointless, as they focus on different aspects.

One of the major problems of Firbas' Theme and CD is that there really is no systematic way of identifying the degree of CD. Firbas mentions a variety of influencing factors but does not explain how these factors interact or how Theme and Rheme can be identified without relying, at least partially, on intuition. From a practical standpoint, Theme in the Hallidayan sense is much easier to identify because it is only based on initial position in English (Davidse 1987: 66). Irrespective of the challenges of CD, Firbas' definitions of Theme and CD certainly have their merits. Unfortunately, this additional system of textual meaning is usually neglected in analyses in the SFL framework.

5.2 The meaning of Theme

5.2.1 Theme as the point of departure of the message

As was shown in the previous sections, Halliday's definition of Theme underwent a series of changes over the years. However, one part of the Theme definition that has stayed consistent since Halliday (1967b: 212) is that Theme is "the point of departure for the clause as a message". If you compare this to the most recent definition of Theme in Halliday and Matthiessen (2014), you notice how the characterization of Theme has become more detailed but, in its essence, has not changed in this respect:

The Theme is the element that serves as the point of departure of the message; it is that which locates and orients the clause within its context. The speaker chooses the Theme as his or her point of departure to guide the addressee in developing an interpretation of the message; by making part of the message prominent as Theme, the speaker enables the addressee to process the message. The remainder of the message, the part in which the Theme is developed, is called in Prague school terminology the Rheme. As a message structure, therefore, a clause consists of a Theme accompanied by a Rheme; and the structure is expressed by the order – whatever is chosen as the Theme is put first. The message thus unfolds from thematic prominence – the part that the speaker has chosen to highlight as the starting point for the addressee – to thematic non-prominence. (Halliday and Matthiessen 2014: 89)

It is important to note that in this extract, Halliday and Matthiessen bring together two aspects of Theme, namely the function of Theme as the point of departure of the message and the realization of Theme (in English) as coming early in the clause. While it is important to identify the formal realizations of any linguistic unit, the function of Theme

must not be confused with its realization (Matthiessen 1992: 75-76). The Theme is the part of the clause which serves as a basis and from which the rest of the message can unfold. By choosing a particular Theme, the speaker offers the hearer a starting ground, with which the message can be interpreted. It just so happens that in English, this point of departure is realized through first position in the clause. Matthiessen considers early position an "iconic" (1992: 75-76) realization of point of departure in the sense that it is natural to depart from a point that comes at the beginning, but this is not a necessary condition of Theme as can be seen in languages such as Japanese⁴³ and Tagalog, which use morphemes to signal thematization (Matthiessen 1992: 75-76).

And yet, the question remains what it means for an element to be the point of departure. Answering this question precisely is surprisingly difficult. One characteristic of Theme that Halliday and Matthiessen (2014: 89) mention is that the Theme guides the addressee in interpreting the message. Accordingly, different Theme choices must also lead to different interpretations of an otherwise identical sentence. This is what Fries (1997: 230-231) attempts to show with the help of this sentence pair:

(52)

- a. *They left their examinations on the table yesterday.*
- b. *Yesterday, they left their examinations on the table.*

While the representation of experience remains the same in both of these sentences, their difference in order changes how they are interpreted. In the first clause, *yesterday* receives the focal attention as it is the last element in the clause and as such the most likely candidate for being the focus of the information unit. In the second clause, *yesterday* is moved to the beginning of the clause, thus making *on the table* the focal unit from an informational standpoint. Instead of receiving focal attention, *yesterday* now provides an orienting context, a frame, on the basis of which the rest of the clause can be interpreted (Fries 1997: 230-231).

By choosing a certain element as Theme, the speaker can thus direct and even manipulate the interpretation of everything that follows without changing its propositional content. This process of thematizing or staging is argued to influence the interpretation of sentences and entire texts (Brown and Yule 1983: 134). This concept of point of departure

⁴³ As will be shown in Section 5.7, it can be argued that position determines Theme in Japanese as well.

and thematic meaning is not just limited to clauses but can also be applied to other linguistic units. For instance, the title of a text has inherent thematic meaning because it is not only the starting point of the entire text but also heavily influences how the rest of the text is interpreted. The same holds true for the first sentences in a paragraph or the first paragraph in a chapter, etc. (Brown and Yule 1983: 133-134).

The choice of Theme is not a question of grammaticality but of appropriateness. Theme appropriateness is largely dependent on context. While one sequence of a given clause can feel acceptable in one context, the same sequence can seem incoherent in a different context (Baker 1992: 124). So, the Theme does not just provide a local context of the clause but also links the clause to the context of neighboring clauses in a text. The context does limit the number of appropriate Theme choices, and yet there is still a great number of options from which the speaker can choose to advance the development of the text (Daneš 1974: 112). It is important to note that the choice of Theme is speaker-dependent. The speaker may factor in the needs or the knowledge of the hearer to facilitate interpretation, but it is ultimately their decision how the message is contextualized.

This idea of point of departure is generally agreed upon by most systemic functional linguists (one notable exception is Fawcett 2007) and also linguists with other linguistic backgrounds (for instance Daneš 1974). Besides the phrase point of departure, a number of other glosses have been introduced to describe the function of Theme: "starting point" (Beneš 1959: 216), "grounding for the message" (Halliday and Greaves 2008: 105), "staging" (Grimes 1975: 323), "anchorage" (Halliday and Matthiessen 2014: 112), "resource for setting up the local context or local semiotic environment" (Matthiessen 1995a: 531), "the peg on which the message is hung" (Halliday 1970: 161), "the setting for the information" (Halliday 1998: 4), and "jumping-off point" (Rashidi 1992: 192).⁴⁴

The functional description of Theme in the Hallidayan sense has been criticized by linguists from other theoretical frameworks (see Brown and Yule 1983; Chafe 1976; Gundel 1988; Hudson 1986; Huddleston 1988) but also by other systemicists (see Fries 1995a; Fawcett 2007). One of the main lines of criticism is its vagueness of meaning, which stems from the fact that the meaning of Theme is typically described in terms of metaphors such as point of departure, wave, pulse, etc. It is not so much that other linguists necessarily

⁴⁴ Interestingly, most of these descriptions are metaphors of space and/or movement, which really captures the way in which Theme is said to operate, namely by identifying a starting point, from which the speaker proceeds to arrive at the rest of the message.

doubt the reality of thematic meaning, it is simply unclear what Theme is actually supposed to mean:

There are many places in the book where he [referring to Halliday and IFG 1] makes a claim which I, for one, was completely unable either to agree with or to disagree with because the categories concerned were so vaguely defined that I could not reliably identify [sic] instances of them. Take the important category 'theme', for example. The following definitions are provided: 'The theme is what the clause is going to be about' (p. 39). 'The Theme is what the speaker selects as his point of departure, the means of development of the clause' (p. 53). Presumably he finds these definitions adequate, because he is able to pick out the themes in any clause and work out the general rules which govern their use [...]. Perhaps he is tuned into language in a way that the rest of us are incapable of, but those of us who can't easily pick out the parts of a clause which define 'what it is going to be about', or its 'point of departure' are simply unable to decide whether any of his claims about themes are right or wrong. (Hudson 1986: 798)

This problem of vagueness is especially apparent when applying THEME to languages other than English. The positional criterion is simply the formal realization of Theme in English and that while the form of Theme in other languages may be different, Theme is not a language-specific concept. And yet, it seems as if the existence of Theme rests to a large extent on the fact that early position in English is meaningful and that the order of clause elements, to some degree, involves a choice. However, if the meaningfulness of a linguistic concept rests so heavily on one particular language, it is questionable if and how Theme can be applied to other languages.

5.2.2 Theme and method of development

The Theme of a clause operates on two levels: the local level and the global level. By choosing one particular unit as the Theme of a clause, the speaker sets up a local context, on the basis of which the rest of the message can be interpreted. At the same time, this choice of Theme also contributes to the development of the rest of the text and the sum of all local clause contexts establishes the general textual orientation.

Daneš (1970, 1974) argues that Theme choices in a text are not random, but that the Themes and Rhemes of different clauses are related. He calls this phenomenon thematic progression (TP), which is a term to describe different systematic ways in which the semantic content of the Theme progresses throughout the text. Daneš postulates three main types of thematic progression: 1. Simple linear TP, where the Theme of the following clause is based on the Rheme of the preceding clause, 2. TP with a continuous (constant) Theme, where the Theme stays the same across a series of clauses, and 3. TP with derived

Themes, where the Themes of multiple clauses are all derived from a so-called hyper-theme (Daneš 1974: 118-120).

Fries (1981, 1995a/b/c) attempts to combine Halliday's original definition of point of departure with thematic progression. He assumes that if the Theme changes the way in which the rest of the message is interpreted, different kinds of Theme choice throughout the text should also influence the way in which the entire text is interpreted (Fries 1995c: 319). Thus, he focuses less on the role of Theme as the local context in individual clauses and more on the relationship between Daneš' thematic progression and the development of the overall text, which he calls method of development. The method of development of a text represents "[t]he way in which a text develops its ideas" and it "affects the reactions of its listeners and readers" (Fries 1995c: 323). In his earliest paper on Theme and method of development, Fries (1981) postulates two hypotheses on the relationship between the two concepts, which he later extends by two additional hypotheses:

1. Different patterns of thematic progression correlate with different genres, i.e. patterns of thematic progression do not occur randomly but are sensitive to genre;
 2. The experiential content of Themes correlates with what is perceived to be the method of development of a text or text segment;
 3. The experiential content of Themes correlates with different genres; and
 4. The experiential content of the Themes of a text correlates with different generic elements of structure within a text.
- (Fries 1995a: 6-7)

Both Fries (1981, 1995c) and other linguists such as Francis (1989, 1990), Berry (1989), Martin (1992a), Davies (1997), and Thompson (2006) have studied this effect of Theme on the method of development. Considering both his own work and that of his colleagues, Fries (1995c) concludes that a relationship between Theme and method of development is evident but that the empirical data does not support all of his original hypotheses. For example, he considers the support of hypothesis 1 only "weak" (Fries 1995c: 320), since Francis (1989, 1990) was able to show that there was some correlation between thematic progression and genre but that these patterns only represented general tendencies and not clear indicators to distinguish different kinds of genres.

Fries' second hypothesis on the relationship between experiential Theme and text interpretation has been tested in a number of studies (for instance Martin 1986, 1989; Benson, Greaves, and Stillar 1992; Halliday 1993). These studies generally show that experiential content does correlate with the evaluation and interpretation of the text by the readers (Fries 1995c: 324-325). In an attempt to demonstrate the meaningfulness of the

method of development in a text, Olsen and Johnson (1989) use traditional readability formulas in order to simplify a text passage and asked readers to rate the comprehensibility of both versions of the text. Surprisingly, readers did not rate the simplified version as more readable, which Olsen and Johnson (1989) attribute to the change in thematic structure and accordingly the change in method of development, which made it more difficult for the readers to understand the text despite its simplified structure.

Studies on language proficiency support the idea that Theme and method of development are crucial for successful communication. Berry (1989) shows that neither children nor advanced level English students are able to master the use of thematic options and for that reason considers Theme one of the topics that need to be further investigated to improve language teaching. Similarly, in her study of university students' essays, Ryshina-Pankova (2006) demonstrates that the complexity of thematic structure is one of the variables that predicted the success of discourse organization and improves cohesion and coherence of texts. Similar results were also presented by Montemayor-Borsinger (2009), who shows that the use of Theme and its complexity are related to the maturity and experience of the authors.

In regards to hypothesis 3, there is a multitude of studies that have explored the relationship between thematic content and genre (for example Berry 1987; Francis 1989, 1990; Ghadessy 1995; Martin 1986, 1989; Whittaker 1995; Corbett 2009; Virtanen 2014). The data shows that the relationship between experiential content and genre relies heavily on which genres are considered and what experiential aspects of the Theme are analyzed. For instance, in their studies on participant roles in Theme position, Francis (1989, 1990) found that only the register of news reports differed significantly with regard to experiential content of the Theme, while the thematic structure of editorials and letters of complaint were comparable. In contrast, Berry (1987) found that place names in Theme position correlated fairly strongly with the registers of coffee parties, committee meetings, travel brochures, and guide books and could be used to distinguish these registers from each other. Other studies like Martin (1986, 1989), and Xiao (1991) report similarly strong correlations between Theme and register. However, the Theme features that were analyzed in each study were in part very different, which is why their comparability is questionable. Also, these earlier studies on the Theme-register relationship used only very limited data (Fries 1995c: 339).

One common thematic difference between registers is the use of marked Themes both in terms of general frequency and type. Ghadessy (1995) shows that guidebooks feature a substantial amount of Place Themes, since one of their main functions is to guide the reader through the spatial realm, whereas sports commentaries are more likely to conceptualize the method of development temporally. Similarly, in an analysis of academic articles on economics and linguistics, Whittaker (1995) shows that circumstance Themes are generally common in both fields of study but that the frequency of specific types of circumstance Themes vary in the data. And lastly, Corbett (2009) demonstrates that Place and Time Themes can be found in popular-scientific writing as well as in academic texts, but that the academic register on average uses fewer circumstance Themes and that their meaning tends to be more abstract, due to the fact that they often do not refer to places and times in the outside world but within the article itself.

Besides limited data, one other shortcoming of these studies is the lack of comparison between experiential Themes and general experiential meaning in different registers. For example, one register may include a higher number of Time Themes than another register simply because it is more common in general to specify temporal locations in that register. Naturally, if a register includes more circumstances of Time on average, the likelihood of Time Themes is also higher. This, however, is then not primarily a difference in Theme or method of development but a byproduct of the differences in experiential representation. Further studies on the relationship between registers and the frequency of particular experiential Themes are thus essential. So, while a certain circumstance type may be comparable in overall frequency between two registers, its frequency as Theme may still be significantly different. Thus, a simple count of experiential Themes does not allow the analyst to decide whether these differences are due to skewed probabilities in the TRANSITIVITY or THEME network.

Freiwald (2016) has considered the relationship between non-Subject elements both in Theme and Rheme position in the register of popular scientific writing. He found that Concession, for example, was only the eighth most frequent circumstance type in the register generally but had the highest likelihood of being thematic if present in the clause at over 80%. Since he only considered a single register, Freiwald (2016) could not comment on whether this effect is purely a register effect or whether circumstances of Concession customarily lend themselves well to be used as Theme. It would be interesting to see

whether these distributions differ between registers or whether marked Themes have a comparable thematic likelihood across different registers.

Regarding Theme and method of development, Fries (1995b) argues for four general functions that Themes can assume in developing the rest of the text: Themes can provide information, which is required to interpret the main message, cancel an assumption which has been established in the previous context, prevent temporal or locational misinterpretation, and highlight the point of elaboration. So, while a one-to-one correspondence between Theme choice and method of development could not always be demonstrated, a relationship between the two on some level is generally accepted in SFL.

Yet, relying on method of development to define Theme functionally is also problematic. One criticism of method of development that has been pointed out repeatedly is that Daneš' types of thematic progression only work for either very short or pre-constructed texts and that any authentic text of a certain length will employ a mixture of many different kinds of progression, so that the analysis of the method of development of a text becomes problematic (Fries 2009: 24). This criticism can be dismissed as a mere methodological and not a conceptual issue. However, if the method of development is generally so complex that it is virtually impossible to analyze it, its usefulness in characterizing Theme is also questionable.

Another common line of criticism regarding Theme is its lack of functional homogeneity. Fries' (1995b) list of Theme functions in relation to method of development works well for both (unmarked) Subject Themes and other kinds of (marked) experiential Themes. Textual Themes and interpersonal Themes, on the other hand, do not contribute to the method of development like experiential Themes. However, Halliday does consider textual and interpersonal elements at the beginning of the clause thematic, which suggests that there must be more to the function of Theme than method of development. Huddleston (1988: 161-162) was among the first to criticize Halliday's notion of Theme, which is supposed to represent a unified function yet is made up of elements that are very different in nature, just because they all happen to be positioned early. And yet, Matthiessen and Martin (1991) claim that non-experiential Themes contribute to the method of development – not in the exact same way that experiential Themes do, but they do further the development of the text:

(53)

- a. *Raju almost sobbed at the thought of the disappointed child – the motherless boy. There was no one to comfort him. Perhaps this ruffian would beat him if he cried too long.* (Narayan, *Malgudi Days*, p. 120)
- b. *This ruffian would perhaps beat him if he cried too long.*
(taken from Matthiessen and Martin 1991: 49)

An interpersonal Theme like *perhaps* in (53a) sets up the local context of modality and signals to the reader that the text developed from a matter of fact to a matter of possibility. Accordingly, while the interpersonal Theme does not develop the experiential nature of the text, the method of development is still influenced by this choice of Theme (Matthiessen and Martin 1991: 49).⁴⁵

Matthiessen and Martin's (1991) argument that the interpretation of the text is influenced by the use of interpersonal Themes like *perhaps* does have merit. However, it is not immediately obvious why this function that is restricted to Theme positions. Clauses such as example (53b) could develop the text in a similar direction of possibility with the interpersonal element positioned in the Rheme. And even if early interpersonal and textual elements develop texts differently from such elements later in the clause, their contribution to the method of development is so different from experiential Themes that it is difficult to argue for a unified function of Theme in terms of method of development that is not so general that it is no longer useful for text analyses. Fries (2009) comes to a similar conclusion when he concedes that the Theme does not solely develop the orientation of the text but that Rheme contributes to the method of development as well. However, Fries is hesitant to lump all parts of the clause together in one functional category just because they function similarly on a very general level:

I worry that the notion of development is being taken in a much more general way than I intended. If the notion gets too general, then all texts develop, and all parts of each text contribute to the way it develops. The result of this interpretation is that the notion of method of development applies to everything in the text, and therefore it becomes unusable. (Fries 2009: 25)

The same argument should be made for non-experiential Themes. While they do contribute to the method of development, they do so in such a different way that a conflation of their function and the function of experiential Themes is not helpful. In summary, method of development is useful in unifying the functions of different experiential Themes throughout a text even if their function within the clause is very different. However, the functions of textual and interpersonal Themes are fundamentally different in terms of text

⁴⁵ Matthiessen and Martin (1991) only comment on Huddleston's criticism of interpersonal Themes and do not make an argument for textual Themes like *and*.

development, which makes a unified definition of all of these early elements problematic. In this context, Fawcett's (2007: 137) assessment is quite accurate when he claims that all of the different thematic roles can only be unified if the function of Theme is defined in such general terms that the definition itself is hardly useful:

However, no great harm would result if one wished to say that the various types of 'theme' in English are concerned with various aspects of the presentation of the Performer's viewpoint (or indeed any of the other proposed generalisations). This is because such generalizations are so general that they do not, in my view, give a genuine sense of explanation. So no great good would come of it either!

5.2.3 Theme as a wave

Another popular metaphor to describe the function and progression of the Theme is the wave metaphor. First introduced by Halliday (1979), it was Christian Matthiessen in particular who popularized the wave metaphor to describe Theme (see for instance Matthiessen 1988, 1992, 1995a).⁴⁶ One area of disagreement in the SFL community concerns the extent of Theme (in English), meaning where to set the boundary between Theme and Rheme (see Section 5.4). To address this issue, Matthiessen compares Theme to the movement of a wave to illustrate the thematic status of different elements in the clause. Thematic prominence is strongest at the very beginning of the clause and decreases as the clause progresses. This is particularly relevant in the case of a marked Theme, where the Subject is oftentimes not analyzed as part of the Theme (Matthiessen 1992: 52). The wave metaphor helps to illustrate that the Subject still bears thematic meaning in such cases but that it does not act as the point of departure as much as the marked Theme does. So, while in language analyses a discrete boundary between Theme and Rheme is necessary, the wave metaphor demonstrates that Theme is rather a question of degree than of either-or (Martin 1992b: 151). Following this approach, the thematic space of a clause can be considered less like a constituent and more like a zone (Gómez-González 2001: 132) containing elements of varying degrees of thematic importance.⁴⁷

The wave metaphor is useful in describing the formal characteristics of Theme (in English), but it also helps to understand Theme and other modes of meaning on a conceptual

⁴⁶ Alternatively, Martin, Matthiessen, and Painter (1997) use the metaphor of a pulse.

⁴⁷ Niemietz, Neumann, and Freiwald (2017) also found the metaphor of a zone useful to describe elements of varying thematic prominence in German.

level. As has been pointed out in Section 5.1, thematic meaning represents a kind of prominence, which operates simultaneously with other kinds of prominence. For example, textually, the clause can be separated in thematic prominence, realized through position, and informational prominence, realized through the tonic. In the unmarked case, these two prominences operate in opposite directions: As thematic prominence decreases, informational prominence increases.

The comparison of Theme to a wave suffers from the same problems as Theme as the point of departure: Any metaphorical description of Theme only helps to vaguely grasp the nature and function of Theme. While it is useful to understand that thematic prominence is a question of degree, the wave metaphor is only meaningful and useful if its starting-point has already been agreed upon.

One other problem of the wave metaphor that, to my knowledge, has not been pointed out in the literature concerns the status of experiential Themes. If thematic prominence is in fact decreasing, that would mean that textual and interpersonal Themes carry greater thematic prominence than experiential Themes in English, given that they have to be positioned before the first experiential element if they are thematic. One could argue that they represent this peak of thematic meaning because some of them are inherently thematic (Halliday and Matthiessen 2014: 109). However, looking at a conjunction like *and* at the start of a clause as the thematic peak seems counter-intuitive, in terms of both point of departure of the message, on the basis of which the rest of the message is to be interpreted, and method of development. This also raises the question of why textual and interpersonal Themes cannot exhaust the thematic potential of a clause and why every Theme must be anchored in the realm of experience (Halliday and Matthiessen 2014: 111-112) if non-experiential Themes do in fact represent the thematic peak.

5.2.4 Other attempts at defining Theme

One common way of distinguishing between Theme and Given is to describe Theme as more speaker-oriented and Given more hearer-oriented. While it is ultimately the speaker's choice what they present as given information with the help of definite markers, word order, and intonation, it is the hearer that they have in mind when they make these choices. The speaker can only present information as given if they know or can at least assume that the hearer can identify what that piece of information refers to. Otherwise,

they risk being ambiguous or outright incomprehensible. Theme, on the other hand, is controlled entirely by the speaker. Of course, situation, context, register, and other dynamic factors restrict the number of appropriate Theme choices, but still the general direction of the discourse is dependent on the speaker's intentions. For this reason, Halliday (1970) proposes another definition of Theme, namely Theme as the speaker's angle. Theme thus represents "what is uppermost in the speaker or writer's mind" (Vasconcellos 2008: 49), which Halliday and Matthiessen (2014) argue is especially apparent in the thematic structure of interrogatives. Both *wh*-interrogatives and *yes/no*-interrogatives are so restricted in their choice of unmarked Theme structure – *wh*-element in *wh*-interrogatives and Finite + Subject in *yes/no*-interrogatives – not because there is anything fundamentally different between them and declaratives but because the interrogative mood signals the speaker's angle much more clearly. If a speaker asks the hearer a question, the information that they are seeking is obviously what they want to talk about. So, this regular pattern has been integrated in the thematic system of interrogatives (Halliday and Matthiessen 2014: 102). Similarly, comment Adjuncts are frequently placed in Theme position in English because they mirror the attitude of the speaker (Fawcett 1981: 168).

Ravelli (1995) understands Theme as the point of departure similar to Halliday (1967b), but she does so on the basis of what she calls a dynamic perspective. From this dynamic perspective, a clause unfolds in real time, during which the hearer has to process and interpret the meaning of that clause. In order to facilitate this processing, the speaker generally chooses a Theme which allows the hearer to quickly identify the mood and accordingly the speech function of the clause. In the case of a declarative, for instance, the speaker will try to quickly arrive at the finite verb as this concludes the mood interpretation by the hearer (Ravelli 1995: 223). In Ravelli's mind, this explains the tight relationship between mood and unmarked Theme structure (Ravelli 1995: 225). Hawkins (1992) makes a similar argument when he claims that syntactic weight is the major determinant of word order variation due to reasons of processing. Syntactically less heavy elements are more likely to come early in the clause because the processing effort of the hearer to identify mood is lower if the mood elements of Subject and Finite come early (Hawkins 1992: 215).

Ravelli's (1995) position on Theme is very interesting, as it understands Theme as point of departure from a very different, more interactive perspective. Nonetheless, it is

quite clear that processing effort and syntactic weight cannot be the only influencing factors to determine the Theme. Speakers do choose marked Themes and Themes of great lengths even if they potentially impede processing. Therefore, the point of departure must have a function that goes beyond the dynamic processing perspective.

In her attempt to analyze the thematic structure of Dari, Rashidi (1992) adopts a cognitive perspective on Theme. Moreover, she uses semantic criteria rather than positional criteria to identify Theme, but that may also be the way in which Theme in Dari is realized. Rashidi makes an interesting point on how, from a cognitive perspective, early position is the most natural place to realize the point of departure of the message (Rashidi 1992: 198). Other than that, Rashidi's interpretation of Theme from a cognitive viewpoint seems like a paraphrase of Halliday's description of Theme as point of departure:

Theme is the clause-level constituent that the encoder uses as the starting point of the message, the constituent that begins moving the decoder towards the core of the communication. Theme is the essential ideational jumping-off point directing the decoder's attention to the ultimate goal of the communication, the kernel of the message, the Rheme. (Rashidi 1992: 192)

5.3 Rheme

This study revolves around differences in thematic structure in English and German and their effects on translation. For this reason, its primary focus is on the meaning and realization of Theme. However, even if the rhematic portion of a clause will not be analyzed in the course of this study, it is still relevant to discuss the meaning of Rheme, as it is the counterpart of Theme in a thematic analysis.

From a formal perspective, identifying the Rheme is rather simple if the Theme of a clause has already been identified because the Rheme is simply "everything else" (Eggs 1994: 275). One problem of this formal realization of Rheme is that while early position is in practice typically restricted to one or two elements in the clause, the Rheme can consist of a large number of different kinds of elements. Quite frequently, the Rheme does not just come late but actually takes up almost the entirety of the clause (Fawcett 2007: 20). Formally, these units are not related other than by not standing at the very beginning of the clause.

In the previous sections, it was shown that defining Theme functionally is a challenging undertaking because it is difficult to unify all the different kinds of thematic meaning in

terms of their function in the clause. A functional definition of Rheme is even more problematic. Both formally and functionally, the Rheme seems like the 'rest' category of everything that is not thematic. It is the development of the message following its point of departure, but this definition is so general that it really does not seem helpful at all. In fact, it is conspicuous that most descriptions of systemic functional grammar discuss the meaning of Theme at length, but do not dedicate more than a couple of sentences to the Rheme – a rule to which this thesis will not be an exception. And yet, Halliday and Matthiessen (2014: 94-96) do speak of a Theme-Rheme structure, which implies that the Rheme is a constituent on some level of delicacy.

While I do believe that the discussion of Theme is meaningful and that elements at the beginning of the clause (in English) may serve a unified function on some level of abstraction, I have to agree with critics like Fawcett (2007) and reject the idea of a constituent called Rheme. Both formally and functionally, the Rheme is simply everything that is not the Theme, in which case it is more meaningful and less ambiguous to leave the rest of the clause blank rather than to give the entire remainder of the clause a single label (Fawcett 2007: 20). I will continue referring to Rheme for reasons of convention but use it as a synonym of non-thematic elements.

5.4 Theme in English

Given the elusiveness of Theme function, it also comes as no surprise that the extent of Theme in English and other languages is a highly debated issue. The function of Theme is not tied to any language in particular, but formal realization is always language specific. In English, Theme in the Hallidayan sense is realized through early position, which is generally agreed upon in the systemic functional framework. However, no consensus has been reached yet on what early position means, in the sense of how far into the clause the Theme extends and where the boundary between Theme and Rheme lies.

Halliday and others have used the wave metaphor to describe Theme, in particular to show that thematic prominence does not just end abruptly but instead decreases as the clause unfolds (Halliday 1994: 337). So, in any case, the boundary between Theme and Rheme will always be fuzzy and the thematic status of an element a matter of degree. Nev-

ertheless, analyzing thematic structures in text requires a clear-cut decision on which elements to still label Theme and which elements are outside of the Theme zone (Thompson and Thompson 2009: 46-47).

Berry (1996) is one of the first and, in my opinion, one of the most successful systemists to try and systematically determine which elements in the clause carry thematic meaning. She establishes the fact that positioning an element at the beginning of the clause in English is one form of prioritizing meaning, which is why this position is meaningful (Berry 1996: 28). But, since Theme is first and foremost a functional unit, its formal realization must also mirror its functional meaning. Hence, if Theme is defined as the point of departure of the message, every element in the clause that functions as this point of departure has to be included in the Theme. Yet, since point of departure is so vague, there really is no objective way of deciding which element(s) serve(s) this function.

To avoid this vagueness, Berry (1996: 18) rather understands Theme as the speaker's or writer's main concern, which she believes can only be determined by asking speaker's and writers themselves or, if that is not possible, by asking the reader (Berry 1996: 23). If Theme is the resource of expressing this kind of prioritized meaning, elements in early position must correspond to what speakers consider their main concern to be. To test for this, Berry used three short passages of authentic texts taken from information material provided to students by the University of Nottingham and consulted with the writers regarding their main concerns while writing the texts. In total, they specified 21 prioritized meanings that they had had in mind while writing the texts. She then summarized ten suggestions from the SFL community on how Theme is formally realized (Berry 1996: 29-31) and went through all main clauses of the texts to determine which hypothesis best captured the meaning specified by the writers. These hypotheses are primarily based on the thematic structure in declarative clauses.

Of these ten hypotheses, only hypotheses 1 to 6 will be discussed at this point.⁴⁸ Hypothesis 1 describes Theme as the very first constituent in the clause. This hypothesis is introduced by Halliday (1985), but only as a working hypothesis to later be rejected in favor of hypothesis 2, the first ideational element hypothesis (Berry 1996: 29). To my knowledge, no systemic functional linguist argues that only the very first position in the

⁴⁸ Hypotheses 7-10 are actually not suggestions on how to identify Theme in a given clause but rather discussions of the nature of Theme. For example, hypothesis 7 represents Matthiessen's (1992) claim that thematic prominence decreases gradually, and that there really is no way of decisively drawing a line between Theme and Rheme. While such discussions are very meaningful, they cannot serve as a consistent basis for annotating Theme in a text.

clause is the position that systematically realizes Theme in English. However, it is a useful approximation of Theme in corpuslinguistics since the very first element of a sentence is easily queriable.

Hypothesis 2, the first ideational element hypothesis, was originally proposed by Halliday (1967a/b) and is still used to describe English Theme realization in Halliday and Matthiessen (2014). According to this hypothesis, "the Theme of a clause extends from the beginning up to, and including, the first element that has an experiential function – that is either participant, circumstance or process" (Halliday and Matthiessen 2014: 112). Halliday and Matthiessen (2014: 110) argue that textual and/or interpersonal elements alone do not take up the full thematic potential of the clause and that only an experiential element can complete the thematic grounding of the message. This is evident from the fact that a circumstantial Adjunct can still be moved to the beginning of the clause to contextualize the rest of the message with or without a textual or interpersonal Theme present (Halliday and Matthiessen 2014: 110). This shows that even with an initial textual or interpersonal element, it is still a thematic choice which of the available experiential elements comes first. In this hypothesis, every Theme includes exactly one experiential element. It is safe to assume that the first ideational element hypothesis is the most well-known hypothesis inside and outside of the SFL framework.

Hypothesis 3, the Subject hypothesis, is another popular strategy to analyze Theme in English. As the name suggests, in this hypothesis, the Theme includes everything up to the Subject of the clause. In contrast to hypothesis 2, the Theme can now also include more than one experiential element in case of a marked Theme. A great number of linguists prefer the Subject hypothesis over the first ideational element hypothesis (for instance Enkvist 1973; Davies 1997; Downing 1991; Ravelli 1995; Fawcett 2007; Montemayor-Borsinger 2009), especially for the purposes of discourse analysis. The Subject often maintains the topic of the discourse and thus lends itself well to analyzing thematic progression and the method of development (Thompson and Thompson 2009: 54). Without the Subject, thematic strings would be interrupted frequently by circumstantial Themes. Moreover, Fawcett (2007: 72) argues that the choices in "the relevant SUBJECT THEME system are independent of the choices in the PROMINENCE system for that Adjunct type". In his mind, every Subject – with few exceptions – carries thematic meaning, and the choice of which experiential element to highlight in Subject position is independent from the question of whether to front a circumstantial Adjunct or a Complement. Ravelli (1995) points

out yet another advantage of treating Subjects as inherently thematic: The hearer can only process the speech function of a clause if they identify beginning and end of the Mood elements. If the Theme in fact serves as the grounding based on which the hearer interprets the rest of the message, this processing aspect should be addressed (Ravelli 1995: 227).

Hypotheses 4 to 6 are all related to the verb phrase. Hypothesis 4, the pre-verb hypothesis, considers everything part of the Theme that is positioned before the finite verb. Compared to hypothesis 3, modal Adjuncts positioned between Subject and Finite would also be included. Berry (1995) uses the pre-verb hypothesis in her analyses.

Hypothesis 5 is the auxiliary verb hypothesis, which was first proposed by Stainton (1993). In this hypothesis, the finite portion of the verb phrase is also considered to be inherently thematic. The advantages of the auxiliary verb hypothesis are that it concludes the Mood of the clause and also captures modal meaning. Lastly, hypothesis 6, the lexical verb hypothesis, was proposed by Berry (1996) and treats the entire verbal unit as part of the Theme.

Berry (1996) considers all of these hypotheses in her discussion of Theme and evaluates how well they match the answers given to her by the authors of the information material. The further thematic meaning is extended in the clause, the more accurately the 21 prioritized meanings are accounted for. In the end, Berry comes to the tentative conclusion that the lexical verb hypothesis is the most promising because it captured 20 out of the 21 prioritized meanings (Berry 1996: 46).

I am not convinced that this is the most plausible way of identifying the Theme in English or any other language. Asking language users to comment on their own language use is problematic in itself, and naturally, the longer the Theme is, the more likely it will include the prioritized meanings specified by the authors. So, in a sense, it was a foregone conclusion that the hypothesis with the longest Theme extent also turned out to be the most promising. Also, Berry does not comment on how many elements that were analyzed as Theme were not part of the author's answers. Nevertheless, at least Berry (1996) attempts to find a systematic way of identifying Theme that is not just based on language intuition. She also shows that early position is in fact associated with the speakers' concern, given that even hypothesis 2, the hypothesis with the shortest Theme zone, accounted for 14 of the 21 prioritized meanings (Berry 1996: 37). Further Theme analyses using an experimental set-up may prove to be fruitful.

In my opinion, Halliday and Matthiessen (2014) have argued convincingly that the Theme in English must contain an experiential element. It is apparent that the choice of Adjunct or Complement prominence is independent of the choice of placing a cohesive or modal Adjunct at the beginning of the clause. Besides, a Theme without any reference to the realm of experience seems to be an inadequate grounding of the message. A single textual element like *moreover* can, to my mind, not serve as a sufficient local context on the basis of which the rest of the message can be interpreted. Additionally, Themes that only consist of a textual or interpersonal element would frequently interrupt thematic progression and render an analysis of the method of development pointless.

Treating the Subject as an inherently thematic element clearly has its advantages. Especially when defining Theme on the basis of method of development, the inclusion of the Subject seems almost unavoidable. That being said, making the Subject thematic by definition arguably takes away from the significance of marked Themes. In a language like English, especially, where word order is fairly fixed, fronting a circumstantial Adjunct or a Complement must be contextually motivated. If the local context of a clause always relied on the Subject, the contextualizing function of initial Adjuncts would be heavily undermined, even if they are argued to be more thematically prominent.

Hypothesis 4 to 6 seem hardly convincing. In practice, the Theme analysis in hypothesis 4 will be mostly identical to the Theme analysis of the Subject hypothesis. The reason why the occasional modal Adjuncts in between Subject and Finite would be inherently thematic remains unclear. Similarly, tense and modality realized through the finite verb are in my opinion not essential for the local contextualization of the clause.

The lexical verb hypothesis is the most questionable hypothesis in my opinion. If the Theme included the entire verb phrase, a considerable number of clauses would consist only of a Theme. In other words, the entirety of a clause would then serve as the point of departure of the message with no destination to depart to. Moreover, the process often represents new information. While THEME and INFORMATION are two independent systems, their conflation in the unmarked case is plausible. It is easier to process the meaning of a message if its point of departure is familiar information. This relationship would have to be re-evaluated if the lexical verb were thematic.

In summary, the first ideational element hypothesis⁴⁹ and the Subject hypothesis are, in my opinion, most useful to analyze Theme in English. The question which of these two hypotheses is preferable largely depends on the type of analysis carried out. As was mentioned in Section 5.2.2, the Theme operates on two levels simultaneously, the local clause level and the global text level. While these two functions are always at play, an analyst can focus more on the clause level or the text level depending on what their unit of analysis is (Rose 2001: 29).

5.5 Multiple Themes

One aspect that all of the Theme hypotheses discussed in the previous section have in common is that they can consist of more than one element. In each of the commonly used formal descriptions of Theme, the Theme must include at least one element that carries experiential meaning in order to conclude the thematic grounding of the message. Notably, the first experiential element does not have to be the very first element in general, which is why it is common to also find textual and interpersonal elements in thematic position in English. Such elements are called textual and interpersonal Themes and belong to the same Theme unit as the obligatory experiential Theme. If a Theme includes more than one element, it is called a multiple Theme.

The order of multiple Themes is typically textual[^]interpersonal[^]experiential (Halliday and Matthiessen 2014: 107; see example (54); textual Theme underlined, interpersonal Theme double-underlined, experiential Theme in bold). Most Themes in English are simple Themes and if a Theme includes more than element, it is usually a textual Theme followed by an experiential Theme. However, in theory, there can be a number of different textual and interpersonal Themes paving the way for the experiential Theme.

(54) And of course, **I** did not fail to visit the fine cathedral, [...].

[E2G_FICTION_009]

Halliday and Matthiessen (2014) distinguish between three different kinds of textual and interpersonal Themes respectively, which will be briefly introduced here: Textual Themes

⁴⁹ In my opinion, the name *first experiential element hypothesis* is more appropriate and will be used throughout the rest of the thesis since the Theme concludes with the first experiential and not the first logical element.

can be categorized as (1) continuatives, (2) conjunctions, and (3) cohesive Adjuncts. Continuatives are words such as *well*, *oh*, and *now* which are used to signal a change of direction in the discourse. Conjunctions and cohesive Adjuncts have similar functions in that they both create a textual link between two clauses/clause complexes (Halliday and Matthiessen 2014: 107-108). Conjunctions refer to the word class conjunctions and are more commonly used between clauses inside of a clause complex but can also be placed at the onset of a clause complex.

The three types of interpersonal Themes include (4) Vocatives, (5) modal Adjuncts, and (6) finite verbal operators. Vocatives are an interpersonal tool to specify which person is being addressed. Modal Adjuncts, which include words like *unfortunately*, *probably*, and *of course*, do not contribute to the experiential meaning of the clause but reflect the personal assessment of the speaker. Lastly, finite verbal operators are finite verbs that are placed in Theme position to signal the *yes/no*-interrogative mood.

If a clause includes one or more of these non-experiential elements, they have a high probability of being thematic. In fact, continuatives and conjunctions are inherently thematic, which means if they are realized in a clause, they must be placed at the very beginning before the first experiential element (Halliday and Matthiessen 2014: 109). All other non-experiential items are characteristically thematic, which means they are likely positioned in the Theme but may also occur in different positions in the clause (Halliday and Matthiessen 2014: 109-110).

The fact that continuatives and conjunctions are inherently thematic in English represents a problem for some systemic functional linguists. One of the essential aspects of THEME and other system networks is that meaning is created based on the choices that the speaker makes. Yet, if an element is thematic by nature, using it as the point of departure of the clause does not represent a choice but is dictated by the language system, and its meaningfulness as a thematic element is therefore questionable (Fries and Francis 1992: 53). Halliday and Matthiessen (2014) argue that while the speaker does not have a choice in using continuatives and conjunctions thematically in English, this inherently thematic status was not established arbitrarily. The language system evolved in such a way because continuatives and conjunctions have such discursive forces that it is most natural for them to function as the point of departure. This inherent thematic nature led to them being moved to the beginning of the clause permanently (Halliday and Matthiessen 2014: 109).

Another issue with non-experiential Themes that has already been touched on in Section 5.2.2 is that they function very differently from experiential Themes. For instance, in terms of method of development, experiential Themes clearly contribute to thematic progression while textual and interpersonal Themes arguably do not or do so in a very different manner. For this reason, Fawcett (2007) questions what textual, interpersonal, and experiential Themes actually have in common besides coming early in the clause, which is a formal and not a functional criterion. Also, it is unclear whether *multiple Theme* refers to one constituent, the Theme, or to a combination of several Themes. And if it is indeed one constituent, the question remains how a multiple Theme can be a meaningful element if its parts serve different functions in discourse (Fawcett 2007: 70).

5.6 Marked Themes

In Section 4.4, it was shown that THEME and MOOD are interrelated and that the prototypical order of elements is largely determined by the mood of a clause. For this reason, some types of Theme are used more frequently than others, depending on the mood of the clause. This distinction between common and uncommon Themes is represented through Theme markedness.

Markedness is a prevalent linguistic concept used by different linguists in various linguistic frameworks (for example Waugh 1982; Greenberg 1990; Croft 1996). The term was first introduced by Trubetzkoy (1931) to specify phonological distinctiveness and has since then been used to compare a variety of linguistic units. As is often the case with popular linguistic concepts, markedness lacks a uniform definition. In fact, Haspelmath (2006) identifies twelve different ways in which markedness is understood in linguistics. A common description of marked linguistic units is that they are rarer, more complex, and less natural compared to their unmarked counterparts (Haspelmath 2006: 26).

The markedness of Theme is also often described based on rarity, with unmarked Themes being more commonly used than marked Themes (see for instance Bloor and Bloor 1995). However, frequency is not the defining property but rather a symptom of Theme markedness: an unmarked Theme is the Theme which is least motivated (Halliday 1967b: 219) and which requires the least amount of organizational effort (Halliday 1976: 178). In other words, the unmarked Theme is the default option, which is used unless the speaker has good reasons to deviate from the norm. Thompson (2007) contrasts three

declarative clauses with the same experiential content but different Themes and argues that the clause with the unmarked Theme (in this case the Subject Theme) is "the one which gives fewest clues to the context in which it is likely to be found" (678). The problem with this distributional markedness, as it is called in Haspelmath (2006), is that there really is no reliable way to decide which pattern requires more or less motivation. That is why Haspelmath (2006: 45) prefers to use the feature of frequency in text to distinguish marked and unmarked structures.

In spoken language, there is yet another strategy to identify marked Themes, which is intonation (Steiner and Teich 2004: 178). As was shown in Section 5.1.1, the focus of the information unit typically falls on the last element in the clause, which constitutes unmarked intonation. However, if a clause features a marked Theme, the focus may also be awarded to an early element in the clause, typically to highlight or contrast that piece of information.⁵⁰

Theme markedness is language-dependent, and in English, it depends on the mood of the clause. As was touched on in Section 4.4, the unmarked Theme choices for each mood type also reflect their most common speech function, so that interrogatives typically start with the information that is sought by the speaker. For this reason, the unmarked Theme in *wh*-interrogatives is the *wh*-element, for *yes/no*-interrogatives, it is the finite verbal operator, expressing polarity, and the Subject, and in imperatives, it is the process or Predicator Theme (Halliday and Matthiessen 2014: 101-103). In the unmarked pattern of declarative clauses, the Subject is conflated with Theme, which makes all other experiential Themes, namely circumstance Themes⁵¹ (see example (55); circumstance Theme in bold), Complement Themes (see example (56); Complement Theme in bold), and Predicator Themes (see example (57); Predicator Theme in bold) marked.

⁵⁰ As I do not work with spoken language in this project, intonation as a proxy of markedness will not be pursued further here.

⁵¹ It would be more appropriate to speak of circumstantial Adjunct Themes to avoid a mixture of MOOD and TRANSITIVITY labels. However, in my annotations, only circumstances map onto circumstantial Adjunct (see Section 6.3), which is why I prefer the label circumstance Theme mostly for reasons of space and convention. Also, for more delicate Theme analyses, I also speak of Actor or Reason Themes, which also refer to experiential categories. A combination of labels is thus unavoidable.

- (55) *By June 30, Iraqi soldiers in the ranks will report up through an Iraqi chain of command to Iraqi generals.*
[E2G_SPEECH_004]
- (56) *Firewood he'd cut and delivered with a vengeance.*
[E2G_FICTION_004]
- (57) *Beckoning with their own gentle magic are the sea bright expanses of Holderness [...].*
[E2G_TOU_010]

While the distinction between marked and unmarked Themes is a binary distinction, it is also clear that markedness in terms of rarity and distributional markedness is a scale. That means not all marked Themes are equally marked. A Subject Theme is used in 70-80% of declarative clauses (Fawcett 2008: 109; Matthiessen 1995a and Freiwald 2016 report similar numbers), which clearly makes it the most frequent Theme in declaratives and the system of Theme markedness a skewed system where one choice is much more probable than every other choice in the system (Halliday 1991). The second most frequent Theme is the circumstance Theme, which most often sets up a temporal or local context (Baker 1992: 132). Halliday and Matthiessen (2014: 99) describe the Complement Theme as the most marked Theme option, but clearly Predicator Themes are even more uncommon. Usually, it is not just the process that is moved into Theme position, but the entire Predicate, including Complements. Theme markedness in terms of frequency is register-dependent. Especially if one considers more delicate Theme options, for instance different kinds of Adjunct Themes, their frequency may vary significantly depending on the register (Matthiessen 1995a: 549). For example, it can be expected that tourism leaflets will feature more Place Themes than scientific articles.

Given the considerable differences in thematization potential between the various marked Themes, Fawcett (2007: 59-60) questions the usefulness of treating circumstance Themes and Complement Themes as part of the same Theme markedness system and proposes to separate them into two subsystems of markedness. And indeed, it seems counter-intuitive to group two types of Themes in the same markedness category when circumstance Themes are forty times more likely to occur in English declaratives than Complement Themes (Freiwald 2016: 46). The reason why Halliday (1967b: 219) treats all non-Subject Themes as part of the same markedness system is that the choice of one marked Theme prohibits the choice of a different marked Theme, which in his mind shows

that they are in a paradigmatic relationship. In other words, if the speaker has decided to use a circumstance Theme, they cannot also move a Complement in front of the Subject of the clause, which makes them part of the same system of choice.⁵² That is also the reason why textual and interpersonal Themes are not part of the same markedness system since the choice of textual and interpersonal Themes does not impede marked Theme choices.

Adjunct and Complement fronting have been studied extensively, and multiple reasons for choosing a more marked word order in English have been identified in the state of the art. Fries (1995b) postulates four major functions that marked Themes can serve in a text:

1. Providing information which is required to interpret the main message
2. Cancelling an assumption that has been established in the previous context
3. Preventing temporal or locational misinterpretation
4. Highlighting the point of elaboration

Above that, Kies (1988) adds the connective function, where marked Themes represent information that was previously mentioned in the text and thus create a connection between the sentence at hand and the rest of the discourse.

The main function of fronted circumstantial Adjuncts has been described as providing a circumstantial, oftentimes temporal, or spatial framework or setting (see for instance Chafe 1976; Brown and Yule 1983; Virtanen 2014). The main difference between a circumstantial Adjunct that comes early in the clause and one that is positioned late is scope. Fronted Adjuncts have the potential to exceed the local context of the clause and take scope over larger stretches of the text until they are cancelled by another circumstantial framework (Verstraete 2004: 819) and may also be used "to signal boundaries between discourse spans" (Crompton 2006: 249). Kies (1988: 60-61) terms this function the scene-setting function where Adjuncts are used to present a new scene of discourse or change an existing scene.

Bestgen and Vonk (2000) had participants read sentences that included either a Time Adjunct in initial position, in final position, or no Time Adjunct at all, and measured average reading time. If the sentence constituted a shift in setting, the sentences that featured an initial Time Adjunct had a lower average reading time than sentences with final or no Time Adjuncts. This suggests that the use of circumstantial Themes facilitates the processing of a shift in setting.

⁵² The results in this thesis support Halliday's claim, as there is no clause in the data that simultaneously has a marked Complement and circumstance Theme.

However, more recent empirical studies on the functions of initial and final Adjuncts call this distinction between local and global scope into question. Verstraete (2004: 838) considers the relationship between final Adjuncts and local frameworks to be "overstated in the literature". He agrees that initial Adjuncts often serve the function of scene-setting but argues that final Adjuncts can take scope over larger stretches of a text as well. Compton (2006) even argues that the evidence between Adjunct position and discourse scope was lacking altogether and that "all adverbials at independent clause level have *potential* scope over a span broader than their host clause" (273; emphasis in the original).

One disadvantage of using frequency to assess the level of motivation between one Theme choice over the other is that it presupposes that every clause element has an equal potential of being thematized. However, it is quite obvious that an element can only be used as the Theme of a clause if that element is present in the clause to begin with. So, while it is factually accurate to say that Subjects are far more likely to be thematic than Adjuncts, this is not necessarily due to a difference in thematic potential but arguably because Subjects are more frequent than Adjuncts generally. Almost every clause contains a Subject so, by default, the speaker always has the option of positioning the Subject first. However, not every English clause contains an Adjunct (in fact only 54% do; see Section 8.1.3), which is one of the reasons why they will also occur less frequently in Theme position. These frequency differences become even grander when considering different kinds of circumstantial Adjuncts. Halliday and Matthiessen (2014) distinguish between 22 different kinds of circumstances, but there are arguably even more.⁵³ Accordingly, the likelihood of finding a circumstance Theme of Product, for example, in a given text is so small that it can be argued that Product Themes are incredibly marked, even more marked than a Goal Complement Theme for example. However, this characterization is misleading if markedness is used to assess the level of motivation of a particular Theme choice. A type of Adjunct that has a low general frequency but that, if present in the clause, is made the Theme in the majority of cases clearly does not require a lot of motivation to be thematic.

This consideration goes back to the concepts of systems and entry conditions (see Section 4.1). While Halliday and Matthiessen (2014: 134) maintain that choices concerning the different metafunctions are made simultaneously, Fawcett (2007: 42) makes a convincing argument that an experiential element has to first be present in the clause before

⁵³ Fawcett (2007: 63) claims that there are over 60 different kinds of Adjuncts and most of them carry experiential meaning.

it can receive the status of Theme. In other words, the speaker has to make the choice of including a certain circumstantial Adjunct in the TRANSITIVITY system before that element can be considered as the experiential Theme of the clause. Hence, the entry condition of the system of Guise Theme is that a circumstance of Guise is part of the experiential representation of the clause. This entry condition is met very rarely, but if it is, the choice between thematic and rhematic status of circumstances of Guise may be equally probable or even in favor of the thematic position.

In fact, the unmarked position of circumstances of Condition, for example, is often argued to be initial (see for example Downing 1991; Ramsay 1987; Ford 1993; Biber et al. 1999; Diessel 2005). Greenberg (1990: 49) even considers it a language universal that conditionals precede the conclusion in normal word order. The reason for this may be that conditionals serve as the most natural point of departure for establishing an unreal situation (Downing 1991: 139). Even though circumstances of Condition can only be found in Theme position in 6.3% of clauses (see Section 8.1.3)⁵⁴, it requires more motivation to place them clause-finally, which makes the Theme position the unmarked position for conditionals. This may be true for other Adjunct types as well, for instance causal Adjuncts (Schiffrin 1992: 193). For this reason, I suggest an alternative approach to Theme markedness, one that accounts for the thematic potential of an element only if that element is actually present in the clause. That means if a certain type of Adjunct is systematically preferred over the Subject as Theme it should also be regarded as less marked.⁵⁵ For instance, in the register of popular scientific texts, Freiwald (2016) showed that if a clause includes a circumstance of Concession or Condition, these circumstances are made the Theme of the clause in 80% and 70% of cases respectively. This suggests that, at least in the register of popular science, the default position of these circumstances is the beginning of the clause.

⁵⁴ It is important to note that this high number of circumstances of Condition is primarily due to the register of instruction manuals, which includes an above-average number of conditionals. In the other registers, the average is only 2.6%, which is consistent with the results in Freiwald (2016).

⁵⁵ Both perspectives on Theme markedness have their merits and answer different kinds of questions concerning Theme frequency. However, in the context of this study, I argue that this alternative approach is to be preferred. One goal of this project is to identify reliable predictors of Theme changes in English-German translations and Theme markedness is hypothesized to be one of these predictors, especially in translations from German to English. However, if the most natural position of circumstances of Condition in English is initial, it is implausible to assume that a Condition Theme in the German original will lead to a change in Theme structure in the English translation. Therefore, accounting for thematic potential of every circumstance in both languages seems most reasonable.

As was mentioned earlier, Complement Themes are far less frequent in English than circumstance Themes partly because there is an alternative way of thematizing experiential meaning that usually maps onto Complements. Instead of being fronted, most Complements (those that are Objects in traditional grammar) can be passivized. Voice is the resource in English that enables the variation of the mapping between Theme and participant roles (Matthiessen 1995a: 590). In active clauses, the Subject is conflated with the participant role that carries out the action or sensing, while passive constructions allow the speaker to thematize the affected participant, for instance the Goal in a material process, while still maintaining an unmarked thematic structure. And in fact, Subject Themes in passive constructions are far more common than Complement Themes in active constructions in English (6.0% to 0.8%; see Sections 8.1.1 and 8.1.2). Nevertheless, Complement Themes do exist and are used despite the possibility of a passive construction, which means that Complement Themes must serve a function that Subject Themes in passive clauses cannot serve.

Chafe (1976: 49) argues that contrastiveness is the main function of marked Themes, in particular of Complement Themes. Complement Themes are often used to establish an opposition between the Theme and another piece of information in order to cancel an assumption or to highlight differences. Furthermore, Complement Themes are also associated with the speaker's feelings and are often emotionally charged. This relation between Complement Themes and emotions is also exemplified by the fact that they often co-occur with desiderative mental processes including verbs like *loving* or *hating* (Fawcett 2007: 56).

Hawkins (1992, 2000) and Diessel (2005) argue that the main reason for thematization is processing. In general, the speaker tries to help the hearer in identifying the mood of the clause, which is why they try to avoid any ambiguity and arrive at the finite verb as quickly as possible. Speakers can override this processing principle by using marked Themes for semantic and pragmatic reason, such that have been explained above; nevertheless, processing motivations are paramount (Diessel 2005: 450-451). This explains why initial hypotactic clauses functioning as Adjuncts in the clause complex are on average shorter than final hypotactic clauses, as long Adjuncts in final position do not hinder processing (Diessel 2005: 458). The processing principle may also be one reason why Complement Themes are so much less frequent than circumstance Themes. Complements, just like Subjects, are typically nominal groups, so placing them initially can lead

to ambiguity in online processing as they may be misinterpreted as the Subject. Adjuncts can be nominal groups as well but are more likely to be prepositional phrases or hypotactic clauses, which makes them less of a processing hindrance.

In their study on the position of concessive Adjunct clauses, Wiechmann and Kerz (2013) tested, among other things, the effect of processing measures like length, complexity, and de-ranking.⁵⁶ They conclude that processing-related variables were statistically significant, so that less complex, shorter concessive Adjunct clauses were more likely to be in initial position. However, compared to semantic and pragmatic/discourse organizational measures, these process-related variables play "only subsidiary roles" (Wiechmann and Kerz 2013: 20).

5.7 Theme in other languages

Up to this point, all discussions of Theme realization were based on the English language system. However, the realization of Theme is entirely language-dependent, and Theme form in one language does not necessarily have to apply to other languages as well. Halliday's systemic functional framework was not designed to solely describe the English language but to also serve as a theory of language in general. And yet, most of the early descriptions of system networks are based on English, which bears the danger of transferring rules from English to other languages. This is particularly relevant for Theme, whose formal realization is a highly discussed topic. To gain a thorough understanding of the nature of Theme, it is fruitful to also study formal realization in languages other than English. Also, given that the thematic structure of German will be discussed in the following section, it will also be helpful to consider strategies to identify Theme in other languages.⁵⁷

In general, languages with Theme-Rheme structures can be divided into positionally marked and morphologically marked Theme languages. English would classify as a positionally marked Theme language given that Theme is always positioned at the beginning

⁵⁶ In this context, de-ranking refers to Adjunct clauses that were either non-finite or verbless, which are arguably more difficult to process than tensed Adjunct clauses.

⁵⁷ When analyzing the textual metafunction of a given language, it is important not to rule out the possibility that a language does not have a THEME system. Halliday and Matthiessen (1999) are careful to not generally assume that the existence of Theme-Rheme structures is language-universal. Instead, they assume that "[a]ll languages display some form of textual organization of the clause" (530) but what this organization looks like is language specific. For instance, Hakulinen (1989) and Baker (1992: 128) have questioned the usefulness of the concept of Theme to describe the Finnish and Arabic language systems respectively.

of the clause. Theme realizations based on sequencing is by far the most common way of identifying Theme in those languages that have already been studied in the systemic functional framework.

The Theme in Chinese is very similar to that of English as it is typically positioned before the Rheme and must include an experiential element. This experiential element can be preceded by textual and interpersonal elements (Halliday and McDonald 2004: 320). Similar to English, thematic elements in Chinese are also typically given or presumed information, while rhematic elements are typically new information (Halliday and McDonald 2004: 323). Textual Themes are less frequent in Chinese because clause links are often implicit, which follows the general Chinese tendency of using more ellipses than English (Kim and Huang 2012: 83). Moreover, Chinese clauses can include absolute Themes, which stand outside of the experiential structure. These absolute Themes can appear both in initial and final position, functioning as an afterthought. In this latter case, the Theme can follow the Rheme in Chinese (Yan, McDonald, and Musheng 1995: 244).

Theme in French is remarkably similar to Theme in Chinese. French Themes are also based on sequencing as they are typically in initial position. French also features absolute Themes which may also come at the end of the clause. While Themes in French typically represent given information, absolute Themes are usually conflated with new information, which is why they have “a marked textual status” (Caffarel 2004: 120). The most common Theme in French declaratives is the Subject, which makes any other experiential element a marked Theme in this mood (Caffarel 2004: 119). In French, the experiential Theme can include more than one experiential element by way of cliticization of participant identities (Rose 2001: 29).

Finnish has a very free word order, which is why most clause elements can be positioned virtually anywhere in the clause. Early position has been traditionally equated with Topic, but Mauraanen (1993: 98) questions whether circumstantial Adjuncts positioned at the beginning of the clause should be considered the Topic of a clause. Such fronted Adjuncts are typically used to set up a contrast, which is similar to the function of some marked Themes in English. She therefore concludes that early position in Finnish represents Theme rather than Topic (Mauraanen 1993: 100). Unlike in English, the Subject is not the only unmarked Theme in Finnish declaratives, given that the Subject is not realized in 30% of clauses and if it is, it does not necessarily occupy first position (Mauraanen 1993: 97-98).

The thematic structure of Germanic languages is particularly relevant for the aims of this thesis given that the clause structure of German is very similar to that of most other Germanic languages. Germanic languages are also positionally marked Theme languages, where early position is associated with thematic prominence. The majority of Germanic languages are finite-second languages (also referred to as verb-second or V2 languages), which means that the finite verb occupies the second position in most declarative clauses.⁵⁸ If the verb phrase is complex and the finite does not conflate with the lexical verb, the rest of the verb phrase is positioned at the end of the clause. The first position can be filled by a variety of different clause elements; therefore, there is no fixed sequence of Subject and Finite. Moreover, the pre-verbal position does not have to include an experiential element and instead can be filled by a textual or interpersonal element only. This makes the identification of Theme in V2 languages considerably more difficult because the first experiential element can be positioned both before and after the finite verb. This leaves the question open how to analyze the finite verb in case there is no experiential material in pre-verbal position.

There are two positions on this question in the state of the art on Theme structure in Germanic languages. In her studies on Theme in Norwegian, Hasselgård (1998, 2004) decides to adopt the Theme criteria of English and considers everything thematic up to and including the first experiential element. If that element comes after the finite verb, she considers the finite as part of the Theme, which she calls a structural Theme (Hasselgård 1998: 148). She describes the thematic role of the Finite in Norwegian as similar to the finite verbal operator in English polar interrogatives. If the finite verb is conflated with the process, it is analyzed as the experiential Theme of the clause. In her comparison of Norwegian and English Theme structure, Hasselgård (2004), unsurprisingly, found a higher number of process Themes and multiple Themes in Norwegian. Moreover, non-Subject Themes in first position were slightly more common in Norwegian (Hasselgård 2004: 188). Rørvik (2004) uses the same strategy for identifying Theme in her study on thematic progression in English and Norwegian. Extending the Theme up to the first experiential element is particularly relevant for thematic progression because a number of clauses may otherwise progress solely through a textual or interpersonal element in V2 languages.

⁵⁸ Interestingly enough, English, as a language with Germanic roots, represents an exception to this rule since the finite verb in English is not tied to any particular position in declarative clauses as long as it follows the Subject in declaratives.

Altenberg (1998) follows a line similar to Hasselgård's (1998, 2004) and also analyzes the finite verb as part of the Theme in Swedish in cases where the first experiential element follows the Finite. He even proposes including an obligatory Subject Theme in both English and Swedish (Altenberg 1998: 119). He distinguishes between Theme and onset, with the onset being the zone before the finite verb, equivalent to the German Forefield. While he regards the Theme in English and Swedish to be realized similarly, their onsets are very different since in English it can include multiple Theme elements before the finite verb, while Swedish can generally have only one. Any additional Theme must then be moved to the Midfield. Apart from the difference in finite positioning, the thematic structure between English and Swedish is roughly the same (Altenberg 1998: 138). Nevertheless, Altenberg regards this disruption of Theme elements by the Finite to bear communicative consequences. The point of departure in English can be signaled earlier compared to Swedish and the message can therefore be developed more quickly. Due to this positional difference, Altenberg raises the question whether the function of Theme is really comparable between the two languages or whether communicative effects of Theme are considerably different for each language (Altenberg 1998: 138-139).

An alternative approach to Theme in V2 languages is introduced by Andersen, Helm Petersen, and Smedegaard (2001) for Danish and by Steiner and Ramm (1995) and Steiner and Teich (2004) for German. In this approach, only the pre-verbal position is considered thematic and any element following the finite verb belongs to the Rheme. This means that unlike all other languages discussed so far, Danish and German would not require an experiential element in the Theme. The lines of argument for this approach will be outlined in the following section.⁵⁹

Korean is an interesting language regarding Theme realization as it seems to be in-between a positionally and morphologically marked Theme language. Korean has two morphological particles that are said to signal thematization: *un/nun* to thematize given in-

⁵⁹ For reasons of clarity, I decided to not describe other positionally marked Theme languages such as Dari (Rashidi 1992), Vietnamese (Thai 2004), Pitjantjatjara (Rose 2004), and Bai (Li 2015) in detail. In all of these languages, Theme occupies early position in the clause and requires an experiential element. In Bai, thematic potential extends up to the process of the clause (Li 2015: 530), which is similar to what Baker (1996) argues for English Theme as well. Theme structure of Pitjantjatjara is similar to that of French in the sense that Pitjantjatjara also allows multiple participants as the experiential Theme through cliticization (Rose 2001: 29). Rashidi considers the Theme in Dari to be “connected with linear development” (1992: 201) but assumes that Theme is not necessarily tied to early position. She even questions whether Theme in Dari has a consistent structural realization (Rashidi 1992: 202). If that is the case, I would in turn question whether Theme is a useful concept to describe the language system of Dari.

formation and *i/ka* to thematize new information (Kim 2007: 134-135). However, the experiential element that carries this particle is typically also in first position and some argue that it is rather this initial position and not the particle that is indicative of thematic status (for instance Sohn 1980). Kim (2007) argues that a Subject which is marked by *un/nun* and comes in first position is the unmarked Theme of a declarative clause. If a different experiential element, like a circumstantial Adjunct or a Complement, comes before the Subject and is marked by the *un/nun* particle, it is the only (marked) Theme of the clause. However, if the Subject and not the fronted marked Theme carries the thematic particle, it can be considered to have textual importance (Kim 2007: 134-135). In other words, a Subject is considered part or not part of the Theme depending on whether it is in initial position and whether it is marked by the thematic particle.

The most commonly used example when discussing different kinds of thematic realizations is Japanese, which uses the particle *wa* to indicate the experiential Theme of the clause. However, *wa* is not only used for thematic purposes but also to express contrastive meaning (Kuno 1973: 39). In general, any kind of experiential element can be marked by *wa*, as long as it has been mentioned in the present discourse, which means it is given information. Generic noun phrases and noun phrases with a unique referent do not have to be mentioned prior to be able to carry a *wa* particle (Kuno 1973: 39). The element marked by *wa* is typically positioned first, but other elements may also precede it, for example conjunctive Adjuncts, interpersonal Adjuncts, and Vocatives but also other experiential elements (Teruya 2004: 231). If such elements precede the experiential Theme marked by *wa*, Teruya considers them part of the Theme as well:

Unlike experiential constituents, neither of these textual and interpersonal elements can be thematized by means of *-wa*, but these elements that precede the experiential Theme are semantically in line with the definition of Theme given above. Because the Theme includes elements that may not be overtly marked by thematic markers, we can tentatively derive the following two recognition criteria for Theme in Japanese (cf. Hasan and Fries 1995[a], for recognition criteria):

1. whatever precedes the element marked by *-wa* or another theme marker is the unmarked Theme; but
2. if a participant or circumstance precedes the element marked by *-wa* or another theme marker, this participant or circumstance is the marked Theme.

In other words the Theme always extends up to and includes the element marked by *-wa* or another theme marker; if a structural or interpersonal element precedes the element marked by *-wa*, this is unmarked, whereas an experiential element that precedes the element marked by *-wa* constitutes a marked Theme. (Teruya 2004: 231-232)

Going by this formal realization of Theme, the *wa* particle does not signal thematization but instead just marks the end of the Theme in Japanese. If that is accurate, the most common example of a language that does not realize Theme positionally, in fact, relies heavily on early position to realize Theme. Thomson (2005: 168-170) argues that *wa* is only

thought to mark the Theme because elements with *wa* typically refer to given information, which are frequently used as points of departure.

Lastly, Tagalog, the most commonly used language in the Philippines, is another prominent example of a morphologically marked Theme language, using the particle *ang* to highlight thematicity. The most common clause element to be marked by *ang* is the Subject and given that Tagalog is a VOS language, the Theme is most commonly positioned last (Martin 2004: 280). The Theme can be positioned first, but that would constitute a marked word order (Martin 2004: 258). Tagalog is the first example where the realization of Theme is not in any way related to first position. Nonetheless, there is debate over whether the particle *ang* is in fact a Theme particle or rather a particle to signal topicality (see for instance Martin 1983).

In conclusion, Theme can be realized very differently, and the extent and content of Theme can vary tremendously between different languages. However, despite these contrastive differences, there are some general Theme tendencies, which, while not language-universal, are the same for the majority of languages discussed here. In many languages, early position in the clause is related to Theme, either as a decisive or a contributive factor of Theme identification. Most languages can have multiple elements in the Theme, which can express meaning from all three metafunctions. And lastly, the Theme of all languages discussed here have to include an experiential element. The only possible exception to this rule may be Germanic V2 languages and in particular German, which will be discussed in the next section.

5.8 Theme in German

Halliday and Matthiessen (1999) are careful about claiming that every language must have a thematic system. They assume that languages must have some resource to organize clauses textually, but this resource does not have to correspond to the English system of THEME (Halliday and Matthiessen 1999: 530). However, if a language does express thematic meaning, the general function of Theme will be comparable to that of other languages, rather than idiosyncratic. The form of Theme, on the other hand, is largely dependent on morphosyntactic and pragmatic rules and can differ noticeably between languages. Nevertheless, even Theme form seems to follow certain tendencies across different languages.

If one were to analyze the Theme realization in a language other than English, especially in the context of a contrastive study, a reasonable way of doing so would be to identify the different functions Theme can have in English and select that part of a clause in the other language that most closely resembles these functions. Such an analysis ensures that whatever is being compared, irrespective of what label it is given, represents the same functional unit.

As was discussed in Section 5.2.1, the Theme is the point of departure of the message, which means that the speakers choose a portion of the clause as the starting-point of their message, on the basis of which the rest of the clause, referred to as Rheme, is developed. The Theme is used to progress the thematic development of the text so that the selection of different Themes throughout the text is indicative of the overall development and direction of the text. Oftentimes, the Theme is used as a textual link to something already mentioned in the text to establish cohesive ties and facilitate the interpretation of the message on the basis of the context. On a more local level, the Theme represents one resource for highlighting a constituent or establishing contrast. Circumstance Themes set up a local context, which takes scope over the clause at hand and possibly over subsequent clauses as well. The Theme can be instantiated by elements from different metafunctions including textual Themes, which create textual links between clauses, and interpersonal Themes, which add meanings of mood and modality to the point of departure. Even though the realization of Theme is language-dependent, the beginning of the clause appears to be a common position for thematic meaning since arguably, there is an inherent relationship between the starting-point of a message and the starting-point of a clause. And lastly, thematic meaning is not the same as information status; hence, there is no one-to-one relationship between Theme/Rheme and Given/New.

Considering the German clause structure, the Forefield seems to be the most likely position that can realize all of these functions. The Forefield is the field before the finite verb, which means it comes first in the clause.⁶⁰ Constituents in Forefield position are not restricted regarding the kinds of meanings they can express and the Forefield can feature both given and new information. The Forefield can be used as a position to highlight or contrast elements and circumstantial Adjuncts in Forefield position can establish a cir-

⁶⁰ Technically the left outfield precedes the Forefield, but for the purposes of the German Theme, the left outfield will be considered a part of the Forefield.

cumstantial framework similar to English. Generally, the Forefield can be filled with a variety of different constituents from different metafunctions, which shows that the speaker is, for the most part, free to choose any point of departure. One of the main functions of the Forefield is to create a link between the local clause and its context either topically or textually (Weinrich 1993: 64).⁶¹ For these reasons, the Forefield seems to correspond functionally to the Theme in English and Steiner and Ramm (1995), Teich (2003), and Steiner and Teich (2004) argue that the Forefield represents the German Theme. Andersen, Helm Petersen, and Smedegaard (2001) make the same argument for Theme in Danish.

However, as was shown in Section 3.3, the Forefield is usually restricted to only one constituent, which can be just a single textual or interpersonal element (see example (58); textual Theme in bold). In that case, the Theme would not have any relationship to the experiential realm and the point of departure would be purely textual or interpersonal.

- (58) ***Ausserdem*** können Sie auf PREMIERE DIREKT die gewünschte Startzeit für Ihre Filmbestellung auswählen.
'**additionally** can you on PREMIERE DIREKT the preferred starting.time for your movie.order choose.'

[G2E_INSTR_003]

In this respect, German and other Germanic languages would be fundamentally different from all other languages discussed in Section 5.7, which all require an experiential Theme. This is especially problematic in respect to thematic progression, which is arguably one of the most central functions of Theme. In the case of a single textual or interpersonal Theme, the thematic progression would be interrupted and would not develop further for potentially longer stretches of texts. It could be argued that the focus is instead on the textual and interpersonal progression of the text; however, the development of the text as a representation of experience clearly does not pause merely because the Forefield is instantiated by a non-experiential element. If the Forefield was truly the equivalent of the English Theme, this would mean that German text development was much less about

⁶¹ Weinrich's (1993: 64) exact quote is: "Das Vorfeld erscheint daher insgesamt als dasjenige Feld, das den textuellen Anschluß an das gegebene Vorwissen garantiert. Das kann insbesondere ein thematischer oder ein argumentativer Anschluß sein." Even though he uses the term *thematisch* (*thematic*), he is not referring to the Hallidayan notion of Theme but rather thematic meaning from the Prague school perspective.

experiential representation and a lot more about textual ties and interpersonal assessments.

This raises the question why to limit German Theme to the Forefield and not search further in the clause for the first experiential element to exhaust the thematic potential. If the Forefield is not conflated with an experiential element, the first constituent in the Midfield, following the Finite, carries experiential meaning in most cases. As was discussed in Section 3.4, this element is usually the Subject but can also be a circumstantial Adjunct or a Complement. Accordingly, the German Theme could be realized the same as in English as everything up to and including the first experiential element, which is sometimes positioned in the Forefield and sometimes in the early Midfield. Steiner and Teich (2004: 172-173) consider this possibility but ultimately decide against the Midfield as a suitable position for a thematic element:

Our description of THEME draws on accounts in Engel (1988), Heidolph, Flämig & Motsch (1981) and Weinrich (1993), as well as, in particular, Hoberg (1981) and Erdmann (1990a, b). In terms of these accounts, we are dealing with *Vorfeld* rather than with *Thema*. That is to say, we will not follow suggestions to assume a functional boundary between *Thema* and *Rhema* in German after the first post-verbal position in indicative clauses, although that seems to be the borderline between "identifiable" and "non-identifiable", as evidenced by the fact that this first post-verbal position is the default place of realization for identifiable discourse referents, such as pronouns and definite phrases (cf. clauses (1), (2), (3) and (5) from Text 3.4 and clause (5) from Text 3.5) above. This, however, seems to be a question of informational, rather than of thematic meaning.

They argue that the first post-verbal position, that is, the first Midfield constituent, is not the appropriate borderline between Theme and Rheme given that this constituent is typically identifiable and definite, which is, loosely speaking, the same as Given. They, thus, argue that this first post-verbal position is chosen based on informational and not thematic considerations, which makes the first Midfield position not a Theme position given that THEME and INFORMATION, while interrelated, are two entirely separate systems in SFL.

It was shown in Section 3.4 that the constituent order of the Midfield is influenced by a number of different criteria, one of which is identifiability/definiteness. Steiner and Teich (2004) are therefore right in claiming that the first post-verbal element often represents identifiable information. However, it was also shown that identifiability is not the only criterion for constituent sequencing in the Midfield and that the default order can usually be altered if the context motivates this change. In example (59), an indefinite nominal group is positioned before a definite nominal group in the Midfield likely because it functions as Subject.

(59) *Beispielsweise könnten gerade aktive Programme oder ein von dieser Partition gestartetes Betriebssystem die Ursache für ein in Verwendung befindliches Dateisystem sein.*

'for example could currently active programs or a from this partition started operation.system the cause for an in usage being file.system be.'

[G2E_INSTR_007]

This discussion of Theme in German is closely related to the question of the relationship between meaning and choice. One of the central principles of SFL is that language is made up of systems which allow the speaker to make choices between different options that stand in a paradigmatic relationship and to create meaning by virtue of these choices. This is exemplified by title of one section in Lyons' Introduction to Theoretical Linguistics: "Having meaning implies choice" (1968: 413). Conversely, any time a speaker cannot choose between two or more options and is 'forced' to use language in a certain way, their decision to follow these language rules is not a meaningful choice. For example, finishing an orthographic sentence with a punctuation mark in a language like English or German is not a choice speakers have to make individually but is dictated by the general rules of that language. The speaker can control the length of the sentence and they can control which punctuation mark is most appropriate in a given context but the decision to use punctuation to mark the boundary between two sentences in written discourse is generally not up to the speaker. As is the case with any rule, the speaker can simply decide to ignore it, but they then run the risk of being ungrammatical or incomprehensible. Some 'choices' in language are dictated by the language system and choices in THEME are no exception to this. Arguably, an element can only be considered thematic if it was deliberately chosen as Theme to function thematically. In other words, if a speaker does not have control over the position of a constituent and is obliged to place it in the Theme, it does not carry thematic meaning even though it comes early in the clause. As Downing concisely puts it, "Theme is a meaningful choice" (1991: 122).

In his earlier works, Halliday (1967b: 220) is adamant that a Theme is only meaningful if it is a result of choice:

The concept of theme, like the other options under discussion, is based on the notion of choice: it represents an option on the part of the speaker, and any clause can be regarded as being in contrast with one or more others differing from it just in the selection of the theme. An item occurring obligatorily in initial position will not, in this sense, be thematic.

And yet in Halliday and Matthiessen (2014), this relationship between Theme, meaning and choice is not explicitly stressed any longer. The reason for this is likely that Halliday and Matthiessen (2014) do consider a variety of units thematic even though their status

as Theme is not a result of choice, for example conjunctions and the finite verbal operator in *yes/no*-interrogatives. Halliday and Matthiessen (2014: 109) argue that some units are inherently thematic and through language development have moved to the front of the clause because of their predisposition as points of departure. Similarly, intransitive clauses without circumstantial Adjuncts also feature an experiential Theme that is not based on choice but on the fact that there is no other element that is capable of being thematic (see example (60)). It is of course the choice of the speaker to use an intransitive verb and to not include circumstantial information, but once these experiential choices are made, the choice for the Theme is predetermined.

(60) *My mother shrieked.*
[E2G_FICTION_008]

The reason why this discussion of Theme and choice is so meaningful for Theme in German is that Steiner and Teich's (2004) argument against the post-verbal element as the boundary between Theme and Rheme rests on the fact that the system of INFORMATION heavily influences constituent order in the Midfield. This raises the question of how freely the speaker has to be able to choose between different Theme options so that this choice can be considered a meaningful Theme choice. Admittedly, there are some sequences where the speaker has no choice whatsoever regarding constituent order in the Midfield, for example in the case of the nominative personal pronoun, which, if not positioned in the Forefield, always comes first in the Midfield (see example (61); nominative personal pronoun in bold).

(61) Außerdem habe **ich** ihm nichts versprochen.
'furthermore have **I** him nothing promised.'

Considering the relationship between meaning and choice, such a Midfield unit should not be considered the Theme of the clause. But if that is the case, the same should also hold true for all English 'Themes' that are not the result of the speaker's choice.⁶² In all other cases, the Midfield word order is arguably heavily influenced by other systems, such

⁶² In fact, Fawcett (2007) makes exactly this argument. He does not consider any element thematic in English which is not positioned early in the clause as a result of thematic choices. This includes the aforementioned conjunctions and finite verbal operators as well as relativizers, preposed Themes and fronted paratactic reported clauses.

as information status, clause function, animacy, heaviness, and others. Then again, it is not uncommon for the choice of Theme to be dependent on choices in other systems. As was discussed in Section 5.6, the probabilities of particular Theme elements in English are also heavily dependent on MOOD. Moreover, while the structure of the German Midfield is influenced by a variety of different factors, in most other cases, there is a sequencing choice involved. This is particularly obvious when one combination of constituents can be ordered differently in the German Midfield. One of these sequences represents the unmarked word order, and yet the speaker can choose a more marked order to highlight or contrast an element, to set up a local context or to develop the thematic progression. In example (62), the dative pronoun (in bold) is the most likely constituent to head the Midfield and yet other Midfield sequences are also possible in the right context. These functions of the first post-verbal element are clearly comparable to what is described as the function of Theme in English and other languages.

(62)

- a. *Zu meiner Überraschung hat **mir** eine Fremde den Rucksack zurückgegeben.*
'to my surprise has **me** a stranger the backpack given.back.'
- b. *Zu meiner Überraschung hat eine Fremde **mir** den Rucksack zurückgegeben.*
'to my surprise has a stranger **me** the backpack given.back.'
- c. *Zu meiner Überraschung hat den Rucksack **mir** eine Fremde zurückgegeben.*
'to my surprise has the backpack **me** a stranger given.back.'

It is for these reasons that I am arguing for an alternative approach to Theme in German, namely that the German Theme zone contains everything up to and including the first experiential element. This Theme realization of German quite obviously mirrors that of English. However, I am not making this argument because it is easier or more practical to use the same formal realization for both languages but because I believe that clause zones best capture the same thematic functions of highlighting, contrasting, contextualizing, linking, and thematically developing. This also avoids the question why German and V2 languages in general are so different from all other languages in that their points of departure do not require any relationship to the experiential realm.

And yet, it must be noted that the thematic structure of Germanic V2 languages is still quite unique even if it includes an obligatory experiential element. System networks are probabilistic in nature, so that different options in a system are always more or less likely than others. In English, the probabilities in the system of THEME change depending on the

decision made in other systems, mainly MOOD but INFORMATION as well. This is the case for German, too, but the impact of other systems varies depending on whether the first experiential Theme is pre-verbal or post-verbal. To illustrate, the choice of experiential Theme in English, both in terms of its role as a clause element and its identifiability, is not affected by the decision to include or not include a textual Theme, which is the whole reason why they are argued to not be part of the same Theme markedness network in English (Halliday and Matthiessen 2014: 110-111). In German, on the other hand, the experiential Theme is more likely to be identifiable, and probably more likely to be the Subject, if the pre-verbal position is occupied by a textual or interpersonal Theme. Hence, by choosing a textual Theme and moving the first experiential element to a position after the Finite, the probabilities between the different Theme options change. If one decides to treat the German Theme like the English Theme and go up to and include the first experiential element, this discrepancy needs to be taken into account.

Changing the formal description of Theme in German leads to yet another interesting difference, one which Hasselgård (1998, 2000, 2004) also addresses, when discussing Theme in Norwegian. If the first experiential Theme is frequently positioned after the finite verb and everything before that experiential element is part of the Theme, the finite verb will frequently be included in the Theme of V2 languages. The Finite does not contribute much to the point of departure of the message. It is very similar to the finite verbal operator in English (and also German) *yes/no*-interrogatives and should therefore be treated as an extra interpersonal Theme, which is what Hasselgård (1998: 148) decided to do for Norwegian.⁶³ Nevertheless, one problematic case potentially arises given that the Finite in German, just as in English, can be conflated with the Predicator in the case of simple aspect. Consequently, if a non-experiential Theme occupies the Forefield and the verbal unit of the declarative is in simple aspect, the first experiential Theme is the process. To illustrate, example (63) is experientially the same but in a. the aspect is simple while in b. the aspect is perfect (process in bold).

(63)

- a. *Und dann **ging** ich nach Hause.*
'and then **went** I to home.'
- b. *Und dann bin ich nach Hause **gegangen**.*
'and then have I to home **gone**.'

⁶³ Hasselgård calls the Finite in Norwegian a "structural Theme" (1998: 148) but draws the comparison to the English finite verbal operator, which suggests that she is referring to an interpersonal Theme.

One potential solution to that problem is to analyze the process as the experiential Theme given that it is the first experiential element of the clause. However, the speaker does not position the process early for thematic reasons but simply because it is mapped onto the Finite and the syntactic rules of the language require the Finite to come in second position. This again relates to the issue of meaning and choice, and since there really is no choice involved at all, the process does not seem to be a meaningful experiential Theme.

In her analysis of English-Norwegian translations, Hasselgård (2000) decides to treat the conflation of the Finite and the process as an experiential Theme and as a consequence not only finds a substantially higher number of process Themes in Norwegian but also various instances of translation shifts, where the experiential Theme of the English original is turned into a process Theme in the Norwegian translations. Hasselgård (2000: 24) does point out that these shifts are just a result of different syntactic rules and that a reasonable alternative would be to treat the fronted verb as an interpersonal Theme as well.

The same problem arises in English if the process of a *yes/no*-interrogative is instantiated as a form of *be* in simple aspect as in example (64a). In his criticism of Halliday (1985), Huddleston (1988: 161) raises this issue and questions whether Halliday believes that the point of departure in this case is the process.

(64)

- a. *Isn't the best idea to join the group?*
- b. *Wouldn't the best idea be to join the group?*

In their reply to Huddleston (1988), Matthiessen and Martin (1991: 48) argue that the finite *be* is not positioned early for experiential reasons but for interpersonal reasons, namely to signal mood, and thus should also be treated as an interpersonal Theme:

Halliday's point is that *wouldn't* is thematic as Finite, i.e. from an interpersonal point of view. But Huddleston seems to have missed this point entirely (cf. for example the thematic analyses on p. 48 of IFG; see also p. 56) for he complains that *isn't the best idea to join the group* 'would have a completely different textual structure, for here *isn't* constitutes the whole of the process and hence would be topical Theme'. In fact, the two examples would have the same thematic structure, as shown in Figure 13.

To re-iterate, it is Finite that is selected as Theme – process is not selected as Theme (contrast the thematic process in *he said he would run and run he did*). Since Finite is co-extensive with process, the latter will also be initial in the clause, but only due to its role as Finite. Consequently, it does not constitute the topical part of the Theme. (Emphasis in original)

Huddleston (1991: 96) does not consider this a satisfying solution given that the process does fulfill the requirements for being the experiential Theme of the clause irrespective

of its status as the Finite. Personally, I side with Matthiessen and Martin (1991) in this case since the process was not fronted so that it acts as the point of departure of the message. The process was put in first position because given the mood of the clause and the aspect of the verb, the speaker had no other choice. This lack of choice should prevent the process in this example from being treated as a meaningful Theme.

The same holds true for German and other V2 languages. A verbal group that conflates Finite and process has to be in second position in declarative clauses not because of THEME but MOOD. German declaratives can have the process as their actual experiential Theme, which represents a very marked, uncommon construction in German (see example (65); process in bold).

- (65) **Kaufen** kann man sie auf der Mülkerstiege in einem der schönsten Geschäfte Wiens.
'buy can you them on the Mülkerstiege in one of the most beautiful stores of Vienna.
[G2E_TOU_022]

If these rare cases were lumped together with the common case of Finite-process conflation in second position, it would diminish the significance of deliberate Predicator Themes.⁶⁴ This separation of the verb's role as the process and its role as the Finite is particularly important for the analysis of translated texts, since otherwise, thematic shifts in German-English translations involving process Themes would occur frequently simply because the translator followed the syntactic rules of the target language. Such cases must be kept separate from deliberate changes to the thematic structure.

Still, it must be noted that the systemic functional state of the art is somewhat selective in what constitutes a meaningful Theme and what does not. The *wouldn't* in example (64b) is also not positioned initially by choice and should therefore not be treated as a meaningful thematic unit, irrespective of whether it has interpersonal or experiential meaning. Halliday and Matthiessen (2014) argue on the basis of thematic inherence, the same argument they use for conjunctions as textual Themes. Notwithstanding, if meaning presupposes choice, finite verbal operators are thematically not meaningful. It may very

⁶⁴ As a counterargument, the speaker may have also deliberately chosen the simple aspect of the verb so that the process does come early in the clause and act as a point of departure without being particularly marked. This argument can only hold true for the difference between Präteritum (simple past) and Perfekt (present perfect) given that there is no progressive aspect in standard German and Präsens (simple present) is the only common verb form to refer to the present. While I do not consider this a convincing argument, since I cannot imagine that decisions regarding tense and aspect in German are influenced by thematic considerations, I can also not prove that this is not the case.

well be that such elements developed to stand at the front of the clause because of their predisposition as point of departure, but with that development they arguably also lost their thematic meaning. The most consistent strategy of analyzing Theme is outlined by Fawcett (2007), where a unit is only considered thematic if the speaker had a choice in the THEME system to position the unit at the beginning of the clause. Everything else is not considered the Theme irrespective of its position in the clause (Fawcett 2007: 72). This arguably poses a great challenge for empirical analyses of Theme but may result in a more plausible account of thematic structure.

Even though German is considered to have a relatively free word order (Hawkins 1986: 44), Steiner and Teich (2004) still distinguish between marked and unmarked German Themes. As was discussed in Section 5.6, Theme markedness can be determined in different ways. Halliday (1967b: 219) argues that the word order that needs to be motivated the least by context represents the most unmarked order, and its first element is the unmarked Theme. However, the problem with the degree of motivation is that there is no reliable way of measuring it, which is why frequency is often used as a proxy (Haspelmath 2006: 45). One other possibility to identify marked word order and marked Theme elements is considering intonation in verbal communication, as more marked Themes also lead to an infrequent intonation pattern (Steiner and Teich 2004: 178).

There is dispute in the state of the art on which Forefield elements can be considered unmarked in German declaratives. Erdmann (1990b: 74) claims that only the Subject represents an unmarked Forefield element, as it is the only constituent that does not need textual motivation to be positioned early. Steiner and Teich argue that German has "a relatively weak notion of markedness" (2004: 169) given that in declarative clauses, the Theme cannot only regularly be conflated with the Subject but also with accusative Complements (see example (66)) and dative Complements (see example (67); Complements in bold) if they take on the role of the first participant of the process.

(66) ***Mir** ist klar, dass hier spezifische Gegebenheiten der Londoner City vorliegen [...].*
'***me**[dat] is clear that here specific factors of the London City exist [...].*'

[G2E_SPEECH_002]

(67) *Da erschreckte **mich** die Frage, was Catherine zu meiner fixen Idee sagen würde.*
'*then frightened **me**[acc] the question what Catherine to my fixed idea say would.*'

[G2E_FICTION_001]

The Theme can also be realized as a circumstantial Adjunct in German declaratives (see example (68); circumstance Theme in bold). While Steiner and Ramm (1995: 80) only mention spatio-temporal circumstances as unmarked Themes in German, Steiner and Teich (2004) list circumstances in general as part of the unmarked Theme category. Akin to English, textual and interpersonal elements are also generally unmarked. However, they consider Complements that do not represent the first participant of the process marked Themes (Steiner and Teich 2004: 177; see example (69); Complement in bold).

(68) ***In Deutschland*** haben wir bisher noch keine Entscheidung über die Einführung von REITs getroffen.
'in Germany have we so far no decision regarding the introduction of REITs made.'
 [G2E_SPEECH_002]

(69) ***Die Ein-/Aus-Taste des TV-Gerätes*** finden Sie hinter der Abdeckung des Bedienteils.
'the on-/off-button of the television find you behind the cover of the controls.'
 [G2E_INSTR_009]

In a study of Theme structure in English-German translations of popular scientific texts, Freiwald (2016) argues that circumstances cannot simply be classified as marked or unmarked Themes in general but have to be assessed on the basis of the kind of circumstantial meaning they express. He found that in German originals, circumstance of Place and circumstances of Means were used as Themes in 9.7% and 7.2% of all clauses respectively, making them the most common circumstance Themes in that register. Nevertheless, Freiwald (2016) also showed that these circumstances were among the most common circumstances in the register generally and were thus also more likely to be thematic due to their high overall frequency. As was argued in Section 5.6, an alternative way of assessing the markedness of individual Theme types is to determine their frequency as Themes relative to their overall frequency. While circumstances of Purpose and Matter were not used as frequently overall, they had a likelihood of 70% and 60% respectively to be the Theme in German if they were included in the clause. Despite being relatively common in the register overall, circumstances of Comparison were made the Theme in only 10% of cases, suggesting that the Forefield is a highly marked position for circumstances of Comparison (Freiwald 2016: 47). With a frequency of 14.8%, Complement Themes were surprisingly common in German original scientific writing, which leaves room to question their degree of markedness. In fact, Kirkwood (1970: 104) considers all

Complement Themes unmarked as long as they are identifiable. Freiwald (2016) did not control for identifiability or definiteness.

In imperatives, the number of unmarked Themes is reduced to the Predicator and in *wh*-interrogatives, the *wh*-element is commonly positioned first in the clause. Polar interrogatives commence with the finite verb, which is conflated with the process in simple aspect, followed by the Subject (Steiner and Teich 2004: 179).

Similar to Theme markedness, the frequency of multiple Themes in German is also entirely dependent on the manner in which Theme is analyzed. According to Steiner and Teich (2004), multiple Themes are fairly restricted in German given the strong Finite-second constraint. The only common case of multiple elements before the finite verb includes constituents that according to Brinkmann (1971) are part of the left outfield, for example conjunctions and Vocatives (see example (70); Vocative underlined, Subject Theme in bold).

- (70) Herr Präsident, **kooperative globale Sicherheit** wird sich an dem ihr gesetzten verbindlichen Rechtsrahmen messen müssen.
'Mr. President, **cooperative global security** will [refl-3sg] by the its set obligatory legal.framework measure must.'
[G2E_SPEECH_003]

As was shown in Section 3.3, the Forefield can host more than one experiential Theme, but such constructions are extremely rare. Also, the Forefield cannot cover Themes of all three metafunctions (Steiner and Teich 2004: 174), unless one allows for parenthetical insertions.

This changes, of course, if the Theme is not limited to the pre-verbal position and potentially extends to the Midfield. In that case, a clause has a multiple Theme whenever the Forefield is occupied by a non-experiential element. This means that multiple Themes are not only more common but can also include elements of all three metafunctions. Yet, the combination of textual Theme, modal Adjunct Theme, and experiential Theme is restricted in German and is only probable if the textual Theme is a conjunction and not a conjunctive Adjunct (see example (71)). Conjunctive Adjuncts take up the entire Forefield position, and the first Midfield position is usually under pressure to feature an element from the TRANSITIVITY system, especially if the Subject or Complements are identifiable.

- (71) *Und | wie bereits erwähnt, | können | Festplattenbackups zur Zeit nur komplett wiederhergestellt werden.*
'and | as already mentioned | can | hard.drive.backups at.the moment only completely recovered be.'
[G2E_INSTR_007]

If the German Theme extends up to the first experiential element, the number of interpersonal Themes consequently increases as well, since most multiple Themes include the finite verb, which has an interpersonal function. Nonetheless, as discussed earlier, the position of the finite verb is fixed and its inclusion in the Theme is unavoidable if the Forefield is filled with a non-experiential element. Its significance as a thematic element should therefore not be overstated.

5.9 English-German comparison of Theme

Theme is a very abstract functional concept and, as a consequence, difficult to identify formally. It is therefore not surprising that there is disagreement over the extent of Theme in both English and German. In both languages, Theme is realized through position by coming early in the clause. In English, the two most commonly used Theme hypotheses are the first experiential element hypothesis, first suggested by Halliday (1967a/b), and the Subject hypothesis (Enkvist 1973; Downing 1991; Ravelli 1995, and others). In German and other Germanic V2 languages, the Theme is argued either to be equivalent to the Forefield (Steiner and Ramm 1995; Andersen, Helm Petersen, and Smedegaard 2001; Steiner and Teich 2004) or to extend to the first experiential element (Hasselgård 1998, 2000, 2004; this thesis). Naturally, a comparison of the thematic structure of English and German depends entirely on which Theme description is favored in each of the languages, which is why a contrastive Theme analysis is such a complicated undertaking. Nevertheless, there are some general Theme similarities and differences that can be observed without assuming a particular Theme description.

In both languages, the Theme can feature elements from all three metafunctions, experiential, interpersonal, and textual Themes. The different sub-categories of textual and interpersonal Themes are identical in English and German, namely conjunctions, con-

junctive Adjuncts, and continuatives, and modal Adjuncts, Vocatives, and finite verbal operators. Continuatives, conjunctions, and Vocatives are noteworthy in German as they do not exhaust the Forefield and allow further elements in preverbal position in V2 constructions. Incidentally, these are the same elements that are inherently thematic in English.

The first participant of the process is a common experiential Theme in both English and German. German allows the first participant to be mapped onto a Complement in mental processes, but such constructions are fairly rare (3.1% of mental clauses and 0.4% of all clauses; see Section 7.1.2). Generally, the first participant of the process is also the Subject of the clause, which makes Subject Themes a frequent choice in declarative clauses for both languages.

Theme markedness in imperatives and interrogatives appears to be the same between English and German. The experiential Theme in imperatives is most likely instantiated as the Predicator.⁶⁵ Both languages have a *wh*-interrogative and the *wh*-element is the most common experiential Theme in this mood. The word order of polar interrogatives is also the same, with the finite verb being positioned first, followed by the Subject. It is more common for the finite verb to be conflated with the process in German polar interrogatives because German, unlike English, does not require an auxiliary verb to form a *yes/no*-question.

One of the main contrastive differences between English and German constitutes Theme markedness in declarative clauses. In English, the Subject represents the only unmarked experiential Theme in declaratives. While circumstantial Adjuncts are generally considered less marked than Complement Themes or Predicator Themes, they are still considered marked Themes. In German, Theme or rather Forefield markedness is a more disputed issue. Four different positions on the matter of Forefield positioning are recognized in the state of the art: Only the Subject is an unmarked Forefield constituent (Erdmann 1990b), Subjects and spatio-temporal circumstances are unmarked (Steiner and Ramm 1995; Teich 2003), Subjects and all kinds of circumstances are unmarked (Steiner and Teich 2004), and Subjects, circumstances, and Complements are all unmarked (Kirkwood 1970; Zifonun, Hoffmann, and Strecker 1997). As an alternative approach, Freiwald (2016) does not consider the entirety of circumstances marked or unmarked Forefield

⁶⁵ German features some imperative constructions that do not exist in English, for example the impersonal imperative: *Einfahrt freihalten* (gloss: 'entry clear.keep'). See Steiner and Teich (2004) for a detailed discussion.

Themes but analyzes them separately according to the experiential meanings they express. Freiwald's (2016) analysis shows that there are both marked and unmarked circumstantial Themes in German as well as in English.

Regardless of whether circumstances and Complements are marked or unmarked Themes in German, their frequencies as Forefield Themes are generally higher in German than in English (Freiwald 2016: 45). Thus, while Theme markedness as a binary category may be the same in both languages, the degree of markedness of circumstance, Complement, and even Predicator Themes, based on frequency, is still different. Inflectional morphology signals many grammatical relations in German, which is why position and sequencing of clause element is not as central to express grammatical meaning. As a consequence, the word order is a lot freer in German than in English, so that almost any constituent can be positioned in the Forefield. Also, the German Subject does not have "a high functional load for expressing Mood" (Steiner and Teich 2004: 180) and it is instead the position of the finite verb which distinguishes between V1, V2, and Vlast constructions. This is also the reason why the Subject can easily occupy a post-verbal position and a different element can take its place as the point of departure of the message. The Subject in English, together with the Finite, makes up the Mood and therefore has to be in a fairly fixed position in the clause. While other elements can still occupy a pre-Subject position, the Subject generally comes rather early in English declaratives, which makes English's Theme structure more syntactically motivated than the Theme structure in German.

The frequency of multiple Themes and the average number of Theme elements is highly dependent on which formal realization of Theme is used. Quite obviously, the Subject hypothesis will produce more Theme elements than the first experiential element hypothesis given that the Theme potentially extends to later constituents in the clause. Similarly, the number of German Themes increases if the early Midfield is also considered potentially thematic. If the two more traditional approaches to Theme are used, namely the first experiential element hypothesis for English and the Forefield hypothesis for German, the frequency of multiple Themes and the average number of Themes is higher in English since German Themes will usually be restricted to one element. In example (72), the early elements in the German translation represent literal translations of the Theme elements of the English original. However, due to the Finite-second constraint the German Theme only consists of a single textual Theme while the English Theme features a

textual and the obligatory experiential Theme (textual Theme underlined, experiential Theme in bold).

(72)

EO: For example, **you** might want to prevent your children from seeing Web sites that contain violent or sexual content.

GT: Beispielsweise möchten Sie Ihre Kinder vor Websites mit Darstellungen von Gewalt oder sexuellen Handlungen schützen.

'for.example might you your children from websites with portrayal of violence or sexual conduct protect.'

[E2G_INSTR_009]

If the same hypotheses are used for both languages, i.e. either the first experiential element hypothesis or the Subject hypothesis, the average number of Theme elements will be higher in German due to the extra finite verbal operator that will be part of most multiple Themes. However, apart from this extra, obligatory Theme element in German, there is no apparent reason to assume that there is a difference in Theme elements in the two languages. If there was, it would mean that one language systematically uses more textual or interpersonal Themes than the other, which might be the case but to my knowledge has not been demonstrated so far. For that same reason, the number of multiple Themes, meaning Themes that consist of more than one constituent, is not assumed to be different in the two languages if Theme is analyzed up to the first experiential element in both languages.

One contrastive difference between English and German that is not a thematic aspect per se but closely linked to Theme is that of Subject agency or sentience (see Section 3.5). Because of the relatively fixed word order, the mapping of semantic meaning and grammatical function is less restricted in English to allow for different semantic sequences without having to use a marked word order. For this reason, inanimate, non-sentient Subjects can be paired with processes that require the first participant to be sentient in English, referred to as non-sentient constructions. While such non-sentient constructions are generally not ungrammatical in German, they are less frequent because the German word order allows for a multitude of unmarked word orders, resulting in the possibility that these inanimate entities can come early in the clause without being conflated with the grammatical role of Subject. One common translation procedure of placing an inanimate entity in the Forefield position is by conflating it with an Adjunct (see example (73); Subjects in bold, circumstance underlined).

(73)

EO: *Arrow down opens a selected menu.*

GT: *Mit der Abwärtspfeiltaste wird ein bereits ausgewähltes Menü geöffnet.
'with the arrow.down-button is an already selected menu opened.'*

[E2G_INSTR_005]

One potential thematic difference between English and German is that of Theme length. As Hawkins (1992: 215) shows, the meaning of a message is more easily processed if the hearer can identify the mood quickly. So, if an English speaker wants to maximize comprehensibility in a declarative clause, they may want to arrive at the Subject followed by the finite verb as quickly as possible and leave the Theme relatively short. In German, on the other hand, mood is primarily signaled by the position of the finite verb. In declaratives, which is a V2 construction in German, the speaker will also want to arrive at the finite verb quickly but if the experiential Theme is in postverbal position, length should not be of any issue from a processing standpoint. Therefore, Theme length of English Themes and German Forefield Themes may be similarly short, whereas German Midfield Themes could on average be longer. Then again, the first Midfield position is the most likely place for personal and demonstrative pronouns, which are made up of only one word and few characters. This theoretical freedom in Theme length might be curtailed due to the interference of part-of-speech requirements of the Midfield.

5.10 Theme in translations

While Theme is a prevalent research topic in the systemic functional community, it has not been in focus in many other linguistic frameworks and, as a consequence, has received only little coverage in translation studies outside of SFL. Lately, the interest in Theme and textual meaning has increased in translation studies after it had been largely overlooked in favor of experiential meaning. Kim and Matthiessen (2015: 336) assume that this is primarily due to the subtleness of textual meaning and the focus of the translator on the meaning of words in sentences rather than on the meaning and structure of the text as a whole.

This neglect of textual meaning in translation studies is surprising given that the textual development is central to the general meaning of a text. Only if a translator understands the organization of a text in the source language, will they be able to "re-create the

same message in another language" (Vasconcellos 1985: 27). Translators may be intuitively aware of differences in thematic structures between source and target language (Lyons 1977: 510) and may be able to adjust their translations accordingly. And yet, textual meaning differences can be very subtle, which is why the analysis of Theme in contrastive and translation studies promises to be quite illuminating.

Translating a text from one language to another is a highly complex activity since the translator has to produce an appropriate translation that ideally captures the meaning on all three metafunctional levels (Kim and Matthiessen 2015: 335-336). In the source language, each metafunction has its own set of systems, which may or may not have equivalent systems in the target language. Depending on the language pair, the translator also has to prioritize between different kinds of meaning. For example, in translations between Arabic and English, an appropriate translation of interpersonal meaning is more important given that Arabic is a Verb-first language. In this case, thematic equivalence has to be largely disregarded, since this would result in an ungrammatical translation (Baker 1992: 128). In fact, grammaticality and syntactic rules are two of the main reasons to deviate from thematic patterns (Vasconcellos 1985: 59)

When translating the thematic structure of the source language, Baker (1992: 128) recognizes three different possibilities:

1. The translator can preserve the thematic pattern of the original, which allows them to stay true to the method of development.
2. The translator cannot preserve the thematic pattern of the original without distorting the target text, in which case an alternative has to be found.
3. The target language does not express thematic meaning, in which case Theme is not a useful category to consider.

In case a translator is not able to preserve the thematic structure as a whole, Baker (1992: 167-171) also suggests a number of different translation procedures which allow the translator to resolve tensions between word order and information structure, for example voice changes, nominalizations, and extra-positioning. One possibility that Baker does not consider explicitly is the case where the translator can generally preserve the thematic pattern of the original, but this pattern represents a more marked pattern in the target language. In this case, the translator has to balance thematic meaning of the original on the one hand and target language appropriateness on the other hand. In this context, Munday (1998: 188) distinguishes between meaningful choices, where the translator deliberately deviates from the thematic structure of the original, and non-meaningful choices, where the translator simply obeys the grammatical rules of the target language.

The state of the art on Theme in translations has largely focused on similarities and differences regarding thematic progression (see for example Hasselgård (1998) and Rørvik (2004) for English-Norwegian translations; Ghadessy and Gao (2000) and Liu and Yang (2013) for English-Chinese translations; and Vasconcellos (2008) for English-Portuguese translations). These studies produced mixed results, as for some translation pairs, thematic progression was largely preserved, while in other pairs, the thematic progression in the translation deviated from the original considerably. Kim and Matthiessen (2015: 339) assume that this discrepancy arises from the fact that some language pairs were typologically similar and allowed for the preservation of thematic progression, whereas other pairs did not.

Alternatively, studies like Hasselgård (1997, 1998, 2000) look at individual sentence pairs of original texts alongside matching translations and examine possible changes in thematic structure without explicitly considering the progression in the text. In translations between English and Norwegian, Hasselgård (1997) shows that the thematic structure of English and Norwegian original texts is noticeably different, but that this difference is minimized in the translations. In other words, translators try to match the thematic structure of the originals as much as possible and as a consequence, the Theme distribution of the translations is more similar to that of the source language than the target language. This effect was stronger for translations from English into Norwegian (Hasselgård 1997: 18). Similar results are presented in Hasselgård (2000), where she compares English originals with Norwegian and German translations. Hasselgård (2000) found that the content of the experiential Theme as well as textual and interpersonal Themes is generally preserved in the translations of both languages. Norwegian translators are more inclined to keep the experiential Theme intact, whereas the German translators more often preserve the very first Theme element, irrespective of whether it is an experiential, interpersonal, or textual Theme. According to Hasselgård, this "may indicate slightly different priorities as to initial vs. topical theme" (Hasselgård 2000: 36) between the two languages.

Neumann (2003) studies, among other things, Theme in English and German original as well as German translated tourist guides. She found that the Theme distribution in the German translations is generally in between that of German and English original texts but correspond more closely to German originals. Teich (2003) reports a difference in Theme

markedness between English and German originals, which also carries over to the translations, falling somewhere in the middle of the originals. Even though the differences between originals and translations in both languages were not statistically significant, Teich (2003) still attributes this effect to shining through and normalization. Corresponding to previous studies, Neumann (2014) also discovers contrastive differences between English and German originals and their translations in the registers of fictional writing and letters to shareholders. Moreover, she demonstrates that the registers do not only influence thematic distributions but also affect the extent to which translation properties such as shining through and normalization can be found. While Themes in translations of fictional texts show more traces of the source texts, the Themes of translated letters to shareholders correspond more closely to the target language register conventions.

In his study of thematic structure between English and German popular scientific writing, Freiwald (2016) compares original clauses and their matching translations in both translation directions. Regarding multiple Themes, he reveals that the number of Theme elements is significantly lower in German translations and significantly higher in English translations compared to the source language. This difference can be attributed primarily to the Finite-second constraint in German (see Section 5.9). While the number of non-Subject Themes is much higher in German originals than in English originals, this difference is again minimized in the translations, with both German and English translations falling somewhere in the middle between the originals (Freiwald 2016: 45). It was hypothesized that in translations into English, marked, non-Subject Themes were a strong predictor of changes in the translations due to the difference in word order freedom between English and German; however, only certain kinds of marked Themes, namely Matter, Quality, Viewpoint, and Complement Themes, were changed often enough to produce statistically significant results (Freiwald 2016: 62). Surprisingly, in translations into German, certain kinds of circumstance Themes, namely Comparison and Concession Themes, also lead to frequent changes in the translations even though German's free word order should allow any kind of element in Theme position. Still, apart from these cases, translators generally tried to preserve the experiential meaning of the Theme.

Niemietz, Neumann and Freiwald (2017) also study Theme structures in popular scientific texts in English to German and German to English translations. In fact, their analyses are largely based on the same data as Freiwald (2016), which is why, unsurprisingly

the overall results on Theme markedness and multiple Themes are consistent with Freiwald (2016). However, they further analyzed multiple Themes and the translation of Subject Themes in particular and found that the combination of textual Theme and Subject is the most common multiple Theme structure in English originals as well as English and German translations (Niemietz, Neumann and Freiwald 2017: 343). Only German originals had a higher number of multiple Themes that include a textual and a circumstance Theme given the high number of circumstantial Adjunct Themes in GO. Subject Themes in GO typically stay Subject Themes in ET and if an experiential Theme is missing in German, a Subject Theme is generally added in the translation. In the opposite translation direction, Subject Themes in EO are often missing in GT. This result is, on the one hand, due to the higher acceptability of non-Subject Themes in German and, on the other hand, due to the fact that the German Forefield typically does not allow multiple elements and many Subject Themes were moved to a post-verbal position.

Kim and Huang (2012) and Liu and Yang (2013) only consider the study of Theme in translations meaningful if it is applied to reveal differences and similarities of thematic progression. A Theme analysis that is simply on the clause level without referring back to the general development of the text is "of very limited value for translation" (Liu and Yang 2013: 275). While I echo the significance and relevance of thematic progression in translations, I disagree with the alleged insignificance of Theme analysis on the clause level. Theme essentially operates on two levels: the clause level, where Theme functions as a local organizer of the message and sets up a context, on the basis of which the rest of the message is interpreted, and the text level, where the sequence of single Themes represents the general progression of the text. Undoubtedly, a study would be maximally informative if it considered both of these levels. However, a study of Theme differences on a more local, clause-based level can also reveal similarities and differences in contextualization that are just as relevant for translations as thematic progression.

5.11 Hypotheses

Based on the contrastive differences in Theme outlined in Section 5.9 and the insights from the state of the art, it is plausible to assume that the Theme of the source text clause has a significant impact on the likelihood of a shift in the translation. While some differ-

ences between German and English thematic structure are hard rules that need to be followed to assure grammaticality, others are a question of preferences and frequencies. In any event, these differences between the original texts are likely to affect translations as well. In the following, two sets of hypotheses will be postulated. The first set is based on a general analysis of original and translated texts in both translation directions. In this analysis, general frequencies between the two sub-corpora in each direction will be compared and tested for significance. The second set of hypotheses is based on inferential statistics. Here, logistic regression models are used to evaluate the effects of different predictor variables on the response variable Change (see Section 6.4). In simpler terms, logistic regression considers all the thematic attributes of all the original clauses and calculates whether any of these attributes reliably predict the presence or absence of a change in the translated clauses on average. These two sets of hypotheses will be fairly congruent. If for example English translations have significantly fewer multiple Themes than German originals in the general analysis, it is plausible that multiple Themes in German originals are a significant predictor variable of Change in the inferential statistics. However, this is not a foregone conclusion since the lower number of Theme elements in the English originals may also be caused by another variable, for example the lower likelihood of marked Themes in English (and as will be shown in Section 9.1.1, this is in fact part of the reason of lower Theme numbers in the Subject hypothesis). For this reason, it is important to test each of these hypotheses separately.

Chapters 7 and 8 analyze the Theme distributions in English and German originals and sheds light on register differences. In Chapter 9, Themes in English and German originals texts are compared to work out all relevant contrastive Theme differences between the two languages. Not only do these analyses offer valuable insight in their own right, they also are highly relevant for translation. Nevertheless, no hypotheses will be formulated for the intralingual and contrastive analyses given that the focus of this thesis rests on translations.

The frequency of multiple Themes and the number of Theme elements in total is highly dependent on the way Theme is formally analyzed. For English, two generally accepted hypotheses have been presented, namely the first experiential element hypothesis and the Subject hypothesis. For German, Steiner and Ramm (1995) and Steiner and Teich (2004) propose the Forefield hypothesis. As was argued for in Section 5.8, the same hypotheses used in English can be applied to German Theme in like manner. Due to this

multitude of possible formal realizations of Theme, three combinations of Theme hypotheses will be tested for in the translation analyses:

1. The Theme up to and including the first experiential element in both languages
2. The Theme up to and including the Subject in both languages
3. The Theme up to and including the first experiential element in English and the Theme as the Forefield in German

For the first two comparisons, the same formal realizations for both languages will be used. The third comparison will consider the first experiential element hypothesis for English and the Forefield hypothesis for German. These represent the two Theme descriptions which are most commonly used by the state of the art for each respective language. And, if German Themes do not need an experiential element, the function of the Forefield comes closest to the functions served by Theme elements in English. A comparison of the two is therefore highly relevant.

For the intralingual as well as the contrastive analysis, a further Theme hypothesis of English will be considered, namely the first element hypothesis. This hypothesis was proposed but subsequently rejected by Halliday (1985) and, as far as I am aware, no systemist argues for the Theme to be instantiated by only the first element in English. That being said, the very first element of a clause undoubtedly carries great significance⁶⁶ even if it does not necessarily conclude the entire point of departure of the message. And since the Forefield hypothesis in German is in most cases restricted to a single constituent a comparison of this very first element in both languages is still relevant. Additionally, the first element of a sentence can be queried easily in a corpuslinguistic approach. Comparing its accuracy with the other Theme hypotheses will be quite illuminating. This Theme hypothesis will only be considered in Chapters 7 to 9. In Chapters 10 and 11, in which translation effects will be analyzed, only the above-mentioned three comparisons will be made, primarily to ensure clarity and manage extent.

This multitude of combination of Theme hypotheses can make the formulation of hypotheses challenging. This is particularly true in the case of average Theme numbers because these numbers naturally change depending on the scope of the Theme. For this reason, each Theme hypothesis combination in both translation directions needs to be considered separately. In translations from English to German, the number of Theme ele-

⁶⁶ The first element arguably carries the greatest thematic significance if the wave metaphor is accurate (see Section 5.2.3).

ments likely decreases if the English Theme is based on the first experiential element hypothesis and the German Theme is based on the Forefield hypothesis. Most Forefields are only instantiated by a single constituent and if the English original includes multiple Theme elements, it is likely for one or more elements to be moved to the Midfield in the German translation. If the Theme is based on the first experiential element hypothesis in both languages, the opposite effect takes place. If the source clause features multiple Theme elements and one of them enters the Midfield in the target clause, the German Theme includes an additional Finite Theme, which separates the Forefield from the Midfield. This additional Finite Theme significantly increases the average number of German Theme elements. The same holds true for the Subject hypothesis.

In the opposite translation direction, the relationships are reversed. The average number of Theme elements increases in English translations if the German Theme represents the Forefield. In this case, the Theme in German originals will be frequently instantiated by a lone textual or interpersonal Theme, which needs to be changed in the English translations, likely by an increase in Theme elements. In the first experiential element hypothesis and the Subject hypothesis, the average Theme number, in turn, decreases because the common Finite Themes in German have to be omitted.

One of the greatest contrastive differences between English and German is their difference in word order freedom with German having a generally freer word order and English being syntactically more restricted. As a consequence, only Subject Themes are considered unmarked Themes in English declaratives while German merely has a weak notion of Theme markedness. However, Freiwald (2016) showed that certain types of circumstance Themes are infrequent in German and as a consequence are changed regularly in German translations. Additionally, he demonstrated that not all non-Subject Themes in translations into English are strong predictors for translation shifts. Whether a certain type of non-Subject Theme is changed regularly in translations is largely due to its thematic potential in original writing. For example, if circumstances of Concession are generally more likely to occupy a Theme position in English originals, they are not likely to be moved out of the Theme in translations into English. Nevertheless, it can still be assumed that Theme markedness as a translation problem is generally more relevant for translations into English.

In addition, processing aspects have been demonstrated to influence word order and constituent structure (see for instance Diessel 2005). As a result, short constituents usually precede long constituents, especially in pre-verbal position, to facilitate online processing. This is not necessarily a contrastive difference between English and German but a consideration in clause structure overall. For this reason, translators may opt to change particularly long experiential Themes as a form of simplification (Baker 1993). At the same time, short experiential Themes may also be prone to change to make their references more explicit in translations.

The semantic mapping onto the Subject is less restricted in English to compensate for the lack of word order freedom. For this reason, English allows non-sentient constructions, which represent a combination of an inanimate Subject and a verb that requires a sentient participant. Kast (2012), Serbina (2015) and Freiwald (2016) already demonstrated that such constructions are a challenge for German translators and as a result are often changed. However, Freiwald (2016) also showed that this effect was only observable for concrete and nonconcrete Inanimates and not for Machine, Place, or Time Subjects (see Section 6.3). Place and Time Subjects were too infrequent in the popular scientific register to allow for generalizations, but the preservation of non-sentient construction including Machines was considered systematic. In translations from German to English, non-sentient constructions will be less relevant given that there is no need for them to be changed in the target language.

Identifiability and definiteness are two of the most reliable factors that govern constituent order in the German Midfield. As a consequence, a non-identifiable Subject Theme in English originals can be problematic for German translators, especially if it is accompanied by additional Theme elements. Normally, the German translators have the option to simply move the Finite in between the first Theme element and the Subject in the translations. If, however, the original Subject is non-identifiable, it would have to be positioned at the onset of the German Midfield, which is the default position of identifiable constituents. In such cases, the translators have to decide between staying close to the original Theme structure and avoiding a marked constituent order in the target language. For this reason, non-identifiable Subject Themes can be expected to undergo more changes than identifiable Subject Themes in translations into German. In the opposite translation direction, there are no restrictions in the target language regarding the use and position of

identifiable and non-identifiable Subject Themes. Notwithstanding, non-identifiable Subject Themes may still be more prone to change if the English translators want to make the referent of the Subject more explicit.

Studies like Hasselgård (1997, 2000), Teich (2003), Neumann (2013), Freiwald (2016), and Niemietz, Neumann, and Freiwald (2017) have shown that the thematic structure in translated language is not entirely like the Theme of the source text nor is it identical to Theme structure in original writing in the target language. Most findings in translations represent a mixture of the two, where the effects of shining through and normalization are observable simultaneously. Depending on the markedness of a feature, one effect may be more prevalent than the other. Interestingly enough, the effect of shining through was observed to be more productive in translations from English compared to translations into English (see for instance Hasselgård 1997; Freiwald 2016; Evert and Neumann 2017). Evert and Neumann (2017) attribute this difference to the difference in prestige between source and target language.

The following hypotheses were formulated based on all considerations above:

I. Analysis of German and English original and translated texts

a. English-German translations

- H_{1.1.1} In the Forefield hypothesis, the average number of Theme elements decreases significantly in German translations due to the Finite-second constraint.
- H_{1.1.2} In the first experiential element hypothesis and the Subject hypothesis, the average number of Theme elements increases significantly in German translations due to the Finite-second constraint.
- H_{1.1.3} Non-Subject Themes are significantly more frequent in German translations across all Theme hypotheses.
- H_{1.1.4} Inanimate Subject Themes in combination with sentient verbs are significantly less frequent in German translations.
- H_{1.1.5} The frequency of middle animate Subject Themes in combination with sentient verbs in German translations do not deviate significantly from English originals.
- H_{1.1.6} The number of identifiable Subject Themes significantly increases in German translations.

b. German-English translations

- H_{1.2.1} In the Forefield hypothesis, the average number of Theme elements increases significantly in English translations due to the Finite-second constraint.
- H_{1.2.2} In the first experiential element hypothesis, the average number of Theme elements decreases significantly in English translations due to the Finite-second constraint.
- H_{1.2.3} In the Subject hypothesis, the average number of Theme elements decreases significantly in English translations due to the Finite-second constraint and the higher number of non-Subject Themes.
- H_{1.2.4} Non-Subject Themes are significantly less frequent in English translations across all Theme hypotheses.
- H_{1.2.5} The number of identifiable Subject Themes significantly increases in English translations.

II. Regression analysis in translations between English and German

a. English-German translations

- H_{2.1.1} High Theme numbers are a significant predictor of Change.
- H_{2.1.2} Experiential Themes that are not of medium length are significant predictors of Change.
- H_{2.1.3} Non-Subject Themes whose thematic potential is lower in German than in English are significant predictors of Change.
- H_{2.1.4} Inanimate Subject Themes in combination with sentient verbs are significant predictors of Change.
- H_{2.1.5} Middle animate Subject Themes, including Machine Subjects, in combination with sentient verbs are not significant predictors of Change.
- H_{2.1.6} Non-identifiable Subject Themes are significant predictors of Change.

b. German-English translations

- H_{2.2.1} High Theme numbers are a significant negative predictor of Change in the Fore-field hypothesis.
- H_{2.2.2} High Theme numbers are a significant positive predictor of Change in the first experiential element hypothesis and the Subject hypothesis.
- H_{2.2.3} Experiential Themes that are not of medium length are significant positive predictors of Change.
- H_{2.2.4} Non-Subject Themes whose thematic potential is lower in English than in German are significant predictors of Change.⁶⁷
- H_{2.2.5} Neither middle animate nor inanimate Subject Themes in combination with sentient verbs are significant predictors of Change.
- H_{2.2.6} Non-identifiable Subject Themes are significant predictors of Change.
- H_{2.3} Translation shifts are more common in translations into English than translations into German.

⁶⁷ H_{2.1.3} and H_{2.2.4} rely on the thematic potential analysis of German and English originals and will be revised following the discussion of contrastive differences.

6 Methodology

This chapter presents the methodology. First, the CroCo corpus is introduced including a description of the registers that were analyzed. Subsequently, annotation tools are presented and the annotation decisions regarding all of the annotated Theme aspects are outlined. The chapter concludes with a description of the statistical tests that were performed.

6.1 The corpus

The corpus that was used in this study is the CroCo Corpus (Hansen-Schirra, Neumann, and Steiner 2012), which was compiled within the DFG-funded CroCo project on linguistic properties of translated texts conducted at Saarland University, Germany. The CroCo Corpus is a bi-directional translation corpus of English and German, meaning that it includes original texts in both English and German as well as matching translations of these texts in both translation directions. Each group of texts is organized as a sub-corpus: the four sub-corpora are thus English originals (EO), German originals (GO), English translations (ET), and German translations (GT). The same eight registers are included in all sub-corpora, which will be further discussed below. The corpus includes roughly one million words in total or 31,250 words per register in each sub-corpus. Each register consists of at least ten different texts, but possibly more if the average text length is generally shorter. Besides these core corpora, the CroCo Corpus also includes two reference corpora of samples of English and German original texts, taken from a variety of different registers. However, these reference corpora were not used in the present study.

The advantages of a bi-directional translation corpus are manifold. The inclusion of original sub-corpora allows analyses of the two language systems both intralingually as well as contrastively. The fact that the translation sub-corpora align with the original corpora limits the number of external factors that may be the cause of differences between source and target texts. Furthermore, it enables the corpus linguist to not only find linguistic patterns that are more or less frequent in the translations but also investigate translation procedures. Thus, the design of the CroCo Corpus is ideal for the purposes of this thesis. Featuring two translation directions also makes it possible to identify effects

in the translations that can be attributed to contrastive differences and interferences as well as effects that can be ascribed to the translation process itself and are more translation-inherent.

The eight registers in the CroCo Corpus are tourism leaflets (TOU), instruction manuals (INSTR), fictional texts (FICTION), political speeches (SPEECH), political essays (ESSAY), popular-scientific texts (POPSCI), letters to shareholders (SHARE), and website texts (WEB). Since almost all annotations were carried out manually (see below), the number of analyzed registers was limited to four. A detailed analysis of all eight registers in both languages as well as both translation directions would have exceeded the scope of this project. TOU, INSTR, FICTION, and SPEECH were analyzed because they represent an appropriate selection of registers of different contextual configurations and accordingly portray different common patterns of language use (Halliday and Hasan 1985). They were also identified as the most likely registers to feature interesting cases of some or all of the previously discussed Theme-related aspects such as different kinds of marked Themes, multiple Themes, non-sentient constructions, etc.

The FICTION register includes contemporary and sophisticated literary texts, which were translated into the respective other language. Since the individual texts are not related to each other in any way, their fields of discourse are quite diverse. The INSTR register represents a collection of instruction manuals which describe and promote a variety of different products, like computer equipment, electric appliances, and tools. The mode of discourse of instruction manuals is unique in the sense that it also includes non-verbal forms of showing. However, since only verbal aspects were included the CroCo corpus, these could not enter the analysis. SPEECH comprises manuscript of political speeches. What sets SPEECH apart from the other three registers is that its texts were written to be spoken. The tenor of discourse is also more diverse since the speakers generally have two groups of audiences, namely the audience that is present during the speech itself as well as the general public. Both of these groups are occasionally referenced by the speaker. The field of discourse covers a wide array of political topics. The register TOU consists of promotional tourism texts, which provide information about a place like a city, a county, or a country. The field of discourse is in so far homogeneous as each leaflet focusses on the features that make the place attractive for travel (Neumann and Hansen-Schirra 2012: 28-30).

For further detail on the corpus generally and the individual registers specifically, see Neumann and Hansen-Schirra (2012), on which this entire section is based. Also see Hansen-Schirra, Neumann, and Steiner (2012), which includes further information about corpus enrichment as well as interesting analyses using the CroCo Corpus.

6.2 Annotations tools

The CroCo Corpus is annotated and aligned on several levels such as word, chunk, clause, and sentence level (Hansen-Schirra and Neumann 2012: 35-36). Unfortunately, for technical reasons, these annotations could not be accessed, which is why only the raw text data of the CroCo Corpus was used in the course of this study. The only alignment layer that could be used was the alignment of source text and target text sentences. This means that every sentence in the original language data was always immediately followed by the matching sentence in the translated text.

The majority of annotations was carried out using the UAM CorpusTool, a text annotation tool that is specifically designed to satisfy the needs of (computational) linguists. The tool offers automatic sentence segmentation as well as statistical analyses. Most importantly, it allows the user to create their own hierarchically structured annotation schemes. Moreover, the tool can be used to automatically annotate segments in one annotation layer based on the tags in another annotation layer. For example, the annotations of Mood elements were used to presort Theme elements into Subject, Complement, and circumstance Themes. For more information on the UAM CorpusTool, see O'Donnell (2008a) and O'Donnell (2008b).

All annotations were carried out manually since, to my knowledge, there is no automatic tool to reliably annotate Theme structure. Halliday and Matthiessen (2014: 70) even go as far to claim that a full-fledged systemic functional annotation cannot be carried out automatically. The UAM CorpusTool does allow an automatic annotation of English Theme, but the level of accuracy and detail was not as high as desired, at least not at the time this project was launched. Moreover, since the annotation of the German Theme had to be conducted manually in any case, it was deemed more consistent and more comparable to carry out the entire set of annotations manually. However, the auto-code function of the UAM CorpusTool did reduce the time spent on annotations significantly.

The eight major Theme analyses that were conducted and that will be discussed in detail in the following chapters include multiple Themes, Theme markedness, Participant Theme, circumstance distribution, Subject Theme sentience, Subject Theme identifiability, and Theme length. The latter two analyses were done in Excel, and Theme length is the one measure that was calculated automatically based on the number of characters in the experiential Theme. Additionally, an analysis of all the Mood elements as well as a complete transitivity analysis were conducted and used as the base schemata to assign segments in the other annotation layers.

The annotations in the UAM CorpusTool were extracted and converted into Excel spreadsheets. As was previously discussed, three different Theme realizations are considered for each language, which required the design of six separate sheets. The different Theme boundaries were identified automatically using *If*-functions in Excel.⁶⁸ Every clause in the original was contrasted with its translation and their Theme categories were compared with each other. On the basis of this comparison, the two Change categories were calculated (see Section 6.4).

6.3 Annotation decisions

In this section, all annotation decisions will be outlined. I decided to be particularly detailed in this section because I want to promote a more comprehensive and transparent exchange in the systemic functional community regarding Theme annotations.

The texts in the CroCo corpus were not filtered beforehand but analyzed whole. Only the Themes of declarative clauses were analyzed. Declaratives represent the by far most common mood type in the four analyzed registers and are in focus in most empirical studies on Theme. Additionally, many thematic aspects are highly dependent on mood, such as Theme markedness and multiple Themes. If this study had also considered other moods like interrogatives and imperatives, they would have to be analyzed and discussed separately from each other. This would have exceeded the scope of this project. Interrogative and imperative clauses were tagged in the annotations, so that their thematic structures may be analyzed in future projects. In some cases, the original clause is declarative

⁶⁸ For example, to accurately identify Theme in German based on the Forefield hypothesis, Excel checked which clause constituent was analyzed as the Finite and subsequently disregarded the Finite and all following elements for the Theme.

but is changed to an interrogative or an imperative in the translations, or the other way around. Although such sentence pairs are very interesting and warrant further analysis, they were also not included in the results.

Apart from imperatives and interrogatives, other sentence pairs were also neglected, for example minor clauses without a process, cases of sentence splittings or merging, cases of missing or erroneous alignment, as well as the occasional instances of sentences in a different language than English or German. Cleft constructions were categorized as predicated Theme, thematic equatives, and thematized comments (see Thompson 2014: 153-157) but ultimately summarized in the results given their overall low rate of occurrence. Finally, clauses with the pattern of '*the ... the*' in English and '*je ... umso/desto*' in German (see example (74)) were only tagged but not annotated further as I was not able to come up with a satisfying Theme annotation.

(74)

EO: *The more questions that were asked, the more people came.*

GT: *Je mehr Fragen gestellt wurden, umso mehr Leute kamen dazu.*
'the more questions asked were, the more people came there.'

[E2G_FICTION_010]

The unit of analysis in this study is not the individual clause and also not the sentence but the T-unit, which comprises an independent clause as well as all hypotactic clauses that are dependent on it (Thompson 2007: 680). So, if a clause complex begins with a hypotactic clause like an *if*-clause, for example, that hypotactic clause is analyzed as a constituent of the independent clause, which is in most cases a circumstantial Adjunct. The T-unit is a commonly used unit of analysis in studies on Theme (see for example Fries 1995b; Martin 1995; Hasselgård 2000; Thomson 2005; Thompson 2006, 2007). For technical reasons, only the very first T-unit in each clause complex was analyzed. That means if a sentence contains two or more paratactic T-units, only the Theme structure of the very first was annotated. The reason for this is when the data from the UAM CorpusTool is converted to Excel, every analyzed Theme is ordered chronologically, one below the other. Since every original clause is immediately followed by its translation in the texts, the two matching Themes could be easily and automatically identified in Excel. If, however, one original Theme could be immediately followed by (an)other original Theme(s), the mapping of original and translated clause would have had to be done manually. This problem could have been avoided, had I previously queried the corpus for individual declarative clauses,

but that was not possible due to a transmission error in the query tool. One advantage of only analyzing the very first T-unit of a sentence is that elliptical experiential Themes were virtually non-existent.

The analysis of systemic functional categories is primarily based on the instructions in Halliday and Matthiessen (2014). For difficult cases, of which there were many, Martin, Painter, and Matthiessen (2006) as well as Bartlett (2014) were consulted, which both include very helpful annotation probes. Steiner and Teich (2004) served as the primary reference for the German analyses, but since the German SFL descriptions are considerably less detailed, I also had to apply and adjust the probes in English for the German annotations. Some of the probes lend themselves well to German analyses, like the *What-do* test (Bartlett 2014: 49) or the projection test (Bartlett 2014: 65); others like the present-in-present probe cannot be easily applied to German grammar. The grammatical Subject was identified through position, verb agreement, and the tag-question-test. Given the vague functional definition of Complements (see Section 4.4), they were not analyzed on the basis of their potential of being the Subject since too many Complements do not meet this requirement. Any constituent that is part of the argument structure of the process but is not the Subject was annotated as a Complement. These include constituents that are analyzed as Subject and Object Predicatives as well as obligatory Adverbials in traditional grammar (Quirk et al. 1999: 145-151). The reason why obligatory Adverbials were not analyzed as circumstantial Adjuncts is because they behave more like Complements regarding their position and are also a lot more marked than conventional circumstance Themes (see example (75); obligatory circumstance in bold).

(75) ***On the table***, he put the book.

The German Theme in the Forefield hypothesis ends with the Finite. If the Theme in German exceeds the Forefield, however, the finite verb will often be positioned before the first experiential element and thus needs to be included in the Theme analysis. In English and German polar interrogatives, the finite verb is fronted and analyzed as an interpersonal Theme because it expresses mood. The second-position Finite in German can also be analyzed this way given that its position is also an expression of mood. However, as in English, the Finite can be conflated with the process of the clause if it is in simple aspect.

Thus, technically, it will often constitute the first experiential element in German declarative clauses, which would mean that German has a high number of presumably unmarked Predicator Themes. This issue was already addressed in Section 5.8, and it was argued against analyzing the conflation of Finite and process as a meaningful experiential Theme for two main reasons. First, if choices regarding word order are made exclusively in different systems and are not a product of decisions in *THEME*, they should not be considered meaningful for Theme. A Finite is positioned early in German not because it has been selected as an appropriate point of departure but because of its role as the Finite in *MOOD*. It is for that same reason that German grammarians like Dürscheid (1989: 7) and Engel (2004: 165) argue against analyzing the field that follows the conflation of finite and lexical verb as the Postfield even though the Postfield is positionally defined as everything following the lexical verb. In their minds, the 'Satzklammer' is still existent even if the lexical verb is not in its regular position and thus, the following field remains the Midfield because it is the finite and not the lexical portion of the verbal group that is relevant for its positioning.

Matthiessen and Martin (1991) make the same argument for *be* in English polar interrogatives, which is technically the first experiential element if it is in simple aspect. Nonetheless, they argue that the form of *be* is positioned early due to its function as Finite and not because the speaker considers the process to be a suitable point of departure and they therefore do not regard it an experiential but an interpersonal Theme (Matthiessen and Martin 1991: 48). This line of argument does have one considerable weakness, which is that the position of the Finite is also not a result of a choice in *THEME* but in *MOOD* and thus should not be considered thematic at all. Still, at least the finite part of the verb can be argued to be inherently thematic like finite verbal operators in interrogatives or Predicator Themes in imperatives as a means to signal the speech function to the hearer early.

The second reason why an analysis of Finites as Predicator Themes is not ideal has to do with the translation focus of this thesis. If German Finites were analyzed as experiential Themes, a very high number of clause pairs in translations from English to German would have to be analyzed as translation shifts because in many cases, Subject Themes would be changed to Predicator Themes in German (see example (76); Finite-process conflation underlined, Subject in bold).

(76)

EO: *Then **she** snatched my hand and whisked me off to the hospital.*

GT: *Dann schnappte **sie** sich meine Hand und schleppte mich ins Krankenhaus.
'then snatched **she** [refl-3sg] my hand and dragged me to.the hospital.'*

[E2G_FICTION_008]

This is exactly what Hasselgård (2004) found in her analysis of Theme differences in English-Norwegian translations. However, these supposed changes would simply be due to the translator adhering to the syntactic rules of the target language and in my eyes would not represent meaningful changes to the experiential Theme of the source clause. For these reasons, all finite verbs in between Forefield and Midfield were analyzed as interpersonal finite verbal operators.

The difference between textual and interpersonal Themes was, for the most part, easily identifiable in English and German. Yet, some fronted constituents arguably create cohesive ties between clauses and comment on the nature of the clause simultaneously. A good example of that is *in fact*, which creates a textual link between clauses but also comments on the way the introduced clause needs to be read. In fact – pardon the pun – Halliday and Matthiessen (2014) list *in fact* as both a verificative conjunctive Adjunct (108) and a mood Adjunct of counterexpectancy (189). I generally leaned towards interpersonal Theme if a case could be made that the speaker was commenting the clause. Other difficult cases were adverbs like *only* and *even* in English, which have interpersonal meaning but can be both comment Adjuncts on the clause level and modifiers inside nominal groups (see example (77)).

(77) *Even the casual tourer cannot help but notice Scotland's wildlife.*

[E2G_TOU_002]

The same problem arises in German with cases like *zumindest* (*at least*) and *vor allem* (*especially*). Conveniently, in the Forefield they are either separated from the next constituent if they function as an interpersonal Adjunct (see example (78)) or they are attached to the constituent and share the Forefield if they are modifiers (see example (79)). The majority of cases in English were not analyzed as interpersonal Themes but rather as part of the following group, unless the adverb clearly took scope over the entire clause.

(78) *Vor allem aber entspricht es der europäischen Gründungs-Idee.*
'especially but corresponds to it the European foundation-idea.'
[G2E_SPEECH_014]

(79) *Vor allem die Staaten an der Südflanke Europas stellt das vor enorme Probleme.*
'especially the states at the south flank of Europe confronts this with enormous problems.'
[G2E_SPEECH_006]

Other possible cases of dispute regarding textual and interpersonal Themes involve fronted expression like *that's why*, *turns out*, or *I guess*, which can be analyzed as the Subject and Predicator of the independent clause or as conventionalized, rank-shifted clauses, functioning as textual or interpersonal elements like *therefore*, *evidently*, and *apparently*. In most cases, they were analyzed as textual and interpersonal Themes, especially if they involved contractions, ellipses, or clippings (which points towards their conventionality) or if they were separated by a comma from the rest of the clause (see example (80)).

(80) *I mean, even I can manage on a good day.*
[E2G_FICTION_006]

Participant role analysis is also based on Halliday and Matthiessen (2014), but some further roles, like a Place participant in material processes, had to be included to account for the aforementioned additional Complements. One general area of disagreement among systemic functional grammarians involves the analysis of metaphorical meaning. When in doubt, I generally relied on the grammatical probes to come to a decision, which usually resulted in annotations on a more literal rather than metaphorical level (see example (81)). I am not arguing that such analyses are generally preferable, but I wanted to stay consistent in my annotations and allow maximum replicability.

(81) *The OSCE plays a critical role in our effort to promote democracy, human rights, and rule of law throughout Eurasia.*⁶⁹
[E2G_SPEECH_006]

⁶⁹ In its literal sense, playing a role is a material act, which is why the *what-do* test and arguably the present-in-present test can be applied. However, in a metaphorical sense, the OSCE, which is a security organization, does not actually play a role but is critical for promoting democracy, which is more like a relational process. Following the probes, this clause was analyzed as material process. That being said, the fact that the speaker chose simple rather than progressive aspect suggests that this lexico-grammatical pattern is somewhere in between the two process types.

Existential processes do not have a first participant, neither in English nor German, because the grammatical Subjects *there* and *es (it)* are semantically empty. That is why the unmarked Theme in both of these constructions is the process itself. Nevertheless, Predicator Themes in existential processes are represented as a sub-category of Subject Themes as I wanted to separate them from other, highly marked Predicator Themes.

The material participant roles Recipient and Client were summarized under Beneficiary because of their rarity, especially in Theme position. In identifying relational processes, the participants Token and Value rather than Identified and Identifier were used in the annotations because their analysis relies less on context and they can be identified more easily by substituting the verb for *represent* and by considering the voice (Halliday and Matthiessen 2014: 283).

Halliday and Matthiessen (2014) offer a very detailed list of different circumstantial meanings, which also fit German circumstances surprisingly well. All 22 types of circumstances are distinguished in the analyses and, for the most part, their distinction was straightforward. However, there were two cases in particular which were repeatedly difficult to analyze. The first involves what is known as detached predicatives or absolute constructions (Biber et al. 1999: 136), which are often fronted in English and give additional information about the Subject (see example (82)). Such constructions also exist in German (see example (83); detached Predicatives in bold) but are a lot less frequent (Behrens and Solfeld 2014: 273).

(82) **Full of character**, Pennine centres Hebden Bridge, Halifax and Huddersfield make a friendly base for exploring the region's industrial heritage.

[E2G_TOU_010]

(83) **Einst Traumziel für lange gereiste Seeleute**, ist dieser Teil der Stadt heute ein "Muss" für alle Touristen.
'**once dream.destination for long traveling sailors**, is this part of the city today a "must" for all tourists.'

[G2E_TOU_008]

One option was to analyze them as part of the Subject;⁷⁰ however, in those cases where the predicative is omitted in the translation – of which there were many in translations

⁷⁰ This is the analysis that Thompson (2014: 164) proposes. He argues that the real starting point of the message is the nominal group while the proposed attributive is just a way of smuggling in extra information.

from English to German – this translation shift is potentially not registered. A second option was to analyze them as circumstances, even though circumstances rather represent background information on the process (Thompson 2014: 114). That being said, the circumstance category that comes closest to detached predicatives is *Guise*, which is also more connected to a nominal group than the entire process. The final option was to create a new category. I decided for the second option because I wanted to avoid further subdivisions of the already very detailed circumstance categorization. Therefore, I analyze detached predicatives primarily as *Guise* and occasionally as *Quality* if the fronted circumstantial information could also be argued to describe the manner in which the process was carried out.

Additional information on tools represents another difficult case with regard to circumstance classification. These meanings are particularly common in INSTR and describe what the reader needs to use to achieve the desired effect. In most cases, these circumstances are prepositional phrases introduced by *with* in English and *mit* (*with*) in German. Instrumental circumstances were generally analyzed as *Means* (see example (84)) unless the question *How?* was not suitable at all, in which case they were categorized as *Comitatives* (see example (85); circumstances in bold). Prepositional phrases introduced by *with* were generally difficult to analyze, which is why the category of *Comitatives* represents a very diverse group of meanings in the analysis.

(84) ***With right arrow*** you open any existing submenus.

[E2G_INSTR_005]

(85) ***With a coastline of 650 miles*** opportunities for sea fishing are limitless [...].

[E2G_TOU_006]

Subject Theme animacy and non-sentient constructions are not a common analysis in systemic functional studies, which is why annotations were based on studies outside of the systemic functional framework. The categorization of the different types of Subjects was based on Zaenen et al. (2004) as well as Garretson (2004), who propose very similar annotation schemes. They distinguish between Human, Animal, Organization, which refers to collectives of humans, Place, Time, concrete Inanimates, and nonconcrete Inanimates. The latter two refer to all entities that are inanimate but cannot be sorted into any of the other categories. Concrete Inanimates refer to concrete objects and substances, which can be perceived with the senses, while the rest is categorized as nonconcrete Inanimates

(Garretson 2004: 29-30). Zaenen et al. (2004) include two further categories, intelligent Machines and Vehicles, and propose an animacy hierarchy that distinguishes between Human, other animates, which include Animal, Organization, Machine, and Vehicle, and inanimates, referring to Place, Time, concrete Inanimates, and nonconcrete Inanimates. In how far this distinction is accurate for English and also applies to German is not immediately relevant here since the annotations are only based on the nine sub-categories. Nevertheless, the labels *middle animate Subjects* and *inanimate Subjects* will be used to refer to these two Subject groups. The animacy list was extended by one additional type, namely Process Subjects, which are predominantly realized as rank-shifted clauses and are part of the category of inanimate Subjects (see example (86); Subject in bold). I would assume Zaenen et al. (2004) and Garretson (2004) analyze these as nonconcrete Inanimates. Nonetheless, I found such processes to be too different from the rest of the category in terms of both form and meaning.

(86) ***Spilling a liquid into any electronic instrument*** will damage the circuitry.
[E2G_INSTR_002]

In most cases, the Subject analyses were non-problematic. At times, it was difficult to decide whether parts of computers were Machines or concrete Inanimates, especially in INSTR. As a general rule, I considered all computer programs Machines, but not entities that were created by the programs, such as menus, icons, and buttons, which I analyzed as concrete Inanimates. In SPEECH and TOU, the distinction between Organization and Place was also sometimes fuzzy if the Subject referred to a country or a town (see example (87); Subject in bold). Whenever the Subject could be understood as metonymically referring to a group of people, I tended towards Organization.

(87) *Seit 1985 vergibt **Augsburg** alle drei Jahre den mit 12 500 Euro dotierten Friedenspreis für Leistungen zur Förderung der Gemeinsamkeiten unter den Konfessionen.*
*'since 1985 awards **Augsburg** every three years the with 12,500 Euros endowed peace.prize for advancements of.the commonalities among the denominations.'*
[G2E_TOU_002]

One of the more challenging analyses involved the agency and sentience requirements of the verbs. These requirements were not analyzed if they were not relevant for non-sen-

tient constructions. This includes all passive clauses and clauses with non-referential Subjects like existential processes and clefts. If the Subject was analyzed as Human, a verb analysis was also skipped since humans can be easily paired with any kind of verb in both languages. Further analysis of the verb was also not necessary if the process was previously analyzed as relational, mental, or behavioral. Relational processes do not make any requirements of their first participants in English or German and should therefore not be relevant for an analysis of non-sentient construction. While this is true for the majority of relational processes, the impression arose that even some relational verbs in combination with inanimate Subjects are more marked in German and regularly lead to translation shifts. In example (88), the relational process (*to*) *hold* is apparently still too agentive and therefore turned into passive in the German translation.

(88)

EO: *Tray 1 (the multipurpose tray) holds up to 125 sheets of paper [...].*

GT: *In Fach 1 (Mehrzweckfach) können bis zu 125 Blatt Papier [...] eingelegt werden.
'in Tray 1 (multipurpose.tray) can up to 125 sheets of.paper put be.'*

[E2G_INSTR_001]

Mental and behavioral processes also did not have to be further analyzed because they require a sentient first participant by default (Halliday and Matthiessen 2014: 250). Thus, whenever this principle was violated and an inanimate Subject was used as a Senser or Behaver, it was automatically annotated as a non-sentient construction. Material and verbal processes do not have general requirements regarding Actors and Sayers, which is why each verb that belonged to these process types had to be analyzed individually.

For the verb semantics annotations in English, I relied on FrameNet, which is a computational lexicography project and online database that links English verbs to the semantic frames that underlie their meanings (Fillmore, Johnson, and Petruck 2003: 235). Also part of these frames are the obligatory and optional participants that carry out or are affected by the verbs. Conveniently for the Subject sentience annotations, each participant is analyzed in terms of their semantic type, which includes the category *Sentient*. The database is quite detailed, which is why most verbs in the CroCo corpus could be either directly identified or clearly associated with other verbs with very similar meanings. The German verb annotations were based on the SALSA database (Burchardt et al. 2009), which is a similar frame semantic lexicon of German verbs and which also uses the same frames and semantic type categories as FrameNet.

The use of FrameNet allowed an analysis that was not annotator-dependent, which furthers replicability. However, I personally did not always agree with the categorizations, especially in the context of contrastive differences between English and German. For example, the frame *Evidence* includes the participant *Support*, which is typically a fact that supports a claim and is thus not necessarily sentient. This frame is tied to verbs such as *show* and *indicate*, and since SALSA is based on the same frames, German verbs like *zeigen* (*show*) are also analyzed as processes that do not require a sentient first participant. Nevertheless, such verbs of evidence in combination with inanimate Subjects are used differently in German and English and have proven to be a challenge in English to German translations (Serbina 2015; Freiwald 2016). This means that either the semantic analyses in FrameNet are too reliant on semantic relations in English or the contrastive differences regarding Subject-Verb combinations go beyond agency and sentience.

The annotations of Subject identifiability were based on Kunz (2010), which includes a detailed account of nominal co-reference in English and German. The two languages draw on the same linguistic resources to signal identifiability of first mention referents, namely definite articles, and proper nouns (Kunz 2010: 118). Besides, English and German can both use demonstrative determiners to refer to either previously mentioned or situationally identifiable referents (Kunz 2010: 135ff.), and textually or situationally evoked references can be established through personal pronouns as well as possessive determiners (Kunz 2010: 61ff.). Kunz makes a distinction between different kinds of identifiability, namely mental, situational, and textual identifiability and she also distinguishes between several sub-categories of each of the markers of identifiability. For the purposes of this thesis, these detailed distinctions were not made. Instead, Subjects that include these definite, inferable, and evoked references were analyzed as identifiable, while non-identifiable Subjects were analyzed based on their indefiniteness. Table 1 lists all linguistic resources to signal identifiability and non-identifiability in English and German (Subjects in bold). Subject Themes that are non-referential were not analyzed in terms of identifiability and instead labelled as *Other*. Such was the case for non-referential Subjects including existential processes as well as cleft and pseudo-cleft constructions. The majority of these Subjects are realized by the personal pronoun *es* (*it*) in German and *it* in English. In both languages, the third person singular nominative pronoun can be used as a semantically empty non-referential Subject but also as a referential personal pronoun for nouns with neuter grammatical gender. Nevertheless, all *es*- and *it*-Subjects were placed in the

category *Other* since this distinction was often not possible due to the lack of context. Moreover, Process Subjects in the form of rank-shifted embedded clauses were also categorized as *Other*.

Linguistic feature		English	German	Translation gloss
Identifiable referents	Definite article	The charging duration is about 6 hours. (E2G_INSTR_007)	Die Ladedauer beträgt ungefähr sechs Stunden.	' the charging.duration is about six hours.'
	Demonstrative determiner	This southern part of the city is good for moderately priced restaurants [...]. (E2G_TOU_001)	Dieser südliche Teil der Stadt beherbergt preiswerte Restaurants [...].	' this southern part of.the city houses cheap restaurants [...].'
	Possessive determiner	Your typing replaces any text currently in the search bar. (E2G_INSTR_006)	Ihr Tippen ersetzt jeden Text, der gerade in der Suchleiste ist.	' your typing replaces any text that currently in the search.bar is.'
	Proper noun	Big Sims called for the check. (E2G_FICTION_003)	Big Sims verlangte die Rechnung.	' Big Sims called.for the check.'
	Personal pronoun	You have asked us about the future. (E2G_SPEECH_004)	Sie haben uns nach der Zukunft gefragt.	' you have us about the future asked.'
	Demonstrative pronoun	This is not a zero-sum game. (E2G_SPEECH_005)	Dies ist kein Nullsummenspiel.	' this is no zero.sum.game.'
Non-identif. referents	Indefinite article	A few minutes went by. (E2G_FICTION_010)	Ein paar Minuten vergingen.	' a few minutes passed.'
	No article	Ferry ports link up with the National Motorway System. (E2G_TOU_008)	Fährhäfen sind am landesweiten Autobahnnetz angeschlossen.	' Ferry.ports are with.the national motorway.system linked.'

Table 1 Linguistic signals of identifiability and non-identifiability in English and German

This distinction between identifiable, non-identifiable, and Other allowed the majority of Subject Themes to be clearly sorted into one of these categories. That being said, there were still plenty of ambiguous cases. Adverbs such as *all/alle*, *everybody/jeder*, and *no one/niemand* do not have immediate referents but invoke entire groups. The decision was made to analyze them as identifiable nevertheless, since they do not represent information that is newly introduced to the text. The same decision was made for the German pronoun *man* and its English equivalents *you* and *one*. The analysis of *one* had to be based on context since it can be used as a general category of 'all people', in which case it was

analyzed as identifiable, or it could be used as numeral with an elliptical nominal group, in which case it was analyzed as non-identifiable (see example (89)).

(89) *The apes had already emptied their chamber-pots on to the dung-heap and rinsed them out under the pump. [...] Now and then, **one** would gesticulate in that measured, urgent fashion of theirs and another would nod or shake its well-brushed head or answer with a little dance of fingers.*

[E2G_FICTION_007]

Theme lengths in the original subcorpora were automatically calculated in Excel based on the number of characters. Only the lengths of experiential Themes were considered. In the Subject hypothesis, where Themes can consist of more than one experiential Theme, only the length of the Subject Themes entered the logistic regressions. Theme length was divided into four categories, namely short, middle, long, and very long Themes. The first three categories represent the three tertiles of the length distribution based on the median. Very long Themes represent outliers, meaning Themes that are 1.5*IQR longer than the category of long Themes. There are no very short Themes as outliers in the other direction would mathematically represent negative characters.

There are two main reasons why length was measured in characters rather than words: The median number of words of experiential Themes is only two in both languages, which leaves little room for categories like short and medium. Furthermore, there are systematic spelling differences between English and German, especially in the case of compound nouns, which are usually spelled together in German and often separated by a space in English. In the statistical models on translation shifts, the separation of a compound noun in English translations would then be marked as a Theme change, even though the translator only followed the spelling conventions of the target language. The length categories in the translations were based on the length categories of the original sub-corpora. Hence, the division between short and medium experiential Themes in German translations is the same as the division in German originals. This reduces the impact of contrastive spelling differences. For reference, Table 2 shows the length divisions in both languages in characters. Circumstance and Complement Themes are generally longer than Subject Themes, and German constituents are generally longer than English constituents except for short German Subject Themes.

	EO		GO	
	First Experiential Theme	Subject Theme	First Experiential Theme	Subject Theme
Short	1-7	1-4	1-8	1-3
Medium	8-18	5-14	9-24	4-19
Long	19-53	15-40	25-68	50-57
Very long	54+	41+	69+	58+

Table 2 Length categories in EO and GO based on characters

6.4 Statistical testing

In Section 5.11, two sets of hypotheses were formulated: The first regards general differences between the original and translation corpora, while the second makes claims on the influence of individual Theme aspects on translation shifts. In order to assess the accuracy of the differences and gauge the impact of each Theme-related category on translations, statistical tests were conducted. No hypotheses were formulated regarding the intralingual analyses and the contrastive analysis. Nevertheless, these chapters will also include some statistical tests to determine whether register differences or differences between English and German are statistically significant.

All statistical tests are based on generalized linear mixed-effects models and were carried out with the help of R and R Studio (R Core Team 2017). Four R packages were required to run the models and test their accuracies: lme4 (Bates et al. 2015), afex (Singmann et al. 2017), MuMIn (Bartoń 2018) and blmeco (Korner-Nievergelt et al. 2015). Generalized mixed models were used because they allow the inclusion of random effects, which was crucial for this data to account for repeated measures from the same texts and thus author idiosyncrasies.

To calculate the effect sizes and p-values of individual Theme aspects with regard to translation shifts, I used the glmer function of the lme4 package (Bates et al. 2015). This test calculates the effect of several predictor variables on a predicted variable against one reference variable of the same super-category. In all of these tests, the predicted variable is Change and the effects of different predictor variables such as Theme number, Participant Theme, and Subject Theme sentience are calculated and tested for statistical significance. Each of these predictor variables consists of a variety of sub-types, for example Actor, Senser, and Carrier Themes in the Participant Theme category. One of these sub-types needs to be chosen as a reference type with which all other sub-types are compared.

The test then calculates the size of the difference between each sub-type and the reference and reports how likely this difference is due to chance, expressed through the p-value. If the p-value is below 0.05, it means that the reported difference is not coincidental with a likelihood of 95% or more, which is considered statistically significant. Here is an example of one of these glmer functions carried out in R Studio:

```
glmer(ChangeBinary ~ scale(Theme.number) + Length + Textual.Theme + Interpersonal.Theme + ExpTheme.Type + Participant.Theme + Register + Identifiability + Sentience + (1|Textfile), family = "binomial")
```

Nine predictor variables entered the model. The choice of the reference types for each of these variables is very consequential for the test results and should therefore be considered closely. Since each model is supposed to gauge the effects on translation changes, I generally tried to choose a reference type which represents an unproblematic Theme type, which implies that a significant, positive deviation from this reference is a more challenging type of Theme to translate. These choices are largely based on the categories that were already identified as less problematic in the hypotheses: medium-sized Themes for *Length*, Subject Themes for *ExpTheme.Type*, identifiable Subject Themes for *Identifiability*, and Human Subjects for *Sentience*. The variables *Textual.Theme* and *Interpersonal.Theme* only distinguish two types, namely the existence or absence of a textual or interpersonal Theme with absence being the reference category. Process type and participant Theme distributions are similar in EO and GO and Actor and Carrier Themes are the most common in both languages. Carrier Themes were ultimately chosen as the reference type for *Participant.Themes* because relational processes increased in number in both translation directions, which points towards Carrier Themes being non-problematic for translators. Since no hypotheses were formulated regarding the different registers, and no register was immediately obvious as a more or less challenging register to translate in terms of Theme, the choice of reference type was random and fell onto FICTION. *Theme.number* did not require a reference category because it is a numerical variable, where the effects are calculated relative to increase in size.

The predicted variable Change was calculated in two different ways: binary Change and numerical Change. Regressions based on binary Change are binomial models while regressions using numerical Change represent Poisson models. In Excel, the Theme in the

originals was compared to the Theme in translations for each of the eight Theme categories, mentioned above, excluding *Register*, which could not change. Binary Change only distinguishes between Themes that remain completely unchanged in all categories and Themes that underwent some kind of change. Given the high number of Theme categories, the majority of Themes were changed in some way (between 48.4% and 68.6% depending on translation direction and Theme hypothesis). For numerical Change, each change instance was given a value of one and these values were added up to arrive at the total change value. Thus, in theory, a numerical Change of up to eight was possible (since a maximum of eight Theme categories could change between the original and translated clause). In practice, the highest numerical Change was seven, which occurred only once in German to English translations in the Subject hypothesis (see example (90)).

(90)

GO: *Plötzlich steht vor dem Parlament eine Gruppe hitzköpfiger Studenten.*
'suddenly stands in front of the parliament a group of hotheaded students.'

ET: *In your mind's eye, you see a group of students in heated discussions in front of the Parliament building.*

[G2E_TOU_020]

Assigning a value to change is inherently problematic. This is due to the fact that there is no way of objectively measuring whether one kind of change weighs more or less heavily than another. For example, a change from a Complement Theme to a Subject Theme or from a Human Theme to a nonconcrete Inanimate Theme may be much more severe than a change from a medium Theme to a small Theme, at least in the eyes of the translator or the reader. Nevertheless, I did not want to arbitrarily assign different values to different change types. And while I do consider a change calculation problematic, I also think that a graduation of change is crucial in order to assess whether a Theme feature is consistently associated with one particular type of change or whether multiple aspects are impacted in the translation. For this reason, each logistic regression was run twice, one with binary and the other with numerical Change as predictor variables. I also conducted separate tests for each translation direction and for each of the three Theme combinations. This totals twelve generalized mixed models in the inferential analysis of Theme in translations.

One advantage of Change as a numerical category is that it facilitates model convergence. Three models were not able to converge with all nine predictor variables included,

which are all the models in translations from German to English that use binary Change as the predicted variable. One of the reasons why these issues only occur in translations into English is likely that more Theme information is missing in the German originals. In the Forefield hypothesis, many German Themes do not contain an experiential Theme, so any variable that is tied to experiential Themes like Theme markedness or Participant Theme is not available. In the first experiential element hypothesis, experiential Theme measures are, by definition, considered. However, since so many Themes in German do not begin with a Subject, the two Subject Theme measures, Subject sentience, and Subject identifiability, are also frequently undefined. That being said, the model with the most converging issues is based on the Subject hypothesis and binary Change, which indicates that there must be additional reasons. It is also apparent that a binary distinction between *Change* and *No Change* is difficult to assess in linear mixed regression since the exact same models were able to converge if the predicted variable was numerical Change. To successfully converge the three models that use binary Change, some of the predictor variables had to be excluded: *Sentience* in the Forefield and first experiential element hypothesis and *Marked.Themes* and *Register* in the Subject Hypothesis.

In addition, I used the function `r.squaredGLMM` of the MuMIN package (Bartoń 2018) to calculate how well each of the linear mixed models describe Theme Change in reference to a null model. The calculated delta values are between 0.1893 and 0.9006, which means that there are considerable discrepancies regarding the explanatory powers of the different models. Nevertheless, the values are generally high enough to consider the models meaningful in terms of Theme Change projection. The function `vif.mer` (Frank 2014) was used to assess the risk of multicollinearity in the regression models. Three variables repeatedly had elevated levels, namely *Theme.number*, *Textual.Theme*, and *Interpersonal.Theme*. This result is hardly surprising since any case of increased Theme numbers in the Forefield and first experiential element hypothesis must necessarily be tied to a textual or an interpersonal Theme, which is why these variables largely predict the same cases of change.

The second series of statistical tests was carried with the help of the mixed function of the `afex` package (Singmann et al. 2017). With this function, p-values of the generalized mixed models were calculated with a likelihood ratio test to assess the explanatory power of entire categories with regard to the predicted variable. This test was employed for the

general analysis of the translation data and also for the intralingual and contrastive analyses. For example, the test was used to assess how well Register in GO predicts the frequency of marked Themes or whether the difference in Theme number between EO and GT is statistically significant. This test was preferred over a conventional t-test or χ^2 test since it also allows the inclusion of random effects.

```
mixed(Theme.number ~ Register + (1|Textfile), family = "Poisson", method = "LRT")
```

7 Theme in German originals

7.1 Results

In this chapter, the results of Theme structures in German originals will be presented. A variety of Theme-related aspects were examined including multiple Themes, marked Themes, participant Themes, process types, circumstance distribution, Subject sentience, and Subject identifiability. Three German Theme hypotheses have been considered: The Forefield hypothesis, the first experiential element hypothesis, and the Subject hypothesis. Some analyses are heavily dependent on the choice of Theme hypothesis, in particular multiple Themes and marked Themes, while other analyses like process types or circumstance distribution are, for the most part, unaffected by the Theme hypothesis. To ensure clarity and avoid redundancy, the differences in Theme hypothesis will thus only be specifically contrasted for multiple Themes and marked Themes. For all other analyses, the results of only one of the three hypotheses will be presented. The choice of Theme hypothesis was based on which hypothesis best displayed the feature in question.⁷¹ That being said, the relative frequencies for all analyses are almost identical between the three hypotheses and it would have made very little difference if other hypotheses were chosen for each of the analyses. In the few cases where there were small differences between the Theme hypotheses, they will be pointed out in the running text. All tables and figures illustrating the results of the other Theme hypotheses are included in the Appendix.

⁷¹ The data for Subject Agency and Subject identifiability will come from the Subject hypothesis, given that unlike in the other two hypotheses, the Subject of every clause is part of the analysis. The Subject hypothesis will also be used for the process type and participant role analysis for the same reasons. The other two hypotheses do not have to have a participant role in their Theme, which means that in the relevant analyses information on process types is often missing. This is the case for 1743 clauses for the Forefield hypothesis and 1384 clauses for the first experiential element hypothesis, which represents 35.6% and 28.2% of the entire data set respectively. Process type distribution analysis based on these hypotheses would be incomplete. For circumstance Themes, I used the first experiential element hypothesis. The Subject hypothesis captures the same circumstance Themes as the first experiential element hypothesis and potentially even additional ones if multiple circumstances are positioned before the Subject. However, in the case of a circumstance Theme, I wanted to highlight the one circumstance that was chosen above all other options as the primary experiential point of departure either as the Forefield or as the very first element in the Midfield.

7.1.1 Multiple Themes and Theme Markedness

7.1.1.1 Forefield hypothesis

	Overall		FICTION		INSTR		SPEECH		TOU	
Total Clauses	4901		1306		755		1580		1261	
Total Themes	5276		1421		758		1772		1327	
Single Themes	4544	92.7%	1195	91.5%	752	99.6%	1401	88.7%	1196	94.8%
Multiple Themes	357	7.3%	111	8.5%	3	0.4%	179	11.3%	65	5.2%
Average # of Themes	1.08		1.09		1.00		1.12		1.05	
Experiential Themes	4475	84.8%	1169	82.3%	682	90.0%	1415	79.9%	1210	91.2%
Subject Themes	2842	63.5%	796	68.1%	379	55.6%	1002	70.8%	665	55.0%
Circumstance Themes	1255	28.0%	288	24.6%	273	40.0%	306	21.6%	389	32.1%
Complement Themes	364	8.1%	81	6.9%	30	4.4%	106	7.5%	147	12.1%
Predicator Themes	13	0.3%	4	0.3%	0%	0%	0	0%	9	0.7%
Textual Themes	528	10.0%	165	11.6%	51	6.7%	218	12.3%	94	7.1%
Interpersonal Themes	234	4.4%	74	5.2%	22	2.9%	117	6.6%	22	1.7%
Cleft	39	0.7%	13	0.9%	3	0.4%	22	1.2%	1	0.1%

Table 3 Basic Theme distribution in GO (Forefield)

Table 3 shows a summary of the Theme distributions of the Forefield hypothesis in GO. The Forefield hypothesis considers only the Forefield, the typological field preceding the Finite, to be the Theme in German. Since the Finite is generally in second position, this field is typically restricted to only one element. This means that if the first element is a textual or interpersonal element, the Theme can lack any experiential meaning. Some columns are divided into two numbers. The first represents absolute frequencies and the second relative distribution. For example, the tourism register included 1210 experiential Themes, which represents 91.2% of all Themes in total.

4901 original German clauses were analyzed. The number of clauses per register varies considerably. There are different reasons for this, which will be outlined in the following. The sizes of the different registers in the CroCo corpus are based on the number of words, so if a register has a higher average number of words per clause, the number of total clauses is smaller. Above that, since only the very first T-unit per sentence was analyzed, there will be fewer analyzed clauses if a register has a high number of paratactic

clause complexes. Lastly, and most importantly, only declarative clauses were considered in the analyses, and the portion of declarative clauses also varies across registers. This is the primary reason why the number of clauses in INSTR is the lowest because this register includes a high number of imperative clauses.

Given the strong Finite-second constraint, it is not surprising that the majority of Themes are simple Themes in the Forefield position. Nevertheless, 7.3% of the Themes are multiple despite the relatively fixed position of the Finite verb in German. Most multiple Themes include an element that scholars like Brinkmann (1971) consider part of the left-outfield. The majority of these are textual Themes like *und* (*and*), which allow an additional Forefield element. However, other, more uncommon Theme combinations, such as multiple circumstance Themes, are possible as well (see example (91); circumstance Themes in bold).

- (91) *1942, | im Dezember, | an einem ungewöhnlich kalten Tag, | spätnachmittags, war er nach Ochsenzoll, wo die SS-Kasernen lagen, hinausgefahren.*
'1942, | in December, | on an unusually cold day, | late.afternoon, was he to Ochsenzoll, where the SS-barracks are, driving.out.'

[G2E_FICTION_008]

The number of multiple Themes varies considerably among the four registers. Only 0.4% of Themes in INSTR are multiple, which is connected to the low number of textual Themes in the register. The highest relative frequency of multiple Themes can be found in SPEECH, which is due to the frequent use of the conjunction *denn* (*for*), which does not exhaust the Forefield position, and the high number of Vocatives which are also not immediately followed by the finite verb (see example (92); Vocative underlined, Subject Theme in bold). Despite this variation, Register is only a marginally significant predictor of Theme number in the Forefield hypothesis ($\chi^2 = 7.69$, $df = 3$, $p\text{-value} = 0.053$).

- (92) *Meine Damen und Herren, auch ich hätte mir eine mutigere erneute Verpflichtung der deutschen Wirtschaft vorstellen können.*
'my ladies and gentlemen, also I have [refl-1sg] a braver anew obligation of.the German economy imagine could.'

[G2E_SPEECH_008]

Unsurprisingly, experiential Themes are the most common Theme type, followed by textual Themes and interpersonal Themes. FICTION and SPEECH demonstrate a similarly

high number of textual and interpersonal Themes, while TOU and INSTR display similarly low relative distributions. The most common, most unmarked type of experiential Theme with 63% overall is the Subject Theme, followed by circumstance Themes with 27.8%, Complement Themes with 8.1% and lastly Predicator Themes with only 0.3% (see example (93); Predicator Theme in bold). Not only are Predicator Themes very rare, but they are also restricted to FICTION and TOU. The percentages found here are slightly higher than the 0.13% reported by Uzonyi and Dabóczy (2016: 56).

- (93) *Aber **rummachen** konnte man.*
*'but **making.out** could you.'*
 [G2E_FICTION_007]

The most common experiential Theme in GO are Subject Themes, which make up 63.5% of experiential Themes. 58.0% of German Forefield include a Subject, which is considerably lower than the 65% reported by Winter (1961). However, Winter (1961) also analyzes some elements as Subjects that are considered Complements in SFL. Marked Themes are also distributed quite dissimilarly across the registers. Circumstance Themes are a lot more common in INSTR and TOU. As will be discussed in detail later, this high number of circumstances in INSTR is due to the frequent use of Condition. Clauses in TOU are often contextualized in space and time, either to direct the reader to a certain location or describe a historical fact. This is partly also the reason for the high number of Complement Themes in this register, which are often circumstantial Attribute Complements of space (see example (94); Complement Themes in bold). Another common type of Complement Theme in TOU is the Existent Complement, which often precedes the existential process (see example (95)). Register is a significant predictor of Theme markedness ($\chi^2 = 16.29$, $df = 3$, $p\text{-value} < 0.001$).

- (94) ***Auf dem Spielplan** stehen vor allem klassische Stücke, allerdings überwiegend in modernen Inszenierungen.*
*'**on the programme** are above all classical pieces, however predominantly in modern orchestration.'*
 [G2E_TOU_006]

- (95) ***Gastronomie** gibt's von 10 bis 1 Uhr*
*'**gastronomy** is.there from 10 to 1 o'clock.'*
 [G2E_TOU_022]

7.1.1.2 First experiential element hypothesis

	Overall		FICTION		INSTR		SPEECH		TOU	
Total Clauses	4901		1306		755		1580		1261	
Total Themes	6083		1680		904		2063		1438	
Single Themes	4204	85.8%	1093	83.7%	683	90.5%	1271	80.4%	1157	91.8%
Multiple Themes	697	14.2%	213	16.3%	72	9.5%	309	19.6%	104	8.2%
Avg. # of Themes	1.24		1.29		1.20		1.31		1.14	
Avg. # of Themes w/o Finite	1.16		1.18		1.10		1.21		1.09	
Experiential Themes	4863	79.9%	1294	77.0%	752	83.2%	1558	75.5%	1260	87.6%
Subject Themes	3200	65.8%	918	70.9%	439	58.4%	1128	72.4%	715	56.7%
Circumstance Themes	1259	25.9%	282	21.8%	279	37.1%	313	20.1%	386	30.6%
Complement Themes	387	8.0%	89	6.9%	32	4.3%	116	7.4%	150	11.9%
Predicator Themes	16	0.3%	5	0.4%	2	0.3%	0	0%	9	0.7%
Textual Themes	527	8.7%	163	9.7%	55	6.1%	217	10.5%	92	6.4%
Interpersonal Themes	655	10.8%	211	12.6%	94	10.4%	266	12.9%	85	5.9%
Finite Themes	416	63.5%	139	65.9%	70	74.5%	149	56.0%	58	68.2%
Modal Adjuncts	239	36.5%	72	34.1%	24	25.5%	117	44.0%	27	31.8%
Cleft	38	0.6	12	0.7%	3	0.3%	22	1.1%	1	0.1%

Table 4 Basic Theme distribution in GO (1st exp. element)

Table 4 shows the same Theme distribution for the first experiential element hypothesis. In the first experiential element hypothesis, everything up to and including the first experiential element is considered the Theme of the clause. Since many clauses in German begin with an experiential element, the Theme of the first experiential element hypothesis is often identical with the Theme of the Forefield hypothesis. However, in the case that the Forefield is occupied entirely by a modal or cohesive Adjunct, the first experiential element is positioned after the finite verb in the Midfield. In these cases, the Theme of the first experiential element hypothesis extends past the Finite. In the majority of cases, the Theme number is thus either the same or higher compared to the Forefield hypothesis. And yet there are a few cases like example 91 where the Forefield features multiple experiential elements, which would only count as a simple Theme in the first experiential element hypothesis.

The number of total Themes has increased, and the number of multiple Themes nearly doubled compared to the previous hypothesis. Above that, the number of interpersonal Themes has increased considerably, which is largely due to the high number of obligatory

Finite Themes. The relative frequencies of multiple Themes have increased evenly across the four registers, with the only exception being INSTR. Under the previous hypothesis, this register had by far the lowest number of multiple Themes but in this hypothesis, the relative frequency jumped from 0.4% to 9.5%, thus surpassing TOU. This significant increase in multiple Themes can be explained by the infrequent use of conjunctions as textual Themes, since fronted conjunctions are the most commonly type of Theme included in multiple Themes in the Forefield hypothesis. Other types of textual Themes like cohesive Adjuncts are relatively common in INSTR, which is why the number of multiple Themes is now closer to the other registers. Example (96), taken from the instruction manual register, shows one of these common Theme structures, where the Theme includes a cohesive Adjunct (in bold), followed by the Finite, followed by the Subject (cohesive Adjunct in bold). Register is a significant predictor of Theme number in the first experiential element hypothesis ($\chi^2 = 19.18$, $df = 3$, $p\text{-value} < 0.001$).

- (96) ***Andernfalls*** | *besteht* | *die Gefahr eines Stromschlags oder anderer Personenschäden.*
'otherwise | *exists* | *the danger of an electric shock or other personal injuries.'*
 [G2E_INSTR_002]

The distribution of Theme types has remained largely unchanged. The relative frequency of Subject Themes has slightly increased and in turn the frequency of circumstance Themes has slightly decreased. The reason for this difference is the increased likelihood of Subjects to be the very first element in the Midfield if they do not already occupy the Forefield (Müller 1999: 11). While it is not ungrammatical to have a circumstance immediately follow the finite verb in German, this choice must typically be motivated by context. In example (97), the circumstance (underlined) precedes the Subject (in bold) in the Midfield, presumably to focus on the German banking sector or set up a contrast to earlier sentences.

- (97) *Wie Sie wissen, hat im deutschen Bankensektor **der Tuningprozess** [...] bereits eingesetzt.*
*'as you know, has in the German banking sector **the tuning process** [...] already begun.'*
 [G2E_SPEECH_004]

For the same reason, the absolute number of textual Themes and modal Adjunct Themes has remained almost identical because they are also usually not positioned at the peak of the Midfield. Register continues to significantly predict Theme markedness in the first experiential element hypothesis ($\chi^2 = 24.27$, $df = 3$, $p\text{-value} = 2.197e\text{-}05$).

7.1.1.3 Subject hypothesis

	Overall		FICTION		INSTR		SPEECH		TOU	
Total Clauses	4901		1306		755		1580		1261	
Total Themes	9625		2463		1546		2997		2623	
Single Themes	2693	54.9%	764	58.5%	381	50.5%	909	57.5%	639	50.7%
Multiple Themes	2208	45.1%	542	41.5%	374	49.5%	671	42.5%	622	49.3%
Avg. # of Themes	1.96		1.89		2.05		1.90		2.08	
Avg. # of Themes w/o Finite	1.56		1.51		1.56		1.55		1.61	
Experiential Themes	6729		1717		1086		2066		1862	
Subject Themes	4855	72.2%	1292	75.2%	752	69.2%	1554	75.2%	1258	67.6%
Circumstance Themes	1370	20.4%	304	17.7%	295	27.2%	340	16.5%	432	23.2%
Complement Themes	425	6.3%	100	5.8%	33	3.0%	133	6.4%	159	8.5%
Predicator Themes	33	0.5%	7	0.4%	3	0.3%	13	0.6%	10	0.5%
Textual Themes	566	5.9%	168	6.8%	58	3.8%	239	8.0%	101	3.9%
Interpersonal Themes	2330	24.2%	578	23.5%	402	26.0%	692	23.1%	660	25.2%
Finite Themes	2001	85.9%	494	85.5%	367	91.3%	549	79.3%	592	89.7%
Modal Adjuncts	329	14.1%	84	14.5%	35	8.7%	143	20.7%	68	10.3%
Cleft	46	0.7%	14	0.8%	3	0.3%	26	1.3%	3	0.2%

Table 5 Basic Theme distribution in GO (Subject)

Table 5 shows the Theme distributions following the Subject hypothesis, where everything is considered thematic up to and including the Subject of the clause. Compared to the other two hypotheses, the number of Theme elements per clause is either identical if the Subject occupies the Forefield (in the case of the Forefield hypothesis) or if the Subject is the very first experiential element of the clause, which are the two most likely cases in German. In all other cases, the Theme element number is necessarily larger. It is rare that Subjects are preceded by another element in the Midfield, though not impossible. The

most likely case of a double experiential Theme is when a German clause includes a circumstance in the Forefield, in which case the Subject must move to a post-Finite position.

As a consequence of this extension of the Theme, the number of multiple Themes now increases noticeably to 45.1%. Accordingly, the average number of Theme elements both with and without the Finite increases as well. Given that every clause must include a Subject Theme, the number of experiential Themes now exceeds the number of total clauses, representing all the cases of multiple experiential elements in the Theme. The reason why the number of Subject Themes is not identical to the number of clauses comes from the fact that clefts were analyzed separately.

Furthermore, the distribution of textual, interpersonal, and experiential Themes differs significantly compared to the other hypotheses. The relative frequencies of textual Themes in particular decreases noticeably. The absolute numbers of textual Themes, of course, increases as well but they do not increase at the same rate as experiential and interpersonal Themes. This is due to the fact that it is not common to commence the Midfield in German with a textual element, as most textual elements are placed in the Forefield.

The two registers that previously had the lowest numbers of multiple Themes, INSTR and TOU now boast the highest proportions of multiple Themes and average Theme numbers overall. This has to do with the high number of circumstances used in both of these registers. What is interesting to note is that the relative frequencies of multiple Themes differ between the registers while the average number of Themes remains relatively constant. This goes to show that each register favors different elements to be positioned before the Subject but the likelihood for any element to precede the Subject is generally the same. Register significantly predicts Theme number ($\chi^2 = 8.37$, $df = 3$, $p\text{-value} = 0.0390$) as well as Theme markedness ($\chi^2 = 26.23$, $df = 3$, $p\text{-value} = 8.528e-06$) in the Subject hypothesis.

7.1.2 Participant Themes and process types

	Overall	FICTION	INSTR	SPEECH	TOU
Total processes	4855	1292	752	1554	1258
Material	1942	444	460	500	538
Subject Actor	1449	405	237	401	406
Subject Goal	397	18	212	77	90
Subject Initiator	72	19	5	17	31
Complement Goal	84	13	17	19	35
Other	64	10	10	20	24
Mental	590	194	33	261	102
Subject Senser	486	169	20	233	64
Subject Phenomenon	104	25	13	28	38
Complement Phenomenon	40	23	4	8	5
Complement Senser	18	9	0	5	4
Relational	1643	400	211	545	488
Attributive	1500	390	176	464	471
Subject Carrier	1321	352	169	408	393
Complement Attribute	164	33	5	51	75
Other	15	5	2	5	3
Identifying	322	48	42	137	95
Subject Token	202	30	32	95	45
Subject Value	115	18	10	38	49
Other	8	1	0	6	1
Verbal	329	109	32	144	44
Subject Sayer	290	104	23	127	36
Subject Verbiage	32	3	7	16	6
Complement Verbiage	24	5	1	12	6
Other	21	3	2	13	3
Behavioral	97	68	0	15	14
Subject Behaver	97	68	0	15	14
Existential	146	28	7	44	67
Process	123	26	7	40	50
Complement Existent	23	2	0	4	17
Empty Subjects	117	48	3	45	21
Dummy Subjects	14	3	6	4	1

Table 6 Participants and process types in GO

Table 6 shows the distribution of the absolute frequencies of participant Themes and process types. Process types were analyzed based on the participant roles of their Themes. Since the Subject hypothesis requires each Theme to include the Subject, it is possible to have two or even more participants in the same Theme. This is the reason why the frequencies of each process type do not add up with the numbers of their participant Themes. For each of the process types, a summary of the most common participant roles is presented while infrequent participant roles were summarized under *Other* to increase

clarity. Such roles include, for example, Attributor and Assigner Themes in relational processes or Beneficiary Themes in material processes.

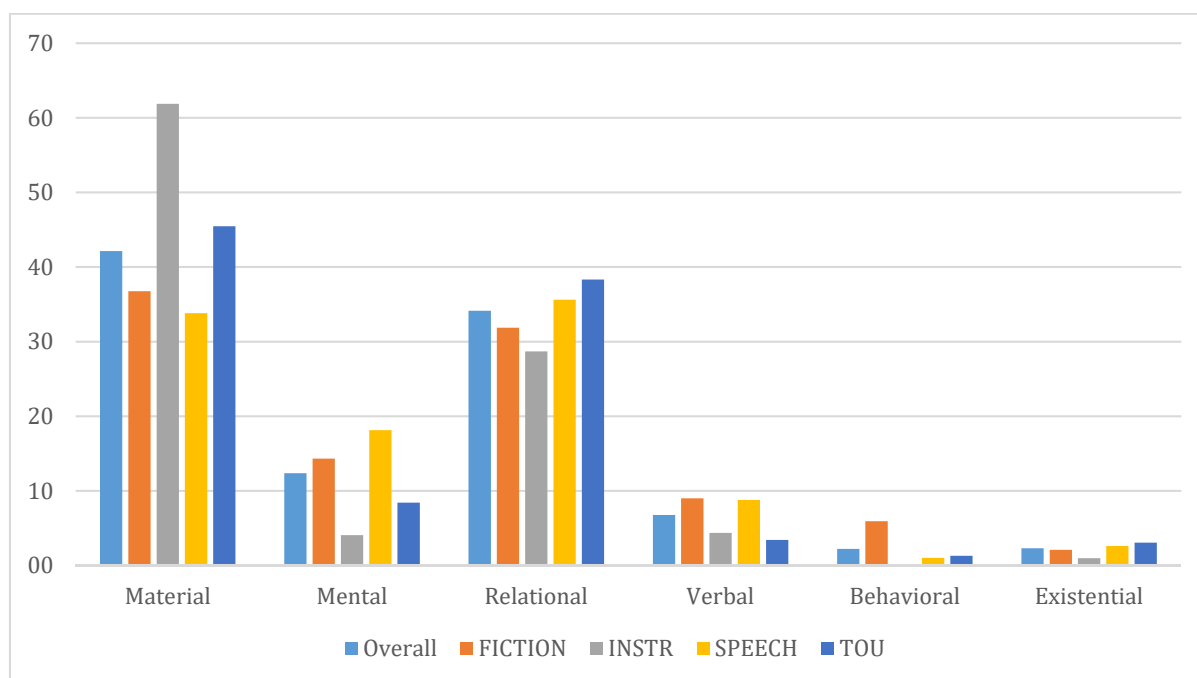


Figure 1 Process type distribution in GO

Figure 1 shows the process type distribution between the registers in relative frequencies. The primary process types, material, relational, and mental processes, are also the most common process types. Material and relational processes are by far the two most common process types both overall and for each individual register. On average, material processes are more common than relational processes, but this can vary between the registers: INSTR and TOU feature more material processes whereas FICTION and SPEECH include more relational processes. The relative frequency of material processes in INSTR is noticeably higher (61.2%) than in all other registers, which explains why INSTR also has the lowest relative frequency for all other types of processes. The relational process is fairly evenly distributed between all four registers and seems to be least dependent on register idiosyncrasies. Mental processes are the third most common process type in all four registers but are primarily used in FICTION and, even more so, in SPEECH. INSTR features very few mental processes (4.4%), approximately the same as verbal processes. Overall, the verbal process is the most common of the subsidiary processes with an average frequency of 6.8%. Compared to the other two subsidiary process types, verbal pro-

cesses are more than three times as common, which calls this distinction between primary and subsidiary process types further into question. Verbal processes are relatively common in all four registers but most common in SPEECH. Behavioral and existential processes have a very similar relative frequency overall. However, while existential processes are similarly common in all registers, behavioral processes only really occur in FICTION and are virtually non-existent in the other registers. Existential processes are distributed evenly, but slightly higher in TOU, where it is more common to talk about entities and places in existence (see example (98)). On the whole, the category Register significantly predicts the choice of participant Themes ($\chi^2 = 13.35$, $df = 3$, $p\text{-value} = 0.00393$).

- (98) *Es gibt das Puppenmuseum, in dem Omas süße, alte Puppen und Opas Blechspielzeug zu bewundern sind, das Spielzeugmuseum und andere mehr.*
'there is the doll.museum, in which grandmother's cute, old dolls and grandfather's tin.toys to admire are, the toy.museum and others more.'

[G2E_TOU_019]

The most common participant roles in Theme position are the first participants like *Senser* in mental processes or *Carrier* in attributive relational processes. In existential processes, it is the process itself that is the most common experiential Theme as they do not have a first participant. Participant Themes are most varied for material processes, which is connected to the fact that material processes also have the highest number of participant roles in general. Material processes also feature more Subject Themes that are not conflated with the first participant compared to the other processes. German mental processes have a relatively high number of first participant realized as Complement Themes. This is due to the fact that in German the Phenomenon can be realized as the Subject of an active clause with the *Senser* being a dative or accusative Object (Steiner 2001b: 3; see example (99); Complement Theme in bold).

- (99) ***Ihnen** ist klar, dass gerade in der Wissensgesellschaft motivierte und leistungsbereite Mitarbeiter das grösste Kapital eines Unternehmens sind.*
'you[dat] is clear that especially in the knowledge.society motivated and hard.working co-workers the greatest asset of.a company are.'

[G2E_SPEECH_005]

Most relational processes are attributive, overall and per register. The number of Complement Themes in attributive relational processes is surprisingly high, higher than for all other process types. These Complement Themes are almost exclusively circumstantial Attribute Themes. In identifying relational processes, the frequencies of Token and Value Themes are relatively similar, which is due to the fact that the two participant roles are easily reversible, and their order is mostly based on context.

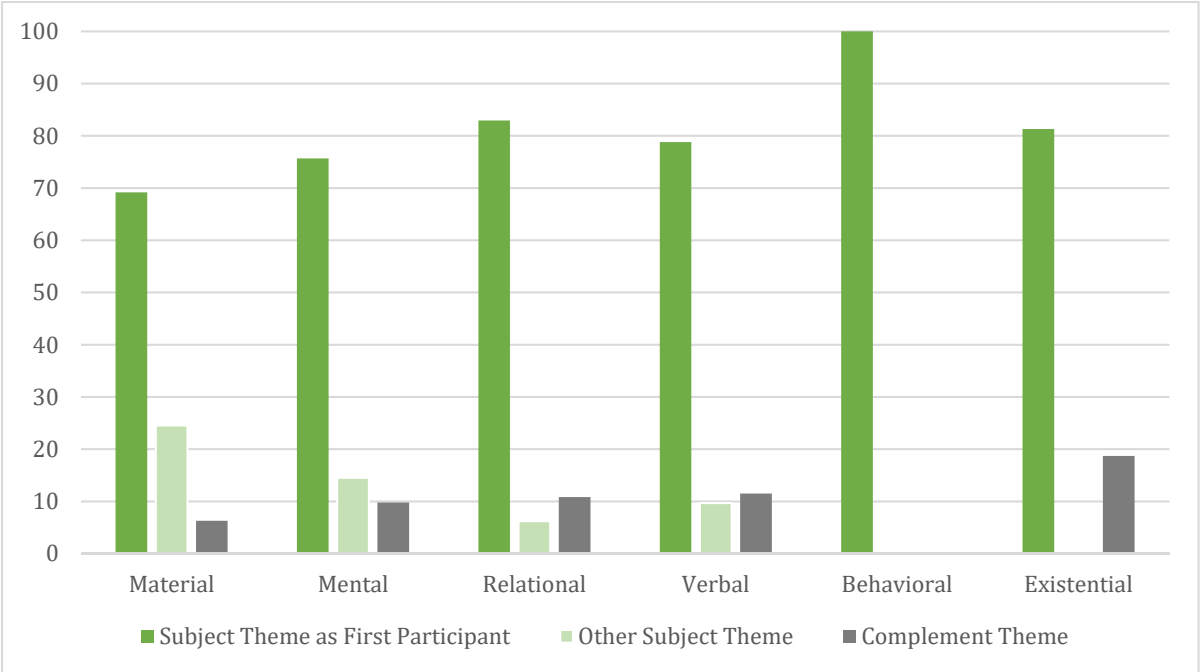


Figure 2 Participant roles and Theme types in GO

Figure 2 shows the distribution of Subject and Complement Themes per process type. These relative frequencies are based on the overall data and are not divided into the different registers. As was pointed out earlier, the most common Theme is the first participant Theme conflated with the Subject in all process types but existential processes, which do not have a first participant.⁷² For behavioral clauses, Behavior conflated with the Subject is the only kind of participant Theme that was found in the entire data set even though a Behavior Subject or Complement Theme is in theory possible. Complement Themes in existential processes were the most common overall. This is likely because the empty *es* Subject has low communicative value and can thus be easily moved out of the Forefield and make room for the Existent Complement in the pre-verbal position (see example (100); Complement Theme in bold).

⁷² For clarity, the process Theme in existential processes, which includes the Subject, was treated as a first participant Subject Theme in this Figure.

(100) ***Infos dazu** gibt es bei den örtlichen Gastgebern und bei der Schwarzwald Tourismus GmbH (STG).*

*'**info about.that** is there at the local hosts and at the Schwarzwald Tourism GmbH (STG).'*

[G2E_TOU_016]

With 22.5% of cases, material processes feature the highest number of Subject Themes that are not mapped onto the first participant. The majority of these cases are Goal Themes. This is mostly due to the high frequency of passive constructions where the Subject can be conflated with the participant undergoing the process. Such constructions are particularly common in material processes because the Goal is actually undergoing an action. While such constructions are grammatical for most other process types as well, a Phenomenon, for example, undergoes the process only in a metaphorical sense as nothing is actually being done to a thought or idea. Another common construction in material processes where the Subject Theme is not conflated with the participant role of Actor are initiating material processes (see example (101); Initiator Theme in bold).⁷³ At the same time, material processes also have the second lowest relative frequency of Complement Themes after behavioral processes. This may be a reciprocal effect since moving a secondary participant in Theme position can be easily accomplished by passive constructions, so that the more marked option of a Complement Theme is not required.

(101) ***Sie** lassen dich hungern.*

*'**they** let you hunger.'*

[G2E_FICTION_002]

Surprisingly, verbal processes have the highest number of Complement Themes after existential processes with 12.7%. However, it should also be noted that Complement Themes are very evenly distributed at around 11% between the four most common process types, which is why these slight differences should not be over-interpreted. Subject Themes that are not the first participant are relatively infrequent in relational processes. This is due to the fact that the majority of relational processes are attributive, and the Attribute cannot be mapped onto the Subject.

⁷³ Given that in an initiating construction, the Initiator arguably becomes the first participant in material processes, such cases could have also been categorized as a first participant Subject Theme.

The data on process types and participant Themes is based on the Subject hypothesis. Compared to the other two Theme hypotheses, there are only very small differences. Naturally, the absolute numbers for all participant Themes increases the further the Theme is extended over the clause. As more elements for Theme are analyzed, the numbers of material processes increase from 35.5% to 40%, while the numbers of relational processes decrease from 40.9% to 33.8%. In fact, in both of the other hypotheses, relational processes are overall more common than material processes and are used most often in all registers except for INSTR. This means that of the 1743 clauses in the Forefield hypothesis and 1384 clauses in the first experiential element hypothesis that are not analyzed in terms of process types, a disproportionate number represents material processes. A similar, though less striking effect can be found for existential processes, which exhibit an increase from 1.6% to 2.5% between the Forefield hypothesis and the Subject hypothesis. This is a seemingly small difference of 0.9% but it is still quite noteworthy given that existential processes are the least common process type overall. Conversely, relational processes are apparently less likely to be paired with non-participant Themes than the other process types. The relative frequencies of mental, verbal, and behavioral processes do not differ in any noticeable way. Furthermore, the distribution of Subject Themes as first participant, Subject Themes as secondary participant, and Complement Themes does not differ substantially for any of the three Theme hypotheses.

7.1.3 Circumstances

In this section, circumstances will be analyzed in detail both in terms of their overall frequency as Themes, their frequency in general, and their likelihood of becoming the Theme if present in a clause. This distinction allows an analysis of Theme markedness from two different perspectives, namely in absolute and in relative terms. As an example, Place Themes are a lot more frequent than Concession Themes and are thus, in absolute terms, also less marked. Nevertheless, this difference in Theme frequency may be solely due to the fact that circumstances of Place are simply more common than circumstances of Concession, irrespective of position. Naturally, if more clauses include circumstances of Place, they also become more likely to be thematic. In relative terms, circumstances of Concession are a less marked option as Theme because if a clause includes a circumstance of Concession, it is the Theme in 58.8% of cases as opposed to the 38.7% of Place. This

calculation of the likelihood of a circumstance to be the Theme of its clause will be called thematic potential (TPot).

	Overall			FICTION			INSTR			SPEECH			TOU		
Total Clauses	4901			1306			755			1580			1261		
Total Circumstances	1259	3235	38.9%	282	716	39.4%	279	625	44.6%	313	920	34%	386	974	39.6%
Circumstances per clause	0.66			0.55			0.83			0.58			0.77		
Likelihood of Circ. Theme	25.7			21.6			37			19.8			30.6		
Additive	15	34	44.1%	1	5	20%	1	5	20%	4	7	57.1%	9	17	52.9%
Behalf	39	85	45.9%	2	5	40%	4	5	80%	12	41	29.3%	21	34	61.8%
Comitative	45	154	29.2%	12	42	28.6%	2	15	13.3%	15	41	36.6%	16	56	28.6%
Concession	20	34	58.8%	12	18	66.7%	0	0	-	7	15	46.7%	1	1	100%
Condition	186	252	73.8%	26	36	72.2%	119	146	81.5%	26	44	59.1%	15	26	57.7%
Comparison	16	55	29.1%	4	30	13.3%	1	1	100%	7	18	38.9%	4	6	66.7%
Distance	2	16	12.5%	0	7	0%	0	0	-	0	1	0%	2	8	25%
Default	0	1	0%	0	0	-	0	0	-	0	1	0%	0	0	-
Degree	1	12	8.3%	0	0	-	0	0	-	0	7	0%	1	5	20%
Duration	35	92	38%	15	34	44.1%	1	9	11.1%	1	13	7.7%	18	36	50%
Frequency	18	48	37.5%	10	18	55.6%	1	5	20%	1	7	14.3%	6	18	33.3%
G guise	19	89	21.3%	6	24	25%	1	15	6.7%	0	20	0%	12	30	40%
Matter	24	51	47.1%	3	6	50%	1	4	25%	15	32	46.9%	5	9	55.6%
Means	88	232	37.9%	7	17	41.2%	33	106	31.1%	37	76	48.7%	12	33	36.4%
Product	0	16	0%	0	1	0%	0	4	0%	0	3	0%	0	8	0%
Place	345	892	38.7%	64	172	37.2%	41	102	40.2%	77	217	35.5%	163	401	40.6%
Purpose	42	134	31.3%	3	16	18.8%	23	44	52.3%	11	46	23.9%	5	28	17.9%
Quality	46	311	14.8%	27	103	26.2%	0	55	0%	6	90	6.7%	13	63	20.6%
Reason	42	109	38.5%	9	27	33.3%	7	16	43.8%	14	44	31.8%	12	22	54.5%
Source	5	13	38.5%	1	2	50%	2	6	33.3%	1	3	33.3%	1	2	50%
Time	267	596	44.8%	78	151	51.7%	41	86	47.7%	78	188	41.5%	70	171	40.9%
Viewpoint	4	9	44.4%	2	2	100%	1	1	100%	1	6	16.7%	0	0	-

Table 7 Circumstances and thematic potential in GO

Table 7 shows a detailed account of all circumstance types, both overall and in Theme position. The first number in each (register) category in each row represents the absolute frequency of that particular circumstance type in Theme position, the second represents the absolute frequency of that circumstance overall, and the last number is the relative likelihood of the circumstance to be chosen as the Theme if it is included in a clause (TPot). For instance, there is a total of 85 circumstances of Behalf in the data, 39 of which are in Theme position. This means that Behalf has a TPot of 45.9%.

In absolute numbers, the distribution of the circumstance types varies substantially. The three most frequent circumstance Themes, Place, Time, and Condition Themes, are more common than all other 19 circumstances combined. Most other circumstance types group somewhere in the middle with 20-50 occurrences each. Six circumstance Themes were found less than ten times overall, and Default and Product Themes do not occur at all.

Furthermore, approximately two thirds of all clauses in GO include at least one circumstance. The relative frequency of circumstance Themes is 25.7% overall, but if a clause does include a circumstance, the likelihood of a circumstance Theme is 38.9%. Hence, even though circumstances are considered a marked Theme in German by some linguists, the difference in probability between a Subject and a circumstance receiving thematic status is only around 60% to 40% if both are present in the clause. TOU and INSTR have a higher average number of circumstances per clause with 0.77 and 0.83 respectively. Circumstances in INSTR also have a higher average thematic potential than in the other three registers with 44.6% compared to the 38.9% average. Even though TOU has more circumstances overall and accordingly also more circumstance Themes, the chance for a circumstance to become the Theme of the clause is identical in comparison to FICTION. FICTION and SPEECH have a similarly low overall number of circumstances and of circumstance Themes compared to the other two registers.

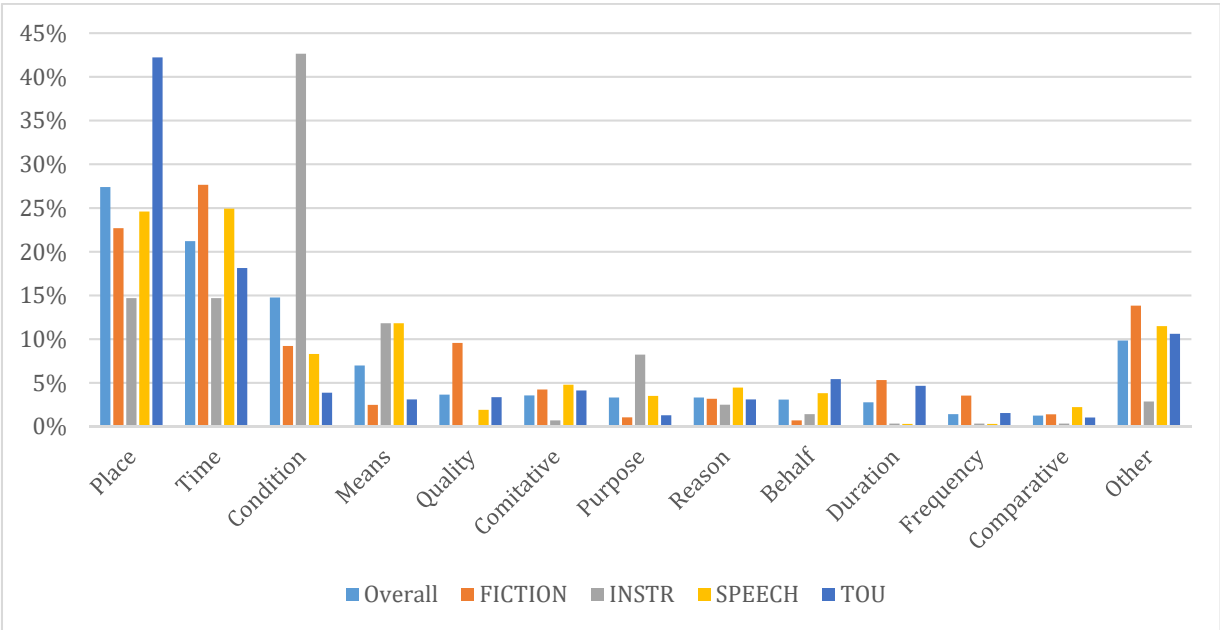


Figure 3 Circumstance Theme distribution in GO

Figure 3 is a visual representation of the relative frequencies of circumstance Themes per register. As an example, Place Themes made up 27.4% of all circumstance Themes overall, but only 14.7% of circumstance Themes in INSTR. The reason why relative frequencies were used over absolute frequencies is because the total number of analyzed clauses per register differed substantially. Only those circumstances that made up at least 2% of all circumstance Themes were included in the graph. All others were summarized under *Other*.

Looking at Figure 3, two interesting aspects stand out: circumstances of Place are the most common type of circumstance Theme overall, but this is largely due to TOU which includes a particularly high number of circumstances of Place. In fact, Time Themes are more common than Place Themes in all of the three other registers. The second striking observation to be made is the tremendously high relative frequency of Condition Themes in INSTR. Condition Themes are more than four times more frequent in instruction manuals than in any of the other registers.

While there are some circumstance Themes that are distributed relatively evenly between the registers, for example Comitative, Behalf, Reason, and to some extent also Time Themes, most other circumstance types are particularly prevalent in one or two registers but not in the others. For example, Means Themes are common in INSTR and SPEECH, Quality Themes are characteristic of FICTION, Purpose Themes particularly common in INSTR, and Duration and Frequency Themes can really only be found in FICTION and TOU. What is also noticeable is that none of the registers group together nicely with any of the other registers. That is to say, FICTION and TOU, for example, are similar in terms of time-related circumstance Themes but are also very different regarding Place and Quality Themes.

Moreover, the register that differs most from all the other registers is INSTR. Regarding relative distribution of circumstance Themes, instruction manuals either score highest or lowest for all circumstance types, except for Behalf Themes. Most noticeably, they feature much lower numbers for the two most common Theme types, Place and Time Themes but include circumstance Themes that are only peripheral in other registers, like Purpose and Means. Their circumstance Theme distribution is also the least diverse out of the four registers, with circumstances in the *Other* category making up only 2.9% of all circumstance Themes in INSTR, compared to the 9.8% average. Register is a significant

predictor of the types of circumstances used as Themes ($\chi^2 = 16.09$, $df = 3$, $p\text{-value} = 0.00108$).

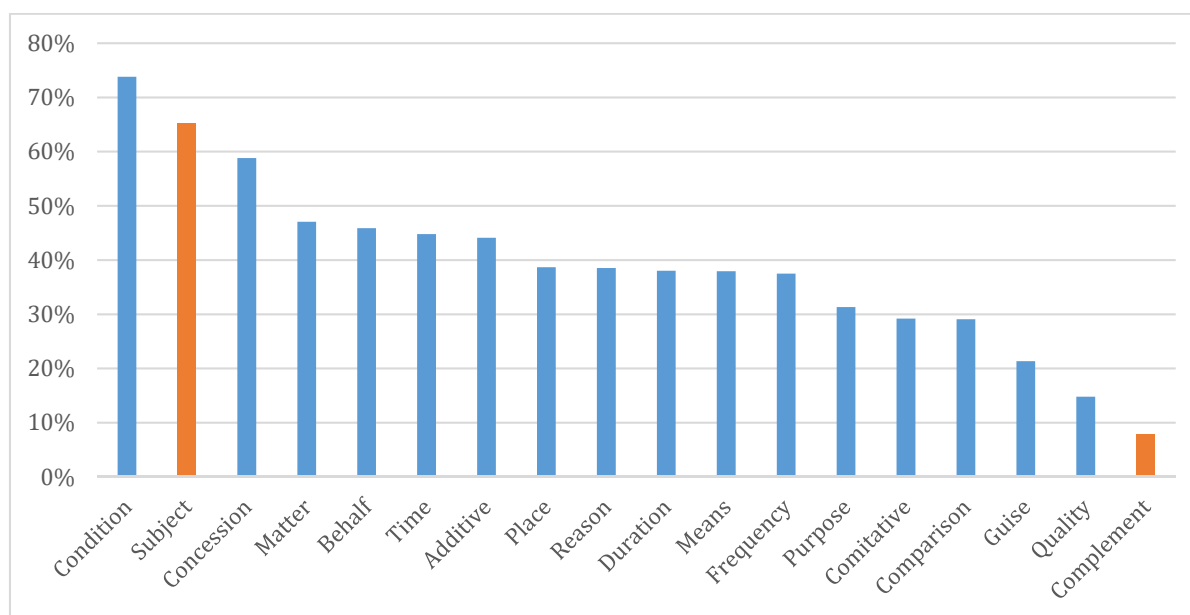


Figure 4 Thematic potential of circumstances in GO overall

Figure 4 shows the thematic potential of circumstances across all four registers. It includes only those circumstances that were found at least 20 times overall. This graph illustrates the likelihood of a particular circumstance to be the Theme of its clause. Accordingly, if a clause includes a circumstance of Reason, it is used as the point of departure of that clause in 38.5% of cases. As a reference point, the TPots of Subjects and Complements were also included.

The two circumstances with the highest TPot are Condition and Concession with 73.8% and 58.8% respectively. In other words, if a clause includes these two circumstances, they are the most likely candidates for Theme. The reason why Subjects have a higher TPot than circumstances of Concession is because there are clauses where only the Subject is a viable option for Theme. Nevertheless, if a clause includes both a Subject and a circumstance of Concession, it is more likely for the Subject to be positioned in the Rheme (see example (102); Concession Theme in bold, Subject underlined).

(102) ***Obwohl es noch warm war von der Tageshitze***, zuckte er zusammen, als ihn der Strahl aus der Handbrause traf.
'***although it still warm was from the day's heat***, flinched he together, as him the stream from the showerhead hit.'

[G2E_FICTION_006]

The majority of circumstances falls between 35% and 45% of TPot. TPot does not seem to be dependent on overall frequency, given that the by far most frequent kinds of circumstances, Place and Time, only have average thematic potentials. In other words, just because a circumstance is generally used more often does not mean that it also has a higher probability to be the point of departure of the clause. The two circumstances with the lowest TPot are Guise (see example (103); circumstance Themes in bold) and Quality (see example (104)). But even the TPot of Quality is still almost twice that of Complements, which makes Complement Themes the by far most marked Theme option in German (besides Predicate Themes of course).

- (103) **Als Pfarrer** betreute er mehrere Gemeinden in Bayern und leitete das Predigerseminar Bayreuth.
'**as priest** supervised he multiple communities in Bavaria and led the preacher.seminar Bayreuth.'
[G2E_TOU_002]

- (104) **Noch stärker** wird nach unseren Vorstellungen die Investmentbranche vom demografischen Wandel profitieren.
'**even stronger** will according.to our imagination the investment.industry from.the demographic change profit.'
[G2E_SPEECH_004]

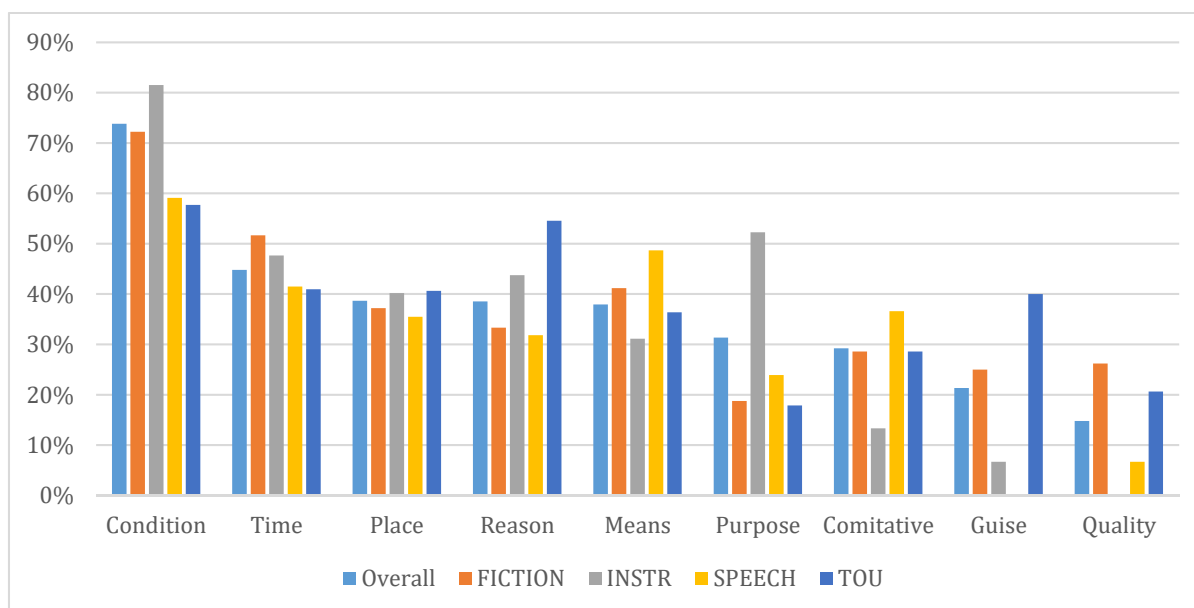


Figure 5 Thematic potential of circumstances in GO overall and per register

Figure 5 shows the same TPot distribution of circumstances, but this time differentiated between the four registers to assess whether register affects the likelihood of a circumstance type to become Theme. For this Figure, only those circumstances were included that were used at least ten times in all four registers to ensure a meaningful comparison between the registers. What becomes immediately apparent is the high number of circumstances that do not meet this threshold. This reveals that the numbers in Figure 4 do not always represent averages across all registers but are more representative of individual registers, for example in the case of circumstances of Matter for SPEECH and Additive for TOU.

For the first few circumstance types, the TPots are reasonably similar among the four registers. What is interesting to note is that there are three circumstances that score a TPot of more than 50% in at least one of the registers, which makes them thematically less marked than Subjects in those registers: Time in FICTION, Reason in TOU, and Purpose in INSTR. Circumstances of Time and Place are remarkably similar across the four registers, which may be related to their universal nature as circumstances.

The four circumstances that have the lowest TPot overall also show the highest degree of divergence between the registers. Circumstances of Purpose have a generally low TPot in FICTION, SPEECH, and TOU yet have a significantly higher probability (52.3%) in INSTR. Conversely, Comitative Theme probability is average in all registers except for INSTR, where only two out of fifteen cases were positioned in the Theme. The distribution of Guise and Quality is the most heterogeneous. Circumstances of Guise are among the

circumstances with the highest TPot in TOU but are used as Theme only once in INSTR and not at all in SPEECH. Similarly, the TPot of Quality is reasonably high in FICTION and TOU but almost zero in the other two registers. Despite there being 55 cases of circumstances of Quality in INSTR, not a single one was used as the point of departure of any clause. Again, circumstance frequency overall seems to be largely independent from TPot. For example, circumstances of Reason are generally more common in SPEECH than in TOU and yet their TPot is noticeably higher in the latter register.

This data on circumstance Themes is based on the first experiential element hypothesis. Between that hypothesis and the Forefield hypothesis, the number of analyzed circumstance Themes change by only four, from 1255 to 1259, which is a remarkably small change even though the Forefield hypothesis does not require an experiential Theme. This difference, or rather the lack thereof, has to do with the structure of the Midfield in German. If the Forefield does not already contain an experiential element, the first experiential element will be positioned at the forefront of the Midfield right behind the Finite verb. However, this position is often reserved for identifiable referents, which makes fronted circumstances a deliberate and motivated choice. Example (105) shows one of these rare circumstance Themes in the Midfield (in bold).

- (105) *Natürlich müssen **in diesen Ländern** noch eine Reihe ganz grundlegender Reformen nicht nur geplant, sondern auch umgesetzt werden.*
*'naturally must **in these countries** still a number entirely fundamental reforms not just planned, but also realised be.'*
[G2E_SPEECH_014]

While this positioning is marked in German, there are not only four but 28 cases of circumstances immediately following the Finite in the first experiential element hypothesis. The reason why the difference is still only four between the two hypotheses is the opposing effect where the Forefield hypothesis actually produces more circumstance Themes than the first experiential element hypothesis. The Forefield hypothesis does not have to include an experiential element but it is also not restricted to only one experiential Theme as the first experiential element hypothesis is. As was pointed out in Section 3.3, a Forefield can include multiple circumstantial elements (see example (106); circumstance Themes in bold), which is the case 23 times (of which one clause had three circumstances

in the Forefield). All but three of these cases include at least one and often several circumstances of Time, which seems to be the most common circumstance type in a multiple circumstance Forefield.

(106) *Heute, | für uns, war Gabor Cziffra in ein gewagt schimmerndes Taubenblau gekleidet.*

'today, | for us, was Gabor Cziffra in a daringly shimmering pigeon.blue dressed.'

[G2E_FICTION_004]

There is a greater increase of circumstance Themes between the first experiential element hypothesis and the Subject hypothesis from 1259 to 1370 cases. These additional 111 circumstance Themes come from clauses that have multiple experiential elements before the Subject of the clause. Despite the difference in absolute numbers, the relative frequencies of individual circumstance Themes do not differ significantly. The biggest difference is a drop of the relative frequency of Condition Themes from 14.8% in the first experiential element hypothesis to 13.6% in the Subject hypothesis. Apart from this, there is nothing else that was noteworthy.

7.1.4 Subject animacy and sentience

	Overall			FICTION			INSTR			SPEECH			TOU		
Total clauses	4901			1306			755			1580			1261		
Relational	1616			393			211			528			485		
Not analyzed	891			148			259			259			225		
Analyzed Subjects	2394			765			285			793			551		
Human	1431			632			106			462			231		
Animal	17	19	89.5%	13	13	100%	0	0	-	0	0	-	4	6	66.7%
Organization	95	153	62.1%	3	3	100%	0	0	-	82	122	67.2%	10	28	35.7%
Machine	26	50	52%	0	0	-	25	49	51%	1	1	100%	0	0	-
Vehicle	3	12	25%	1	4	25%	0	0	-	0	0	-	2	8	25%
Concrete Inanimate	53	222	23.9%	12	59	20.3%	6	87	6.9%	5	11	45.5%	30	65	46.2%
Nonconcrete Inanimate	89	335	26.6%	13	46	28.3%	11	31	35.5%	37	179	20.7%	28	79	35.4%
Place	43	128	33.6%	0	0	-	0	0	-	0	1	0%	43	127	33.9%
Time	2	7	28.6%	1	1	100%	0	3	-	0	1	0%	1	2	50%
Process	6	37	16.2%	3	7	42.9%	0	9	0%	2	16	12.5%	1	5	20%

Table 8 Subject Theme animacy and sentience in GO

This section will examine the relationship between Subject animacy and the semantic requirements of the process with a particular focus on non-sentient constructions. Table 8 shows the results of all Subjects that are part of either the middle animate or inanimate Subject category and the processes that they were paired with. In total, less than half of all Subjects were analyzed. Excluded from the analysis were relational and existential processes, cleft constructions, semantically empty Subjects, and passives. The Subjects of the remaining 2394 clauses were analyzed following Zaenen et al.'s animacy hierarchy (2004). If the Subject was human, the process was not further considered since humans are animate and sentient and are not restricted in the processes that they can be combined with (see Section 6.3).

The results for each middle animate and inanimate Subject are divided into three numbers. The first number stands for the number of cases in which the Subject type was paired with a sentient verb and thus part of a non-sentient construction. The second number represents the overall occurrence. Note that this does not accurately reflect the frequency of animate and inanimate Subjects in the corpus since relational processes were disregarded entirely. The final number is the relative frequency of non-sentient constructions for that Subject type. To illustrate, a total of 128 Place Subjects were analyzed, of

which 43 were part of a non-sentient construction, which equals a relative frequency of 33.6%.

As was to be expected, the different Subject types are not distributed evenly in terms of overall frequency. Among the inanimate Subjects, concrete Inanimates, nonconcrete Inanimates are the most frequent types, while Organization is the most common middle animate Subject. Time, Vehicle and Animal Subjects are used the least overall. It is important to note that the different Subject types are very unequally distributed among the four registers. Animal Subjects come almost entirely from FICTION, Organization Subjects are primarily from SPEECH, Machine Subjects prevail in INSTR, and all but one case of Place Subjects was found in TOU. For many Subject types, the overall numbers thus do not represent a meaningful global average but rather represent individual registers. A choice of different registers could therefore produce very different results. That being said, Register is a significant predictor of Subject animacy and non-sentient constructions ($\chi^2 = 52.32$, $df = 3$, $p\text{-value} = 2.555e-11$).

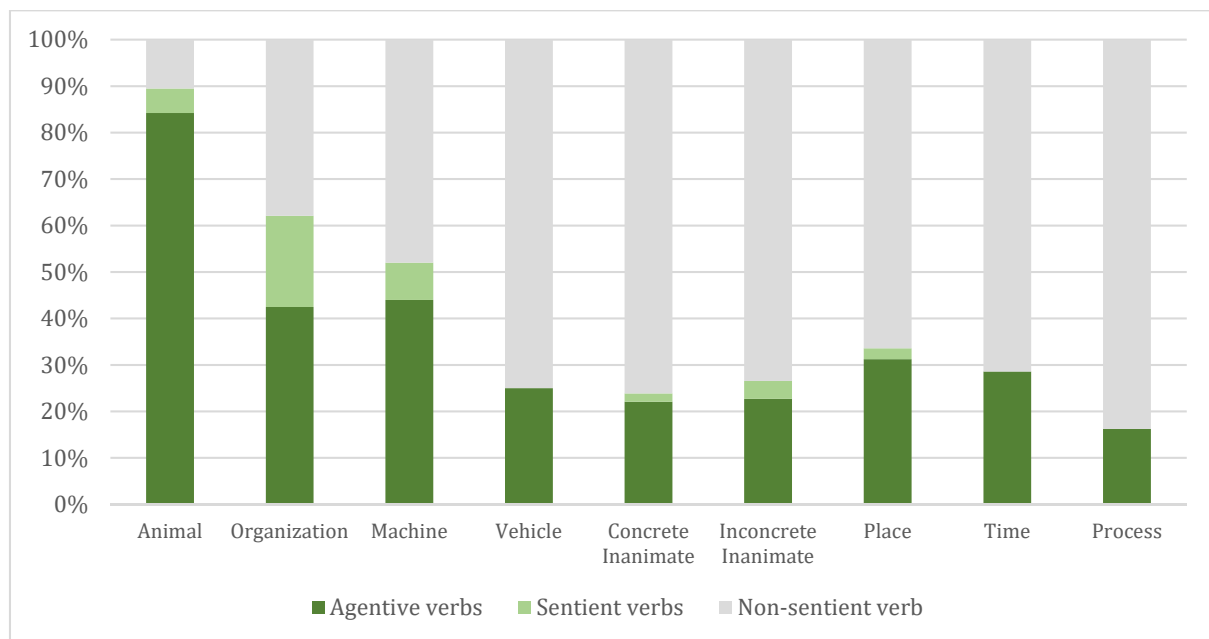


Figure 6 Subject Themes and verb types in GO

Figure 6 illustrates the relative frequencies of different Subjects being paired with sentient and non-sentient verbs. Since relational processes were disregarded, the actual distribution will look very similar, so this breakdown only serves as a means to compare the different Subjects with each other and does not represent an accurate division of sentient and non-sentient processes. In this Figure, verbs were also divided into agentive verbs,

mostly material and verbal processes, and other sentient verbs that do not require agency, mostly mental but also some behavioral processes.

The two types of Subjects that are part of the middle animate category, Animal and Organization, are also the ones that are most likely to be grouped with sentient verbs. Animals in particular are far more likely to carry out processes that require a sentient first participant. Among the (supposedly) inanimate Subjects, Machine Subjects are involved most often in non-sentient constructions with over 50%, which is considerably higher than for the rest of the inanimate Subjects. A combination of a Machine and a verb that requires sentience seems to not be particularly marked in German originals (see example (107), Machine Subject in bold). The number of non-sentient construction that involve Place Subjects is also reasonably high with 33.6% (see example (108); Place Subject in bold). All other inanimate Subjects are distributed similarly in their likelihood of being part of a non-sentient construction at roughly 25%.

(107) *In diesem Fall wird sich **das Nero-Wiederherstellungsprogramm** weigern, ein solches Backup wieder herzustellen!*

*'in this case will [refl-3sg] **the Nero-restoration.programme** refuse a such backup again to.make!'*

[G2E_INSTR_007]

(108) ***Die Bühne** an der Kirchenallee blickt auf eine glanzvolle, über 100-jährige Geschichte zurück.*

*'**the stage** at the Kirchenallee looks at a brilliant, over 100-years.old history back.'*

[G2E_TOU_006]

The distinction between agentive and mental processes shows that, as expected, mental processes are more common and more relevant for Subjects of middle animacy than for Subjects of inanimacy. Organizations in particular were very commonly paired with mental verbs. In example (109), the Subject *mein Haus* (*my house*) referring to a political party belongs to the Subject type Organization (in bold) and is paired with the verb *verfolgen* (*follow*), which in this context means 'watching closely' and this requires a sentient first participant. The use of such constructions is to be expected since Organization Subjects metonymically stand for groups of people that are capable of mental activities.

- (109) **Mein Haus** verfolgt aber aufmerksam die Entwicklung in anderen Ländern [...].
'**my house** follows however attentively the development in other countries [...].'
[G2E_SPEECH_002]

Only looking at non-agentive constructions rather than non-sentient constructions, it is noteworthy that Machine Subjects actually have a higher relative frequency compared to Organization Subjects (44% to 42.5%). Except for Machines, all other inanimate Subjects are very unlikely to be the first participant in a mental event. Vehicle, Time, and Process Subjects are not part of such constructions at all. Surprisingly, nonconcrete Inanimates have a reasonably high likelihood of taking on the role of Senser in mental processes with 3.9%. Most of these nonconcrete Inanimates come from SPEECH and refer to concepts like stability or change that 'need' or 'consider' something (see example (110); nonconcrete Inanimate Subject in bold).

- (110) **Sie** [die neue Weltwirtschaftsordnung] muss die Belange aller - der sich entwickelnden wie der entwickelten Welt - berücksichtigen.
'**it** [the new world economic order] must the needs of.all - the [refl-3sg] developing like the developed world - consider.'
[G2E_SPEECH_003]

7.1.5 Subject identifiability

Subject identifiability is particularly relevant for the German Theme because different topological fields are subject to different rules regarding content and order. One criterion that especially affects the order of the Midfield is the identifiability of the referents of the clause elements. If a referent is identifiable, it is much more likely to be placed at the onset of the Midfield in German, provided it does not already take up the Forefield position. In the case of the first person singular personal pronoun *ich* (*I*), it is almost impossible to place anything before this pronoun in the Midfield. The Theme in German is thus more restricted by information criteria in the Midfield than in the Forefield.

	Overall		FICTION		INSTR		SPEECH		TOU	
Subjects in total	4901		1292		752		1554		1258	
Identifiable	3438	80.2%	1017	90.6%	567	80.8%	1169	84.6%	686	63.3%
Non-Identifiable	850	19.8%	105	9.4%	135	19.2%	213	15.4%	397	36.7%
Other	613		184		53		198		178	
Pre-verbal Subjects	2900		812		388		1031		669	
Identifiable	2109	81.2%	650	92.2%	280	79.5%	800	86.8%	379	61.3%
Non-Identifiable	488	18.8%	55	7.8%	72	20.5%	122	13.2%	239	38.7%
Post-verbal Subjects	2001		494		367		549		592	
Identifiable	1329	78.6%	367	88.0%	287	82.0%	369	80.2%	307	66.0%
Non-Identifiable	362	21.4%	50	12.0%	63	18.0%	91	19.8%	158	34.0%
Subjects in immediate post-verbal position	1649		438		322		431		459	
Identifiable	1195	82.1%	345	90.8%	265	84.9%	331	84.9%	255	68.2%
Non-Identifiable	260	17.9%	35	9.2%	47	15.1%	59	15.1%	119	31.8%

Table 9 Subject Theme identifiability in GO

Table 9 shows the distribution of identifiable and non-identifiable German Subjects in different positions in the clause. Overall, identifiable Subjects are much more common than non-identifiable Subjects, irrespective of position. In the Forefield, 81.2% of Subjects represent identifiable information, which is considerably more than the 43%, reported by Engel (1974). The ratio of identifiable and non-identifiable Subjects varies greatly between the different registers. FICTION has the highest ratio of identifiable referents realized as Subjects with over 90%. This is due to the frequent use of proper nouns and personal pronouns referring to previously introduced characters in the story. With only a 63.3% identifiability rate, Subjects in TOU differed substantially from those in other registers. The majority of non-identifiable Subjects in TOU refer either to some features that a place can offer (see example (111) ; non-identifiable Subject Themes in bold) or to people that will enjoy a place or activity (see example (112)). Register significantly predicts the number of identifiable Subject Themes ($\chi^2 = 37.87$, $df = 3$, $p\text{-value} = 3.012e-08$).

(111) *Zahlreiche Gasthäuser am Wasser und in der Höhe laden zur regionalen Küche und einem Glas heimischen Wein oder Bier ein.*
'**numerous inns at the water and in the elevation** invite to regional cuisine and a glass domestic wine or beer in.'

[G2E_TOU_004]

(112) *Musical-Liebhaber strömen schon seit mehr als 15 Jahren ins Operettenhaus.*
'**musical-lovers** flock already since more than 15 years into the opera house.'

[G2E_TOU_008]

The state of the art claims that the choice of the element in Forefield position is largely uninfluenced by identifiability, while the Midfield follows a stricter order of given elements followed by newsworthy elements (Götze and Hess-Lüttich 2002: 485). It was therefore assumed that Subject Themes in the Midfield are more likely to be identifiable than Subject Themes in the Forefield. However, surprisingly, this is not the case. While the distribution is fairly similar between pre- and post-verbal Subjects, the number of identifiable Subjects actually decreases in the Midfield overall. Again, this is not consistent across all registers since identifiability goes up for post-verbal Subjects in INSTR and TOU. Nevertheless, the assumption that Subject identifiability generally increases in the Midfield could not be confirmed.

This may have to do with Subject position, since non-identifiable Subjects could simply be moved to the back of the Midfield to still adhere to information value order. In example (113), the non-identifiable Subject (in bold) is moved to a position after the circumstance in the Midfield.

- (113) *Nun erscheint in der unteren Bildhälfte **eine Einblendung mit der aktuellen Uhrzeit**.*
*'now appears in the lower image.half **an overlay with the current time.**'*
[G2E_INSTR_003]

For this reason, Subjects that take up the very first position after the finite verb, at the onset of the Midfield, are also considered. In fact, the ratio of identifiable Subjects increases and surpasses that of Forefield Subjects overall. This is, however, not the case across all registers as the two registers that have a lower rate of post-verbal identifiable Subjects in general, FICTION and SPEECH, still have a lower rate even if one only considers the first Midfield position. What is consistent across all registers, however, is the fact that Subjects that open up the Midfield are more likely to be identifiable than Subjects in any position of the Midfield, so the general rule of identifiable referents preceding non-identifiable referents in the Midfield is accurate. And yet, even if the Subject is identifiable, other elements (mostly identifiable Complements and circumstances) can easily be positioned in between the finite verb and the Subject (see example (114); identifiable Subject in bold, identifiable Complement underlined). In summary, the assumption that Subject Theme choice is more restricted by identifiability in the Midfield than in the Forefield is not reflected by the data.

(114) *Gleichermassen scharf sehen uns vermutlich **die drei anderen**.*
*'equally sharp see us presumably **the three others**.'*

[G2E_FICTION_004]

7.2 Discussion

One of the primary goals of this Ph.D. project is to consider a variety of different Theme hypotheses and to compare the differences in Theme that emanate from varied structural realizations. The extent of Theme in any given language is of course not a purely empirical issue. Whether the Subject is an obligatory part of a clause's point of departure is primarily a theoretical question and also depends which aspect of Theme is in focus. Nevertheless, it is worthwhile to examine what compositional differences arise from a theoretical choice.

Naturally, the earlier the borderline between Theme and Rheme is set, the fewer Theme elements overall will be included on average. Thus, it comes as no surprise that Theme numbers increase from the Forefield hypothesis to the first experiential element hypothesis and from the first experiential element hypothesis to the Subject hypothesis for every type of Theme. What is noteworthy is that the absolute numbers do not increase evenly for each of the different kinds of Themes. This is due to positional differences. Circumstances, for instance, are not commonly placed between the Finite and the Subject, which is why most of the circumstance Themes in the Subject hypothesis are also already included in the Forefield and first experiential element hypothesis. Similarly, most textual Themes are already accounted for by the Forefield hypothesis, as most textual Themes have a strong preference for being placed at the very beginning of the clause. Regardless of where a boundary between Theme and Rheme is best placed, this goes to show that different positions within the Theme do not fulfill the same purpose. If every Theme position served as the point of departure in a similar fashion, there would be little reason why textual Themes could not follow other Themes in the Forefield or take up the beginning of the Midfield. What is positioned at the very beginning of the Theme is obviously meaningful. This suggests that Matthiessen's (1992) assumption that the Theme in English is a cline with decreasing levels of thematicity may also apply to German.

The relative frequency of multiple Themes is quite clearly dependent on the chosen Theme hypothesis. Nevertheless, even the Forefield hypothesis yields a surprisingly high number of multiple Themes even though one of the most common claims made about the Forefield in German is that it is only made up of a single element. Whether one believes a distinction between Forefield and left outfield is meaningful or not, the pre-verbal position in German is clearly not one of only a single unit. And even if one disregards all

Themes that include a left outfield, there would still be 2.3% of clauses left which indisputably have more than one constituent in the Forefield. Example (115) for instance includes two circumstance Themes (in bold) in the Forefield, which are clearly independent constituents. As was pointed out by Bassola and Schwinn (2016), studies on topological fields in German have to address the possibility of multiple Forefield positioning.

(115) ***Irgendwann**, | **unbemerkt von uns**, verwandeln diese Alltage sich in gelebte Zeit. 'sometime, | **unnoticed by us**, transform these daily.routines into lived time.'*

[G2E_FICTION_010]

The observation that average Theme numbers increase between the three Theme hypotheses is accurate for all registers. However, it is quite noteworthy how this rate of increase differs between the four registers. INSTR and TOU have by far the lowest numbers of multiple Themes in the Forefield and first experiential element hypothesis yet not only catch up but even surpass the other two registers in the Subject hypothesis. As was pointed out, this has to do with the lower numbers of textual and interpersonal Themes and the higher number of circumstance Themes.

Furthermore, the frequencies of non-experiential Themes and circumstance Themes appear to be connected in German, as there is an inversely proportional effect between the two in that the more non-experiential Themes a register has, the fewer circumstantial Themes can be found. In general, one would expect that a high relative frequency of any Theme also has a negative effect on the frequency of other types of Themes. However, the same inversely proportional relationship does not exist between textual and interpersonal Themes and Complement Themes or Subject Themes, so circumstance Themes seem to be uniquely affected by non-experiential Themes. Also, this effect is very consistent across the four registers, meaning that the register with the highest number of non-experiential Themes, SPEECH, has the lowest number of circumstance Themes, the register with the second highest number of non-experiential Themes, FICTION, has the second lowest number of circumstance Themes, and so on.

This result could be purely coincidental and turn out completely different, had other registers been considered. Nonetheless, there are at least two other possible explanations for this effect: For one, some circumstances fulfill functions very similar to certain textual elements to the point where drawing a clear line between them is relatively arbitrary. This is particularly true for circumstances of Time and textual elements that also express

a chronological order like *dann* (*then*) or *schließlich* (*finally*). A similar case can be made for some circumstances of Reason and textual elements like *deswegen* (*therefore*). However, the register with the highest number of Time Themes, FICTION, is also the register with the highest number of textual Themes. Thus, this cannot be the only reason for the effect.

Another, maybe even more crucial reason for this inversely proportional relationship is that of competition for position. Steiner and Teich (2004) only consider the Forefield the Theme in German, but in this thesis, it was argued that the early Midfield position(s) could be deemed thematic as well. Circumstances and textual and interpersonal elements behave fairly similarly regarding their position in the German clause, as both can occur in the Forefield and neither are likely to take up the onset of the Midfield. Circumstances occur frequently in the middle and the back of the Midfield, more so than textual and interpersonal elements. Nevertheless, if these elements are meant to come early, they take up the same space. The same cannot be said for Subjects and Complements. They can occupy the Forefield, where they would compete with textual and interpersonal Themes as well as circumstance Themes. However, they can also open up the Midfield and allow another element to be thematic in the Forefield. To clarify, consider example (116):

(116)

- a. *Meine Mutter hat mich am Freitag überraschenderweise besucht.*
'my mother has me on Friday surprisingly visited.'
- b. *Mich hat meine Mutter am Freitag überraschenderweise besucht.*
'me has my mother on Friday surprisingly visited.'
- c. *Überraschenderweise hat mich meine Mutter am Freitag besucht.*
'surprisingly has me my mother on Friday visited.'
- d. *Am Freitag hat meine Mutter mich überraschenderweise besucht.*
'on Friday has my mother me surprisingly visited.'
- e. *?Am Freitag hat überraschenderweise mich meine Mutter besucht.*
'on Friday has surprisingly me my mother visited.'
- f. *?Überraschenderweise hat am Freitag meine Mutter mich besucht.*⁷⁴
'surprisingly has on Friday my mother me visited.'

Example (116) includes four elements that can easily be positioned in the Forefield: A Subject, a Complement, a circumstance, and an interpersonal comment Adjunct. So, if only the Forefield was thematic, all four elements would have to compete for this one Theme

⁷⁴ Personally, I would not consider e) or f) ungrammatical, but they do sound marked due to the position of the circumstance and the interpersonal element. This is supported by the fact that circumstances and interpersonal elements are very rarely found in the onset of the Midfield in GO (see below).

position. However, as (116c) and (116d) show, both the Subject and the Complement can occupy the first Midfield position and thus do not have to be in the Forefield to come early in the clause. The circumstance and the interpersonal element, on the other hand, can only be positioned in the Forefield or they have to be moved to the middle or the end of the Midfield. Clauses like (116e) and (116f) are probably not outright ungrammatical but need to be highly motivated, which is why such sequences were not found frequently in the data. The same holds true for textual Themes that are not conjunctions, which are also either in Forefield position or in the back of the Midfield. I believe that the number of non-experiential Themes and circumstance Themes are inversely proportional because they compete for the same early position, while Subjects and Complements do not. In other words, a speaker has to decide whether they want to have a non-experiential element or a circumstance early in the clause, given that the Midfield order does not allow either of them to immediately follow the Finite. Speakers can, however, have both a non-experiential element and a Subject or a Complement early because Subjects and Complements can immediately follow the Finite.

If this interpretation is accurate, it would suggest that the German Theme does cover more ground than just the Forefield. The frequency of Complement Themes is not affected as much by the number of textual and interpersonal Themes even though, in most cases, they cannot share the Forefield. However, Complements can more easily follow the finite verb compared to circumstances. If this onset of the Midfield did not carry any thematic meaning, the relative frequencies for Complement Themes and circumstance Themes in the Forefield hypothesis should be affected equally by the frequencies of non-experiential Themes, but they are not. Additionally, the number of circumstance Themes decreases noticeably between the Forefield hypothesis and the first experiential element hypothesis, while the frequency of Complement Themes remains constant. This also supports the claim that Complements have two viable Theme positions in German (Forefield and early Midfield), while circumstances have only one (Forefield). This is of course not to say that it is impossible to position circumstances at the start of the Midfield. Still, in this data only 6.7% of early circumstances were positioned after the finite verb compared to 33.6% of Subjects and 16.8% of Complements.

Multiple Themes are heavily affected by the Theme hypothesis, as was to be expected. The relative frequencies of marked Themes remain fairly unaffected by differences in Theme hypothesis. Between the Forefield hypothesis and the first experiential element

hypothesis, the relative frequencies of circumstances decrease slightly due to the aforementioned positional restrictions of circumstances in the Midfield. The numbers of both circumstance Themes and Complement Themes seemingly drop further in the Subject hypothesis. However, this is misleading because in the Subject hypothesis, every clause contains an (additional) Subject Theme by definition, which inevitably causes a decrease in the relative frequencies of other experiential Themes.

The Subject is the most common experiential Theme in all Theme hypotheses. In the overall statistics, the frequencies between Subject Themes and circumstance Themes, which is the next most common experiential Theme, diverge clearly. From a purely frequency-based point of view, circumstance Themes can be viewed as marked in German. However, Theme markedness is very dependent on register given that INSTR has an almost even distribution of Subject Themes at 55.6% and circumstance Themes at 40.0% in the Forefield hypothesis. Freiwald (2016: 45) even reports a higher number of circumstance Themes in the register of popular scientific texts, albeit on a limited data set. The frequency of Complement Themes also fluctuates between the register, but their marked status can hardly be denied.

Neumann (2014: 303) also analyzed the register FICTION of the CroCo Corpus regarding the very first element in the sentence. Her results differ noticeably from the results here. However, she also includes interrogative and imperative clauses in her analysis, which explains her high number of Finite and Predicator Themes. Moreover, Theme elements are categorized differently as for example all Adjunct Themes are part of the same Theme category.

Regarding process types, participant roles, and circumstances, the choice in Theme hypotheses also does not play a major role in relative distributions. Naturally, the absolute numbers increase for almost any measure in the Subject hypothesis but their distributions between each other remain relatively constant. The division between the different participant types and their grammatical functions in the clause are nearly identical for all three hypotheses.

In terms of process types generally, the frequency of material and existential processes steadily increases from Finite to Subject hypothesis whereas relational processes steadily decrease. The fact that the Forefield and first experiential element hypothesis do not accurately account for material and existential processes can only mean that other, non-participant Themes are more likely to occur in these process types, namely circumstance

Themes and/or textual and interpersonal Themes. The frequencies of textual and interpersonal Themes in material processes are fairly comparable with other process types at around 7.8%. For existential processes, non-experiential Themes even occur less often than average at 4.9%. This observed difference is caused purely by circumstance Themes, which are most commonly used in existential and material processes.⁷⁵ The high number of material processes with circumstance Themes comes primarily from INSTR. A very common Theme structure in this register includes a circumstance of Condition in Forefield position followed by the Finite followed by an Actor Subject Theme (see example (117); circumstance Theme underlined, Subject in bold).

- (117) *Findet zwei Minuten kein Wiegevorgang statt, schaltet **die Waage** automatisch ab.*
*'takes two minutes no weighing.process place, switches **the scale** automatically off.'*
 [G2E_INSTR_001]

The fact that existential processes feature the most circumstance Themes has to do with the register of tourism leaflets, which contributes the highest number of existential processes to the overall statistics. A common type of clause in TOU is a clause like (118), where first a place is introduced in the form of a circumstance of Place in Forefield position (in bold) followed by a sight that can be found at that place.

- (118) ***In allen Ostseebädern** gibt es bewachte Strandabschnitte mit Strandkörben und den verschiedensten Serviceeinrichtungen.*
*'in all **Baltic.sea.baths** is there surveillanced beach.sections with beach.chairs and the most.diverse service.facilities.'*
 [G2E_TOU_011]

Relational processes feature the second lowest number of textual and interpersonal Themes at 4.9% and the lowest number of circumstance Themes at 17%, which explains why they are overrepresented in the Forefield and first experiential element hypothesis. The low number of textual Themes may be explained by the fact that relational processes describe states and are not as often (chrono)logically ordered as processes that describe series of events like material and behavioral processes. And given that relational processes have their own circumstantial subtype which already expresses relationships of

⁷⁵ 31.7% of existential processes and 26.4% of material processes contain a circumstance Theme, while the average across all process types is 21.7%.

space and time, that may explain why these circumstance Themes are particularly infrequent in this process type.

The effects of the different Theme realizations are marginal in regards to the distribution of circumstance types. There was no notable change between the Forefield hypothesis and the first experiential element hypothesis. In the Subject hypothesis, a number of circumstance Themes are added but these additions made almost no difference to the relative distributions between the different circumstance types. As was pointed out, the only noticeable difference was a drop in relative frequencies of Condition Themes. In fact, the absolute number of Condition Themes does not change at all between the First Experiential Element and the Subject hypothesis. This means that not a single circumstance of Condition is positioned in between another experiential element and the Subject. This again has to do with positioning rules and formal aspects. Circumstances are already marked in the first Midfield position between the finite verb and the Subject. If they are positioned there, the vast majority of these circumstances are short phrases, usually consisting of only one or two words. In fact, of the 92 circumstances that are positioned first in the Midfield⁷⁶, only one was realized as a clause (see example (119); circumstance in bold), which also sounds more like a parenthesis than a regular circumstance.

- (119) *In diesem Sinne wird, **auch wenn uns das heute noch schwer vorstellbar erscheint**, die UNO als der Ort der Entscheidungen an Bedeutung zunehmen.*
*'in this sense will, **even if to.us that today still hard to.imagine seems**, the UNO as the place of.the decisions in significance gain.'*
- [G2E_SPEECH_010]

The majority of Conditions are clauses introduced by *wenn* or *falls (if)*, which explains why they do not often assume a post-verbal Theme position. It is difficult to say whether this has to do with the formal difference between phrases and clauses or just with length in general, given that clauses are typically longer than phrases. Upon further review, it turns out that circumstance Themes in the Forefield position are on average almost twice as long (average length of 31.6 characters) as circumstance Themes in the Midfield (av-

⁷⁶ In Section 7.1.3, I reported only 28 cases of circumstances that immediately follow the finite verb in the first experiential element hypothesis. However, the Theme analysis in this hypothesis ends as soon as the first experiential element occurs. Therefore, cases where a clause is opened up by a circumstance or Complement, followed by the Finite, followed by a(nother) circumstance are not accounted for. This pattern makes up these additional 64 circumstances in the onset of the Midfield.

erage length of 16.5 characters). This difference in length hints towards processing effects so that the speaker tries to arrive at the Subject earlier if it is in a post-verbal position, even if its modal responsibility is limited. This result on circumstance length is the opposite to those in English, where circumstances are more likely to be fronted if they are shorter (Wiechmann and Kerz 2013: 20).

All in all, it was surprising how little difference the Theme hypotheses made for the statistics. A difference in Theme numbers was expected and observable in the data. As reported, there are also a few other differences between the three hypotheses but given the large number of Theme aspects analyzed here, some variation is inevitable. I therefore come to the conclusion that apart from multiple Themes and average Theme number, the choice of Theme hypothesis in German is largely inconsequential.

What is not inconsequential at all regarding Theme structure is the choice of register. The data shows remarkable differences between the four registers in almost all Theme-related aspects and Register as a category significantly predicted all of the analyzed Theme aspects except for Theme number in the Forefield hypothesis. This clearly shows that the choice of register has an enormous influence on the organization and content of Theme in German. As discussed in 5.2.2, Fries (1995a) formulates four hypotheses regarding Theme, genre, and thematic potential. His third hypothesis states that experiential content and genre correlate, which has only partially been confirmed by the state of the art. However, on the basis of the registers here, this relationship between experiential Themes and genre/register is indisputable at least in German.

Following SFL terminology, the thematic differences between the registers can be largely attributed to differences in field, tenor, and mode of discourse (Halliday and Hasan 1985) between the four registers. In the following, each register will be considered separately and their most striking characteristics and their effects on Theme will be pointed out. But first, some general register effects that were not tied to one or more particular registers will be addressed.

One of the most interesting effects of register variation shows in circumstance thematic potential. It was assumed that the overall frequency of different types of circumstance Themes would be highly register-dependent, which is accurate, but that the general thematic potential of a circumstance type is independent of register. This assumption turned out to be true for some circumstances, while other circumstance TPots showed great variation between different registers. The TPot of some circumstances, like Place

and Time, is completely unaffected by register differences, while the TPots of other circumstances like Purpose or Guise vary greatly between the different registers. For some circumstance types, a high general frequency also contributes to a high TPot as is the case for Conditions and Purpose Themes in INSTR or Quality Themes in FICTION. For other circumstances, the exact opposite effects can be observed, for example for circumstances of Means, which are by far most common in INSTR but also have the lowest TPot there. The same holds true for Comitative Themes, which are most often found in TOU, where they also have the second lowest TPot.

Just looking at descriptive statistics makes it difficult to assess the reasons for these variations in circumstance TPot. Most likely, these differences are partially caused by formal or functional differences between circumstances of the same type. Even though two circumstances can be sorted into the same general circumstance category, the meanings that they express and the form that they have can still be very different and as a result can be more or less suitable as a point of departure. For instance, circumstances of Means in INSTR are often realized as prepositional phrases introduced by *mit* (*with*), referring to an instrument or device that is used to achieve the desired outcome (see example (120)). In SPEECH, circumstances of Means are often introduced by *durch* (*by*), referring to steps that have been or need to be taken in order to achieve the desired outcome (see example (121); circumstances of Means in bold). While both can generally be asked for using *How?*, a distinction could also be made between asking *Using what?* and *Doing what?*. These subtle differences in meaning may be the reasons for the differences in thematic potentials.

(120) *In diesem Menü können Sie **mit dem Einstellrad** eine voreingestellte Farbtemperatur [...] auswählen*
*'in this menu can you **with the adjustment.wheel** a pre-set colour.temperature [...] choose.'*

[G2E_INSTR_002.]

(121) ***Durch die Zusammenlegung von Arbeitslosenhilfe und Sozialhilfe** gelingt es uns zunächst, deutlich zu machen, dass wir sie als Menschen begreifen, die nicht vergessen sind.*
*'**through the merging of unemployment.benefit and social.welfare** succeeds it us initially, explicit to make, that we them as humans see, who not forgotten are.'*

[G2E_SPEECH_013]

In Section 5.9, five different positions on Theme markedness in German were discussed. As was also discussed at length in 5.6, the difficulty of assessing markedness is that it can be based on different criteria. Halliday (1967b: 219) suggests that Theme markedness can be based on contextual motivation, with unmarked Themes requiring the least and marked Themes requiring the most amount of contextual motivation. However, motivation is difficult to measure, which is why frequency is usually used instead. Freiwald (2016) suggests an alternative method of assessing markedness based on the thematic potential, which has also been argued for in this present thesis.

In terms of raw frequencies, the suggestion that Complements are unmarked Theme units (Kirkwood 1970; Zifonun, Hoffmann, and Strecker 1997) can be rejected. At 8.1%, they are not textual abnormalities but are considerably less common than other experiential Themes. Circumstances of Place and Time are the most common types of circumstance Themes, but together only make up 12.5% of all clauses, which is why I would not consider them unmarked on their own. At around 28.0%, the frequency of circumstance Themes is right in between Subject and Complement Themes. The binary distinction between marked and unmarked is not ideal given that markedness is clearly a scale. But since these categories exist and are commonly referred to, I would side with Erdmann (1990b) and consider circumstance Themes rather part of the marked than the unmarked Theme category.

This evaluation changes completely when thematic markedness is not based on plain frequencies but on thematic potential. In this case, two circumstance types are clearly unmarked given that they are more likely to function as the Theme of their clauses, namely circumstances of Condition and Concession. Additionally, there is a number of circumstance types that require only little contextual motivation, amongst others Place and Time but also Matter, Behalf and Additive. Quality and Guise Themes, on the other hand, are clearly marked in German since their thematic potential almost matches that of Complement Themes. Freiwald's (2016) analysis of thematic potential in popular scientific texts also showed that Matter and Condition had high thematic potentials alongside other circumstances like Time and Purpose. Moreover, Quality Themes were also among the circumstances with the lowest TPot in popular science texts together with Additives and Comparison. This shows that there is a certain consistency of TPot across different registers.

One other general register-related effect that is not limited to one register in particular regards the frequency and type of non-sentient constructions. Non-sentient constructions are generally difficult to compare between the different registers because the types of Subjects used are highly dependent on the kinds of experiences that are discussed in each register, the fields of discourse. As a consequence, most non-Human Subjects are very common in some registers and virtually or actually non-existent in other registers. The only two kinds of inanimate Subjects that are comparably common in all registers are concrete and nonconcrete Inanimates. The likelihood of these inanimate Subjects to be part of a non-sentient construction varies between the different registers. This variation is, however, not always consistent within one register. For example, INSTR has the highest relative frequency of non-sentient constructions for nonconcrete Inanimates, yet the lowest frequency of non-sentient constructions for concrete Inanimates. Similarly, SPEECH has the second highest number of concrete Inanimates in non-sentient construction but also the lowest for nonconcrete Inanimates.

These inconsistent distributions of non-sentient constructions are most likely due to differences in the lexical meanings of these Subjects. Concrete and nonconcrete Inanimates are two rather general descriptions of a variety of different meanings that a Subject can express. For instance, in INSTR, many of the concrete Inanimate Subjects refer to displays and pop-up windows (see example (122); concrete Inanimates in bold), whereas concrete Inanimates in TOU mostly refer to (hiking) trails, parts of buildings, and types of food (see example (123)).

(122) *Sobald Nero den Schreibvorgang beendet hat, erscheint **ein Fenster mit der Meldung 'Brennvorgang erfolgreich'**.*

*'as.soon.as Nero the writing.process finished has, appears **a window with the message 'burning.process successful'**.'*

[G2E_INSTR_007]

(123) ***Vielfältige Wege und Pfade** führen die Menschen bergauf in dichte Mischwälder mit einer ganz eigenen Flora und Fauna.*

*'**diverse roads and paths** lead the humans uphill in dense mixed forests with a wholly unique flora and fauna.'*

[G2E_TOU_004]

While all of these meanings fall under the category of concrete Inanimates, it is hardly surprising that there would be a difference between them regarding their potential of

appearing in a non-sentient construction. A potential solution to this problem would be to sub-categorize concrete and nonconcrete Inanimates but, since these groups are so varied, it is improbable to end up with categories that are used consistently in all of the registers.

Nevertheless, in general, the results on Subject Theme sentience have confirmed what has been originally assumed: The perceived level of animacy and agency of Subjects represents a cline rather than distinct categories and the use of sentient and non-sentient constructions with different Subject types is a question of probability. The results also show that Zaenen et al.'s (2004) and Garretson's (2004) animacy categories for English can be applied to German as well since Subjects that are placed higher in the animacy hierarchy are also more likely to be paired with a sentient verb.

The results on Subject positioning and identifiability were likely the most astounding of all. There are some register differences, which will be commented on below, the overall picture looks quite similar across all registers, which is that there is no notable difference in terms of Subject identifiability between Forefield and Midfield. This finding goes against all topological descriptions of German that were referenced in this thesis and also goes against my personal native speaker intuition which deems some non-identifiable Subjects more marked in the Midfield than the Forefield position.

The overall lower number of identifiable Subjects in the Midfield compared to the Forefield was already attributed to the fact that there may be elements in between the Subject and the Finite which are identifiable and thus take precedence over the Subject in the Midfield sequence. In fact, when only considering Subjects immediately following the finite verb, the number of identifiable Subjects goes up in all registers, which suggests that many of the non-identifiable Subjects are positioned late in the Midfield preceded by either a circumstance or a Complement. This generally confirms Götze and Hess-Lüttich's (2002: 485) assumption of the topological verbal distance, which locates elements which are more newsworthy closer to the lexical verb. And yet, while this is more common for non-identifiable Subjects, identifiable Subjects could also easily be headed by another element in the Midfield (see example (124); identifiable Subject in bold, Complement underlined).

- (124) *Nur zehn Wochen später, am 11. Juni, folgte ihm **Sicardsburg**, wohl wegen gebrochenen Herzens, ins Grab.*
'only ten weeks later, on the 11th of June, followed him **Sicardsburg**, probably because of a broken heart, to the grave.'

[G2E_TOU_021]

Looking at some of the cases with late identifiable Subjects, circumstances that preceded the Subject in the Midfield are not noticeably different from pre-verbal circumstance Themes other than with respect to their shorter length, which was commented on earlier. What is noteworthy, however, is that every Complement Theme positioned between the Finite and an identifiable Subject is without exception an identifiable personal pronoun. The identifiable Subjects which follow these identifiable Complement pronouns are never personal pronouns but always noun phrases with a definite article or demonstrative determiner or proper nouns (see example (125); Subject in bold, Complement underlined). This shows that the positional rules outlined by the state of the art are generally at play. If a clause element is realized as a personal pronoun it has a much higher probability to open up the Midfield than other identifiable referents irrespective of its grammatical function (König and Gast 2009: 167-168).

- (125) *In gleichmässig knappen Intervallen kippt ihm **das schlanke Fläschchen** vor dem Mund in die Höhe.*
'in evenly short intervals tilts him **the slim bottle** before the mouth into the height.'

[G2E_FICTION_004]

And yet, the question still remains as to why the differences in identifiability are almost negligible between Subject Themes in Forefield and first Midfield position. I do believe that the Midfield sequencing patterns outlined by the state of the art are generally correct. Factors like identifiability, length, agency, grammatical function, form, etc. can in theory compete with each other and be the cause of particular Midfield patterns. Such cases were observed in the data and the Midfield rules such as pronoun-first were followed. However, in practice these considerations are often not relevant. The majority of German clauses only have one or two meaningful Theme candidates, of which one occupies the Forefield and the other the early Midfield without any real competition for Midfield positioning. Cases where a number of clause elements are eligible for Forefield and Midfield thematization are just very rare, at least in the four registers analyzed here, which is why

they do not have a stronger effect on the statistics. And even if competition for position exist, speakers still have a choice between different sequences in most cases and can deviate from the unmarked order if the context motivates it.

Turning to the individual registers, the one register which is the most unique in almost every Theme related measure is INSTR. It has the lowest average in multiple Themes and Theme number in the Forefield hypothesis and yet the highest average in both categories in the Subject hypothesis for reasons explained above. It features the highest number of circumstance Themes and also the lowest number of Complement Themes. With a total of 61.2%, INSTR includes the by far highest frequency of material processes across all registers, almost twice as high as FICTION and SPEECH. At the same time, it also has the lowest frequency of all other process type. In terms of circumstance Theme types, INSTR differs most from all of the other registers. 42.7% of all circumstance Themes in instruction manuals are circumstances of Condition, which is more than ten times as many compared to TOU. Given that there are 22 circumstance types in total, it is quite remarkable how one single type can make up almost half of the circumstances in the entire register. Means and Purpose Themes are relatively common in INSTR as well, while generally common circumstances like Place and Time are surprisingly rare compared to the other registers. Regarding circumstance TPot, INSTR again stands out since the register scores either highest or lowest for almost every circumstance type. INSTR is essentially the only register that includes Machine Subjects, which have an almost even spread of verbs that require a sentient first participant and verbs that do not. Based on these results, Machine is arguably not part of the marked, non-sentient group in German and should receive the status of middle animacy.

Most of these thematic characteristics can be explained by the unique field and tenor of discourse in instruction manuals. Instruction manuals predominantly describe different applications of the product that was bought and explain cause and effect relationships. However, thematically, neither the product nor the user is consistently in focus of the clauses in INSTR. Oftentimes, the point of departure is not the machine or what it does but rather the conditions, requirements, and purposes for its use. In other words, the focal point of many clauses in this register is less on providing information and more on the conditions under which the information is relevant to the reader. Grammatically, this translates to the high frequency of marked circumstance Themes and in particular the incredibly high number of Condition Themes. Many clauses in INSTR are opened up by a

conditional statement followed by a description of how the machine reacts if the condition is met. Other frequent thematic structures involve circumstances of Means and Purpose to highlight either how or for what purpose the product can be used.

In this sense, the register is a lot more formulaic than the other three registers. Not only are particular lexico-grammatical patterns used more consistently, but oftentimes entire sentences are also repeated almost word-for-word (see example (126)). These repetitions can be attributed to the need in instruction manuals to be particularly specific and not leave out any possible application or circumstance surrounding the use of the product.

(126) *Bei spürbarem und hörbarem Ratschen springt Zeiger B um jeweils 90°. Bei spürbarem und hörbarem Ratschen springt Zeiger E um jeweils 90°. [...] Bei spürbarem und hörbarem Ratschen springt Zeiger H um jeweils 90°. Bei spürbarem und hörbarem Ratschen springt Zeiger L um jeweils 90°.*
'at palpable and audible rattling jumps indicator b about respectively 90°. at palpable and audible rattling jumps indicator e about respectively 90°. [...] at palpable and audible rattling jumps indicator h about respectively 90°. at palpable and audible rattling jumps indicator l about respectively 90°.'

[G2E_INSTR_014]

Since instruction manuals revolve around a particular product, the numbers of Machine Subjects and concrete Inanimate Subjects are higher than in other registers. Non-sentient constructions are common in this register as German speakers apparently assign machines a certain level of agency and sentience. This is particularly true for intelligent machines like computers, computer programs, and robots, which often do act autonomously with little to no input from an animate agent. This is partly also the reason for the high frequency of material processes, which are primarily about doings and happenings. For the same reason, mental processes are comparatively infrequent. While there are cases of Machine Subjects functioning as the Sensor of a mental event, such constructions are the exception. The low number of relational processes is surprising as it is the lowest out of all four registers. It was expected that relational processes would be commonly used in INSTR to describe an attribute or feature of the product. But as it turns out, instruction manuals are less about what the product is and more about what it does or can do.

The tenor of discourse is also unique in instruction manuals because the reader is often addressed directly, which suggests a smaller distance between speaker and hearer than, for example, in FICTION. Also, the communicative role of the speaker is most diverse in

this register because it often shifts between information provider and demander of goods and services, as many clauses are about what users should and must do or not do. As a consequence, the number of imperative clauses is exceptionally high in INSTR. The form of address is quite direct in INSTR. Most demands for goods and services are realized as imperatives and those that are phrased as declarative often include a modal verb of obligation without much hedging (see example (127)). These demands, which are almost exclusively events of doing, are another reason for the high number of material processes.

(127) *Um den Bedienkomfort zu erhöhen, müssen Sie den PIN-Code in einer Betriebsphase jedoch nur einmal eingeben.*
'in.order the serving.comfort to raise, must you the PIN-code in an operating.phase however just once enter.'

[G2E_INSTR_003]

The second register that deviates noticeably from the overall average for most Theme measures is tourism leaflets. In many respects, TOU turned out to be similar to INSTR, given its high number of marked circumstance Themes and low number of textual and interpersonal Themes. However, unlike INSTR, TOU also has the highest number of marked Complement Themes, making it the most thematically marked register of all. TOU has the most even distribution of material and relational processes, with material processes being only slightly more common. The number of relational processes is the highest across all registers. On top of that, existential processes occur comparatively frequently in this register despite its status as a peripheral process type. All other process types have a below average rate of occurrence in the register. In terms of circumstance Themes, TOU stands out due to the particularly high number of circumstances of Place, making up over 42.2% of all circumstance Themes. Yet, unlike INSTR, there is no other circumstance Theme that is particularly characteristic of the register. Surprisingly, Time Themes are also less common than in other registers, even though many entries in TOU deal with historical facts. Circumstances of Guise and Reason have a TPot that is substantially higher than average while neither circumstance type is exceptionally common in the register.

TOU is the only register with inanimate Place Subjects, which occur in non-sentient constructions surprisingly frequently; they are in fact the most common inanimate Subject type that is part of such constructions if Machines are considered part of middle animate Subjects. Organization Subjects also occur relatively frequently in the register and

they mostly refer to tourism agencies. Across registers, the number of identifiable and non-identifiable Subjects does not vary noticeably, neither in pre- nor post-verbal position. The only exception to this rule is TOU, which has a lower number of identifiable Subjects in any position of the clause.

The field of discourse in tourism leaflets mostly covers different touristic places and the attractions they have to offer. Thematically, this is largely accomplished by two lexico-grammatical patterns: One involves a description of what people in general, and the reader in particular, can and should do at that place. Such clauses usually involve material processes, which are often introduced by a circumstance of Place. In succession, these circumstances of Place as points of departure create a chronological and geographical order of events, almost like a trip itself (see example (128); Place Themes in bold), which explains the high frequency of this circumstance Theme. It also explains why textual Themes are less relevant in the register. Sometimes, these clauses are also realized as imperatives, similar to how a tour guide would address their travel groups.

(128) ***Außerhalb der Altstadt*** imponieren die Gründerzeitvillen in der Weststadt. ***Auf der anderen Neckarseite*** finden Geschichtsinteressierte Spuren des Mittelalters.
'***outside of the old town*** impress the Gründerzeit.mansions in the western.city. ***on the other neckar.side*** find history.interested traces of.the middle.ages.'

[G2E_TOU_005]

The other common thematic structure in tourism leaflets involves the establishment of a feature in space. There are two lexico-grammatical realizations that accomplish that. One consists of a relational process, oftentimes of the circumstantial subtype. The circumstantial Attribute is also often fronted, drawing more thematic attention to the place than to the feature itself (see example (129); Complement Theme in bold). This is one of the reasons for the high number of Complement Themes.

(129) ***Neben dem Kunstverein*** liegt das beliebte Szene-Restaurant "Jena Paradies" [...].
'***next to the art.club*** lies the popular scene-restaurant "Jena Paradise" [...]'

[G2E_TOU_007]

The other lexico-grammatical pattern comprises an existential process, which is frequently paired with a circumstance of Place as its initial point of departure (see example (130); Place Theme in bold). This result corresponds to Neumann (2003), who also found

a high number of spatial circumstance Themes in German tourist guides. The Existent in existential processes can also be fronted easily, which is a second reason for the frequent use of Complement Themes. In a sense, both of these thematic patterns accomplish the same goal. The difference is that existential processes with Place Themes draw more attention to the place itself rather than the feature, while relational processes more often use the feature or the sight as their point of departure. Fronted circumstantial Complements do invert this focus again but are generally the more marked choice. This frequent switch between moving to a new location and then describing distinctive features of that location results in the even distribution of material and relational processes.

- (130) *In der gesamten Lahn-Region gibt es mehr als 240 Kilometer Wasserwanderwege, [...].*
'in the entire Lahn-region are there more than 240 kilometers of.water.hiking.trails [...].'
[G2E_TOU_004]

In terms of tenor of discourse, the social distance between speaker and hearer is low in TOU. The reader is often addressed directly and is frequently grouped together with the speaker using the first personal plural pronoun *wir* (*we*) (see example (131); Subject in bold). This contributes to the feeling that the speaker is actually present during the reader's trip to act as a tour guide, reinforced by the use of informal imperative clauses.

- (131) *Durch die Buchkremerstraße gelangen **wir** in das alte Kurggebiet der Stadt und sehen rechts ein Hotel, das den alten Namen der Stadt trägt: Aquis Grana.*
*'through the Buchkremerstrasse get **we** into the old spa.area of.the city and see to.the.right a hotel, that the old name of.the city carries: Aquis Grana.'*
[G2E_TOU_001]

Since the tourism leaflets register is mostly about places, it is hardly surprising that many of the Subjects are Places. What is surprising is the high likelihood of these Place Subjects to be used in non-sentient constructions. In many cases, the places are presented as animate participants, inviting the reader to visit them (see example (132); Place Subject in bold). I believe that this level of ascribed sentience to places is rather unique to TOU and is not representative of Place Subjects in German as a whole.

- (132) **Hübsche Hofläden, Brennereien und Gaststätten mit regionaler Speisekarte laden zu einer Entdeckungsreise für Zunge und Gaumen.**
'**pretty farm.shops, distilleries, and restaurants with regional menu** invite to an expedition for tongue and palate.'

[G2E_TOU_012]

The low number of identifiable Subjects in TOU is striking, especially in the light of the high number of second person singular and first-person plural pronoun Subjects found in the register. Nonetheless, in many cases, the reader is not directly addressed in the clause, and instead a group of people, to which the reader may or may not belong, is referenced in the Subject. These groups of people usually share a hobby like hiking or fishing and the tourism leaflet comments on what the venue can offer them (see example (133); non-identifiable Subjects in bold). This supports the universal appeal of tourism leaflets as they can seemingly offer something to any kind of reader. Besides groups of people, many of the sights used as Subjects are also not uniquely identifiable since the reader is presumably not familiar with them yet (see example (134)).

- (133) **Selbst Wildwassersportler** werden im Gelbachtal, einem Seitenzufluss der Lahn, fündig.
'**even white.water.athletes** will in.the Gelbachtal, a side.river of.the Lahn, discover.'

[G2E_TOU_004]

- (134) Auf der gegenüberliegenden Seite der Ursulinerstraße steht **eine moderne Bronzeplastik** [...].
'on the opposite site of.the Ursulinerstrasse is **a modern bronze.sculpture** [...].'

[G2E_TOU_001]

The Theme results in FICTION are surprising in that they are not very surprising. Even though fiction represents creative writing, most thematic aspects in FICTION turn out to be remarkably similar to the overall distribution. FICTION is closest to the overall average in terms of multiple Themes, average Theme number, frequency of Subject Themes, frequency of circumstance Themes, frequency of textual and interpersonal Themes, and second closest in terms of Complement Themes. Judging from the four registers analyzed in this study, FICTION is as average as it gets thematically. This impression holds also for process type and participant Theme distribution, where the register either scores closest

or second closest to the average for five of the six process types. Only behavioral processes are more commonly found in FICTION. Out of all the registers, FICTION has the most even spread of the different process types, with material and relational processes being the most common and all other processes being distributed fairly evenly. It is one of two registers that uses mental processes regularly.

In terms of circumstance Themes, FICTION is again close to the average for many of the circumstance Theme types, but occasionally it also scores highest or lowest for others. For example, FICTION has the highest number of temporal circumstance Themes, namely Time, Duration, and Frequency Themes. In addition, Quality Themes are a lot more common in this register compared to the rest of the corpus. On the flipside, Means, Purpose, and Behalf Themes have the lowest frequency here. FICTION is also the register that has the highest number of *Other* circumstance Themes, which underpins the impression of a relatively even distribution of Themes in the register. Regarding circumstance TPot, the register is again remarkably average with only Quality having noticeably high and Purpose noticeably low thematic potentials. High thematic potentials correspond to high occurrence rates in FICTION.

FICTION has the overall lowest relative frequency of inanimate Subjects. It is the only register that includes Animal Subjects, which are paired with sentient verbs in 100% of cases. With 90.6% overall, Subjects have the highest probability of being identifiable in the fictional register.

It is difficult to speak of a homogeneous field of discourse in FICTION since the experiences in the story world can differ significantly between two texts. In this respect, FICTION is probably the most diverse of the four registers. However, generally speaking, the fictional texts in the CroCo Corpus deal with characters that do something actively in the fictional world. At the same time, a detailed description of the characters and the surroundings is essential to fictional storytelling to inspire the reader's imagination. Lexicogrammatically, this is accomplished through the use of material processes and relational processes respectively, which explains the relatively even distribution of these two process types. However, FICTION often illustrates a range of different events, from the inner workings of the characters to outward representations of these mental processes, to direct and indirect communication. The field of discourse is quite diverse, and this diversity also corresponds to a varied range of process types. The fact that mental and behavioral processes occur relatively frequently are testament to this diversity.

This diversity of experiences can also be observed in the use of marked circumstance Themes. Unlike INSTR and TOU, FICTION is not dominated by one particular circumstance Theme but has a comparably even distribution between many different types. Place and Time Themes are by far the most frequent but apart from these, most other circumstances have an occurrence rate that is close to the overall average. Time-related circumstance Themes are more common in FICTION than in other registers. This has to do with the fact that fictional texts are typically events that unfold in time where a description of temporal relationships is paramount to understand the plot. These shifts in time or duration are often made the point of departure to guide the reader through the unfolding of the story. Similarly, orientation and re-orientation in space are essential for the illustration of a plot, which explains why circumstances of Place are a common thematic choice. FICTION deviates most notably from the other registers in its frequent use of Quality Themes (see example (135); Quality Theme in bold). Circumstances of Quality are generally common in all registers, but they make up a substantially larger percentage of the thematic space in FICTION and also have a significantly higher TPot, at least in comparison to INSTR and SPEECH. Since fictional stories are oftentimes concerned with human behavior, writers use Quality Themes to draw attention to the way in which such doings are carried out, which is less common in the other registers.

- (135) *Andächtig* hörten wir ihm zu [...].
'**devoutly** listened we him to [...].'
[G2E_FICTION_004]

The communicative role of the speaker as a provider of information is relatively constant in the register. Nevertheless, since such stories often illustrate the interactions between characters, the characters can become the speaker of single clauses, taking on a variety of different communicative roles. For this reason, FICTION has the highest number of interrogatives among the four registers. The register offers very little direct interaction with the reader.

Animal Subjects occur only in FICTION. However, it must be noted that 12 of the 13 Animal Subjects all come from the same text, which is about a boy who owns a dog. Animal Subjects are thus not generally frequent in FICTION but only occur in selected texts. This is different from other registers, for example TOU, where Place Subjects are used in most texts. The fact that the Animal Subject in the one text is a dog likely also plays a role

in the distribution of sentient and non-sentient processes. Speakers probably ascribe a higher degree of agency to mammals, especially domesticated animals, than to insects, for example. The high number of identifiable Subjects can be attributed to the frequent use of proper nouns and pronouns referring to characters in the story.

SPEECH is very much the opposite of TOU and INSTR in terms of basic Theme distribution. It has the lowest number of circumstance Themes and marked Themes overall and at the same time the highest number of non-experiential Themes, which, as pointed out earlier, is an interconnected issue. As a result, SPEECH has the highest number of multiple Themes and Theme elements in the Forefield and first experiential element hypothesis. While SPEECH leads the statistics in both types of non-experiential Themes, interpersonal Themes are particularly frequent, especially in view of their generally low rate of occurrence in other registers.

Participants of relational processes are the most common participant Themes in SPEECH. Material processes are the second most frequent process type in the register but have the overall lowest relative frequency compared to the other three registers. On the other hand, mental and verbal processes are surprisingly frequent in political speeches, scoring highest across all four registers. Overall, the process type distribution is comparable to that of FICTION apart from behavioral processes, which are basically non-existent in political speeches.

In terms of circumstance Theme types, SPEECH is also quite similar to FICTION in that they are relatively balanced across many different types and do not predominantly use one specific circumstance Theme. As is the case for all registers apart from INSTR, circumstances of Place and Time are the most frequent circumstance Themes. A variety of different circumstance types have a fairly similar rate of occurrence with no type standing out in particular. SPEECH has the highest percentage of Comitative, Reason, and Comparative Themes but only by a small margin, making none of these circumstances representative of the register. Interestingly enough, the circumstance Themes that are particularly infrequent in SPEECH are the same types of circumstances that are also infrequent in INSTR, namely Quality, Duration, and Frequency. Regarding circumstance TPot, the numbers in SPEECH largely correspond to the average; however, while generally close, all circumstance TPots but two are below the global average. Hence, circumstances in SPEECH generally have the lowest likelihood of being made the Theme of their clauses. The register has the highest number of middle animate and inanimate Subjects, including

a lot of Organization and nonconcrete Inanimate Subjects. Even though the number of nonconcrete Inanimates is highest in SPEECH, the probability of them being part of a non-sentient construction is lowest.

Up to this point, mode of discourse has not been discussed, given that it is generally very similar in all of the previously discussed registers. However, SPEECH has a unique mode of discourse compared to the other three registers as the texts in SPEECH are written but meant to be spoken. They are not transcripts of oral speeches but pre-planned written versions of these speeches. Thus, they lack many of the linguistic features common in oral communication, like false starts. At the same time, some textual and interpersonal Themes are uniquely used in political speeches, like continuatives (see example (136); continuative in bold) and Vocatives, which are characteristic of the oral, interactive mode of discourse.

(136) ***Nein**, wer nicht mit Scheuklappen durch die Gegend läuft, der kann sehr wohl mitbekommen, wie zum Beispiel im Osten unseres Landes durch betriebliche Bündnisse Ansiedlungserfolge erreicht worden sind.*

'no, who not with blinders through the area runs, he can very well notice, how for example in.the east of.our country through company alliances settlement.successes reached been have.'

[G2E_SPEECH_013]

This is partly the reason for the increase in non-experiential Themes: the register allows the use of Themes that would be uncharacteristic or meaningless in a purely written register. Vocatives and continuatives also have the effect of increasing the number of multiple Themes in the Forefield hypothesis because, like coordinating conjunctions, they do not exhaust the Forefield position, allowing for an additional Theme element.

The tenor of discourse is also distinct in SPEECH compared to the other three registers. Political speeches generally have two groups of audiences: the participants of the congress or parliament assembly, whom the speaker can interact with to some degree, and the general public, which is not present and is exposed to the speech through the media. The first group is frequently addressed during the speeches, often by name, which is represented by the high number of Vocatives. The second group is also addressed directly by the speaker, using mostly first-person plural pronouns.

There are two types of experiences that are primarily shared in SPEECH: the state of either a country, the economy or government programs, and the actions of politicians and

governments. These experiences map nicely onto relational processes and material processes. At the same time, the numbers of mental and verbal processes are surprisingly high. In most of the mental clauses, the speakers comment on the needs or desires of either themselves or the entire country with the help of first person singular and plural pronouns (see example (137)). Participant Themes from verbal processes typically represent the speakers themselves having conversations with other governments but also events that have taught them something, which is a more metaphorical kind of verbal process (see example (138); Subjects in bold).

(137) *Wir brauchen einen nationalen Aufbruch für Bildung, Forschung und Familie [...].*
'we need a national departure for education, research and family [...]'

[G2E_SPEECH_007]

(138) ***Das verheerende Seebeben vom 26. Dezember** hat uns furchtbar deutlich gemacht, dass eine Naturkatastrophe globale Auswirkungen haben kann.*
'the devastating sea.quake from.the 26th of.December has to.us terribly clear made, that a natural.disaster global effects have can.'

[G2E_SPEECH_006]

The distribution of circumstantial Themes is arguably least characteristic in SPEECH. While FICTION also has a relatively even spread of circumstance Themes, at least temporal circumstances are representative of the register to some degree. In SPEECH, there is no single circumstance that is representative of the register, however. Its characteristic is the even distribution between many circumstance types, which speaks to the variety of experiences that are shared in SPEECH. Place Themes are used to shift the focus from one country or governmental space to the next, while circumstances of Time are typically used to refer to the historical past or the future. In clauses like (139), Means Themes (in bold) act as a focal point on the previous accomplishments of the politician or government through which a goal has been achieved. The fact that SPEECH corresponds to FICTION in terms of process types and circumstantial Theme distribution suggests that speeches also represent a sort of storytelling; albeit one of very different experiences.

(139) ***Durch den Reformprozess** haben wir dafür gesorgt, dass Menschen, die in der Vergangenheit im System staatlicher Leistungen versteckt wurden, aus der Anonymität geholt worden sind.*

'through the reform.process have we for.that cared, that humans, who in the past in.the system national achievements' hidden were, out.of the anonymity taken been have.'

[G2E_SPEECH_012]

Organization Subjects used in SPEECH are mostly governments or countries, metonymically referring to the people or the political leaders of these countries. The expression of desires is a common event addressed in political speeches, which is the reason for the high number of non-sentient (as opposed to non-agentive) constructions for Organization Subjects. Nonconcrete Inanimates in this register refer to a large variety of different, often highly abstract entities like the law, political programs, and freedom. Given the abstract nature of many of these Subjects, it is not surprising that nonconcrete Inanimates have a lower probability to be paired with a sentient verb compared to other registers.

8 Theme in English originals

8.1 Results

This chapter will outline the results of the Theme structure analysis in English original writing. The same Theme-related aspects as in German will be considered here with the exception of Subject identifiability, which is not particularly relevant in an intralingual analysis of EO. Again, three Theme hypotheses for English were analyzed, namely the first element hypothesis, the first experiential element hypothesis, and the Subject hypothesis. The variables of multiple Themes and marked Themes are very much dependent on the selected Theme hypothesis, which is why the results of all three hypotheses will be contrasted in detail. The relative distributions of other Theme measures like participant Themes or Subject Theme sentience are less affected by the Theme hypothesis. For these analyses, the results of only one of the three hypotheses will be presented. The choice of Theme hypothesis is the same as in the analysis of German originals (see Chapter 7.1).⁷⁷ In practice, the relative frequencies for all analyses are again almost identical between the three hypotheses and it would have made very little difference if other hypotheses were chosen for each of the analyses. The few significant differences between the hypotheses will be pointed out in the running text. All tables and figures illustrating the results of the other Theme hypotheses are included in the Appendix.

⁷⁷ The results for Subject Agency, process types, and participant role analysis will be based on the Subject hypothesis since this hypothesis ensures that every Subject and at least one participant role per process is being considered. This does not have to be the case for the First Element and first experiential element hypothesis, which can be instantiated by just a non-experiential or circumstantial Theme. For circumstance Theme distributions, I used the first experiential element hypothesis, which considers all those circumstance Themes that make up the first experiential element in the clause. Since the Subject hypothesis allows for multiple experiential Themes, it also includes additional circumstance Themes, which are not part of the Theme in the first experiential element hypothesis. However, I wanted to put the focus on the very first circumstance Theme as the primary point of departure and therefore decided against considering multiple circumstance Themes for this analysis.

8.1.1 Multiple Themes and Theme Markedness

8.1.1.1 First element hypothesis

	Overall		FICTION		INSTR		SPEECH		TOU	
Total Clauses	4496		1241		845		1430		980	
Experiential Themes	3955	88.0%	985	79.4%	803	95.0%	1248	87.3%	919	93.8%
Subject Themes	3142	79.4%	829	84.2%	612	76.2%	982	78.7%	719	78.2%
Circumstance Themes	731	18.5%	127	12.9%	187	23.3%	244	19.6%	173	18.8%
Complement Themes	42	1.1%	16	1.6%	0	0%	4	0.3%	22	2.4%
Predicator Themes	4	0.1%	1	0.1%	0	0%	0	0%	3	0.3%
Textual Themes	411	9.1%	215	17.3%	30	3.6%	134	9.4%	32	3.3%
Interpersonal Themes	130	2.9%	41	3.3%	12	1.4%	48	3.4%	29	3.0%
Cleft	36	0.9%	12	1.2%	4	0.5%	18	1.4%	2	0.2%

Table 10 Basic Theme distribution in EO (1st element)

Table 10 shows a summary of the basic Theme distributions of the first element hypothesis of EO. The first element hypothesis considers only the very first element in the English clause, which can be almost any experiential or non-experiential constituent. The question of multiple Themes and average Theme length is irrelevant for this hypothesis since the Theme always consists of just one element by definition. Since the Theme is arguably a wave with decreasing thematic meaning, the nature of the very first element is quite crucial even for other Theme hypotheses. Some columns in Table 10 are divided into two numbers, with the first representing absolute frequencies and the second relative distribution. For example, INSTR include 803 experiential Themes, which represent 95% of all Themes in total.

4496 clauses of the EO sub-corpus were analyzed. The absolute numbers per register differ noticeably, which is because the entire register was analyzed, and the number of relevant clauses varied. For one, the registers are not normalized according to number of clauses but according to word count, so that registers with higher word number per clause will also feature fewer clauses. For technical reasons, only the very first independent clause per clause complex was considered, so that registers with longer clause complexes will also have fewer analyzed clauses. Lastly, only independent declarative clauses

were considered. Theme structures of interrogative and imperative clauses were not analyzed. This methodological decision affected the registers dissimilarly.

Despite the fact that only the opening element of every clause was analyzed, and this position represents the default place for many textual and interpersonal elements, the relative frequency of experiential Themes is still by far the highest with 88% of all Themes. Textual Themes are the most common non-experiential Theme type, making up 9.1% of all Themes, followed by interpersonal Themes with 2.9%. Subject Themes, unsurprisingly, account for the majority of experiential Themes with 79.4%. These percentages are in line with the results of Matthiessen (1995a) and Fawcett (2008). The only relatively common marked Theme type is circumstance Themes, which represent 18.5% of all experiential Themes. At 1.1% and 0.1% respectively, Complement Themes (see example (140); Complement Theme in bold) and Predicator Themes (see example (141); Predicator Theme in bold) were used as the point of departure in only few cases.⁷⁸

(140) **Of cash**, he had none.
[E2G_FICTION_004]

(141) **Overlaying** it all is the hand of the Viking earls, who held sway in the islands until the 15th century.
[E2G_TOU_002]

There is some variation in Theme types between the registers. Most affected by register characteristics are textual Themes, whose relative frequencies range from 3.3% in TOU to 17.3% in FICTION. The number of interpersonal Themes is relatively stable at around 3% in all registers except for INSTR, where it is by far the lowest with 1.4%. Notably, Theme markedness is surprisingly even across all registers.

The unmarked Theme choice of Subject is the most common, making up around 78% of experiential Themes in all registers. Only FICTION diverges slightly with a relatively high percentage of Subject Themes at 84.2%. The frequency of marked Themes in total is similar in the four registers, but the distribution of Theme types is uneven. Circumstance Themes are most common in INSTR, which do not feature a single other marked Theme.

⁷⁸ The reason why the relative frequencies of the individual experiential Themes do not add up to 100 is because of cleft constructions, which arguably have an experiential Theme. However, their Theme structure is too unique to simply count them as Subject Themes, which is why they are listed separately.

SPEECH and TOU have a comparable number of circumstance Themes, but TOU also includes a relatively high number of Complement Themes. FICTION has the lowest number of marked Themes generally and circumstance Themes in particular, which corresponds to its high number of unmarked Themes. Register is a significant predictor of experiential Theme types ($\chi^2 = 14.199$, $df = 3$, $p\text{-value} = 0.00265$).

8.1.1.2 First experiential element hypothesis

	Overall		FICTION		INSTR		SPEECH		TOU	
Total Clauses	4496		1241		845		1430		980	
Total Themes	5097		1543		887		1625		1042	
Single Themes	3955	88.0%	985	79.4%	803	95.0%	1248	87.3%	919	93.8%
Multiple Themes	541	12.0%	256	20.6%	42	5.0%	182	12.7%	61	6.2%
Avg. # of Themes	1.13		1.24		1.05		1.14		1.06	
Experiential Themes	4496		1241		845		1430		980	
Subject Themes	3568	79.4%	1039	83.7%	641	75.9%	1124	78.6%	764	78.0%
Circumstance Themes	829	18.4%	166	13.4%	200	23.7%	277	19.4%	186	19.0%
Complement Themes	46	1.0%	18	1.5%	0	0.0%	4	0.3%	24	2.4%
Predicator Themes	4	0.1%	1	0.1%	0	0.0%	0	0%	3	0.3%
Textual Themes	431	8.5%	233	15.1%	30	3.4%	136	8.4%	32	3.1%
Interpersonal Themes	170	3.3%	69	4.5%	12	1.4%	59	3.6%	30	2.9%
Cleft	49	1.1%	17	1.4%	4	0.5%	25	1.7%	3	0.3%

Table 11 Basic Theme distribution in EO (1st exp. element)

Table 11 demonstrates the same Theme distributions based on the first experiential element hypothesis. For this Theme hypothesis, everything up to and including the first experiential element is considered, which is the most commonly used Theme hypothesis for English, advocated by Halliday and Matthiessen (2014). This hypothesis allows for multiple Themes whenever a textual or interpersonal element is positioned at the beginning of the clause. Naturally, the number of Themes overall has increased compared to the first element hypothesis and the number of experiential Themes is now identical to the total number of clauses. 12% of English clauses have at least one non-experiential Theme element, representing the number of multiple Themes. The distribution of experiential

Theme has remained basically identical overall. The percentage of textual Themes has dropped, but interpersonal Themes have surprisingly increased. In fact, 23.5% of interpersonal Themes are paired with a textual or another interpersonal Theme, which is the highest among all the different Theme types. To compare, only 11.9% of Subject Themes, 11.8% of circumstance Themes, and 8.7% of Complement Themes share the thematic space with a non-experiential Theme.

The number of multiple Themes and average Theme count shape up quite differently across the four registers. FICTION has a much higher number of multiple Themes at 20.6%, while INSTR and TOU boast only 5% and 6.2% of multiple Themes respectively. Register is a significant predictor of Theme number in the first experiential element hypothesis ($\chi^2 = 22.699$, $df = 3$, $p\text{-value} = 4.665e-05$). These distributions are directly linked to the use of non-experiential Themes. FICTION and SPEECH have more cases of non-experiential Themes than multiple Themes, while the numbers in INSTR and TOU are virtually the same. This means that the latter two registers only allow a maximum of one textual or interpersonal Theme in their Themes, while FICTION and SPEECH can also combine multiple non-experiential Themes (see example (142); textual Themes underlined, interpersonal Themes in bold). Interestingly enough, while the overall frequency of circumstance Themes has remained almost the same, their numbers have increased in all registers but SPEECH. This increase is, however, only slight. Thus, it is not surprising that Register remains a significant predictor variable of experiential Theme types in the first experiential element hypothesis ($\chi^2 = 12.062$, $df = 3$, $p\text{-value} = 0.00717$).

(142) But | then, | **of course**, | **I hasten to add**, | *there are many valets who would never dream of indulging in this sort of folly.*

[E2G_FICTION_006]

8.1.1.3 Subject hypothesis

	Overall		FICTION		INSTR		SPEECH		TOU	
Total Clauses	4496		1241		845		1430		980	
Total Themes	6067		1752		1088		1936		1291	
Single Themes	3178	70.7%	842	67.8%	616	72.9%	1000	69.9%	720	73.5%
Multiple Themes	1317	29.3%	399	32.2%	229	27.1%	430	30.1%	259	26.5%
Avg. # of Themes	1.35		1.41		1.29		1.35		1.32	
Experiential Themes	5413		1439		1046		1731		1197	
Subject Themes	4437	82.0%	1222	84.9%	839	80.2%	1400	80.9%	976	81.5%
Circumstance Themes	865	16.0%	179	12.4%	201	19.2%	296	17.1%	189	15.8%
Complement Themes	46	0.8%	18	1.3%	0	0%	4	0.2%	24	2.0%
Predicator Themes	8	0.1%	2	0.1%	0	0%	1	0.1%	5	0.4%
Textual Themes	434	7.2%	233	13.3%	30	2.8%	137	7.1%	34	2.6%
Interpersonal Themes	220	3.6%	80	4.6%	12	1.1%	68	3.5%	60	4.6%
Cleft	57	1.1%	18	1.3%	6	0.6%	30	1.7%	3	0.3%

Table 12 Basic Theme distribution in EO (Subject)

Table 12 represents the Theme distribution of the Subject hypothesis, where everything is considered thematic up to and including the Subject of the clause. In this Theme hypothesis, the number of multiple Themes becomes quite large, as any Theme is considered multiple if it does not open precisely with the Subject. That being said, it is quite surprising that still over 70% of Themes are simple, which is an index of the high number of clauses that begin with the Subject in English.

The relative frequency of Subject Themes has increased, and conversely circumstance Themes, Complement Themes, and textual Themes have decreased. This effect is hardly surprising as now every clause must contain a Subject Theme, which dilutes the number of any other Theme type. The only exception are interpersonal Themes, which have again increased in relative frequency compared to the previous Theme hypothesis. Register continues to significantly predict the distribution of experiential Theme types ($\chi^2 = 12.078$, $df = 3$, $p\text{-value} = 0.00712$).

The two registers that previously had the lowest number of multiple Themes and average Theme count, INSTR and TOU, also have the lowest numbers in the Subject hypothesis. However, while the numbers were drastically different in the first experiential element hypothesis, they have almost evened out in the Subject hypothesis. Multiple Themes

in INSTR, in particular, have increased noticeably. This is due to a high number of circumstance Themes in this register, which previously represented simple Themes but now include an additional Subject Theme as well. In the Subject Hypothesis, Register is not a significant predictor of Theme number any longer ($\chi^2 = 6.1753$, $df = 3$, $p\text{-value} = 0.1034$).

8.1.2 Participant Themes and process types

	Overall	FICTION	INSTR	SPEECH	TOU
Total processes	4452	1214	838	1402	998
Material	1747	379	521	568	279
Subject Actor	1423	351	380	478	214
Subject Goal	196	19	76	49	52
Subject Initiator	112	5	63	34	10
Complement Goal	3	2	0	0	1
Other	13	2	2	7	2
Mental	517	230	46	182	59
Subject Senser	468	219	38	174	37
Subject Phenomenon	48	10	8	8	22
Complement Phenomenon	1	1	0	0	0
Relational	1651	402	227	478	544
Attributive	1318	318	186	365	449
Subject Carrier	1281	306	186	361	428
Complement Attribute	37	12	0	4	21
Identifying	333	84	41	113	95
Subject Token	244	43	36	90	75
Subject Value	81	37	5	20	19
Other	8	4	0	3	1
Verbal	301	103	28	151	19
Subject Sayer	273	97	21	142	13
Subject Verbiage	11	2	3	1	5
Subject Receiver	14	2	4	7	1
Complement Verbiage	1	1	0	0	0
Other	2	1	0	1	0
Behavioral	78	63	2	3	10
Subject Behaver	78	63	2	3	10
Existential	158	37	14	20	87
Process	158	37	14	20	87
Empty Subjects	31	26	1	2	2

Table 13 Participants and process types in EO

Table 13 shows the absolute distribution of participant Themes and process types. Process types are analyzed based on the participant roles of their Themes, which means these results do not include those clauses that do not have a participant Theme, in particular in the case of empty Subject Theme. Table 13 lists the most common participant Themes per process type. Infrequent participant roles were summarized under *Other*, which included, for instance, Beneficiary Themes in material processes and Target Themes in verbal processes.

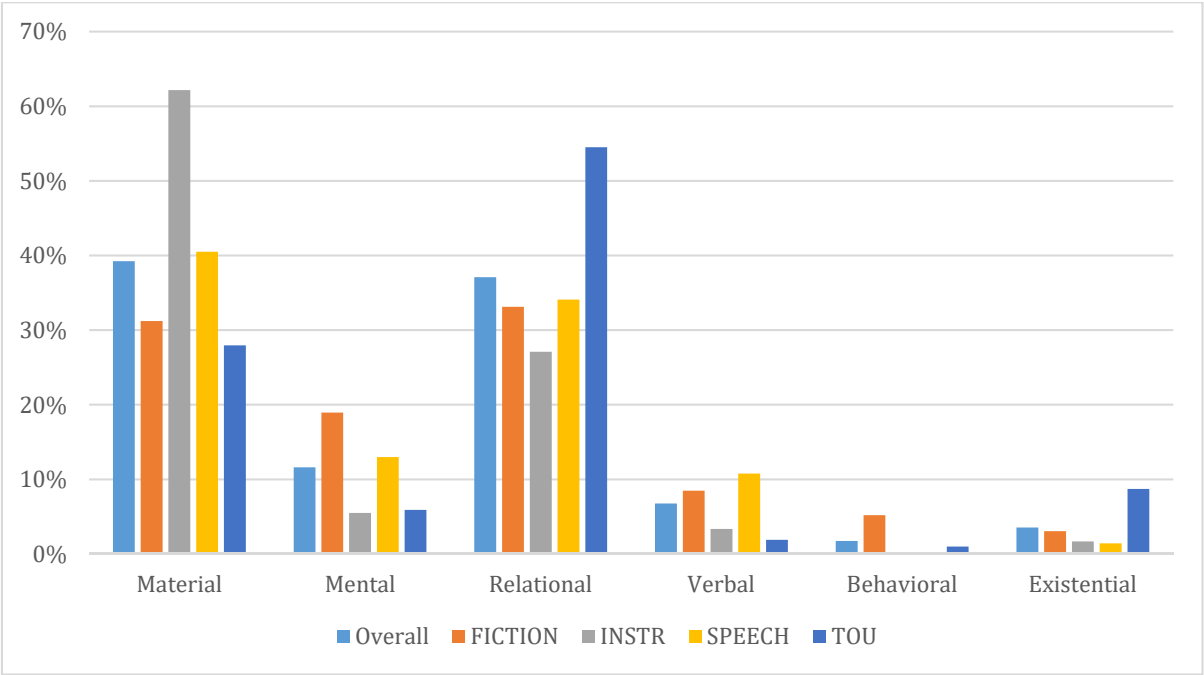


Figure 7 Process type distribution in EO

Figure 7 shows relative distributions of process types. Two types are clearly the most common, namely material, and relational processes, which are almost identical in number, at 39.2% and 37.1% respectively. Mental processes come in third place, with 11.6% overall, followed by verbal processes with 6.8%. Existential processes (3.5%) and behavioral processes (1.8%) are by far the most peripheral process types.

The divergences in terms of process type frequencies between the four registers is very noticeable. INSTR has by far the highest number of material processes at 62.2%, while also featuring the lowest number of relational processes. TOU is the exact opposite with an exceptionally high number of relational processes at 54.5% but also the lowest number of material processes. FICTION and SPEECH are distributed more evenly with similar percentages for the two dominant process types. Mental and verbal processes are

used most in FICTION and SPEECH and merely make up a small part of the process type distribution of INSTR and TOU. Behavioral processes can really only be found in FICTION, where they make up a surprisingly large percentage of 5.2%. Existential processes are evenly distributed among the four registers with the only exception being TOU, which uses an above average number of existential processes at 8.7%. Register is a very strong predictor of process types overall ($\chi^2 = 41.236$, $df = 3$, $p\text{-value} = 5.828e-09$) as well as participant Theme distribution ($\chi^2 = 37.228$, $df = 3$, $p\text{-value} = 4.118e-08$).

Participant Themes realized as Subjects are by far the most common types of participant Themes. They are mostly realized as the first participant of a process like Actor or Sayer. Existential processes do not have a first participant, so the semantically empty Subject *there* combined with the process is the unmarked Theme choice. The number of different participant Themes is highest in material processes due to their large number of participant roles in general. Most relational processes are attributive, which is true overall and per register. Even though Token and Value participants are reversible in identifying relational processes, the majority of Subject Themes represent Tokens. This seems to suggest that it is more common to use the concrete rather than the abstract entity as the point of departure in EO.

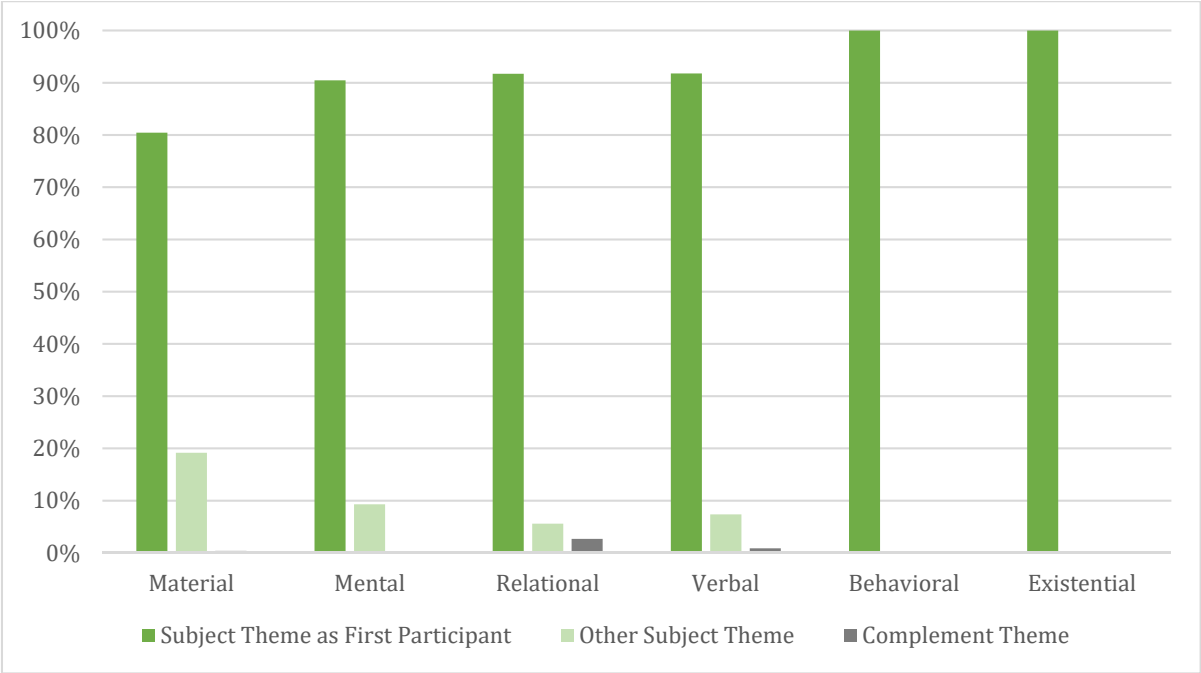


Figure 8 Participant roles and Theme types in EO

Figure 8 visualizes the distribution of the different participant Themes, separated into Subject Themes as first participants, Subject Themes as other participants, and Complement Themes. Figure 8 only shows the overall data and is not divided into the different registers. Predicator Themes in existential processes, which technically include the Subject, were counted as first participant Subject Themes.

Both behavioral and existential processes have only the first participant as participant Themes. The process type with the highest number of Subject Themes that are not the first participant is the material process at 19.2%. Passive constructions are common in material processes, resulting in a high number of Subject Themes that map onto Goal participants. Furthermore, material processes allow initiating constructions where the Initiator participant makes or enables the Actor to carry out the process.⁷⁹ Such constructions, like (143) were especially common in INSTR, where a computer program or a device allows the user to perform an action.

(143) *Kopete lets you create groups to sort your contacts.*

[E2G_INSTR_004]

The number of passive constructions in verbal and mental processes is similarly high, making up the majority of *Other Subject Themes*. Passive constructions in relational processes are only possible in one particular subtype and are thus quite rare. The majority of *Other Subject Themes* comes from Value Subjects in identifying relational processes. Complement Themes are uncommon in all process types except for relational processes, where they make up 2.3% of participant Themes. The majority of these Complement Themes are circumstantial Attributes, formally realized as prepositional phrases (see example (144); Complement Theme in bold). It looks as if verbal processes feature a higher number of Complement Themes compared to the other processes. In reality, there are only two cases of Complement Themes, which amount to 0.9% due to the generally lower frequencies of verbal processes.

(144) ***Along the famous rock-bound coast of Maine** are lighthouses, sandy beaches, quiet fishing villages and thousands of offshore islands.*

[E2G_TOU_003]

⁷⁹ Since the Initiator is the unmarked choice as the Subject Theme in such initiating material processes, it could have been analyzed as first participant as well.

The data for process types and participant Themes was based on the Subject hypothesis. Compared to the other two Theme hypotheses, there are only very small differences. Absolute numbers obviously vary between first element hypothesis, first experiential element hypothesis, and Subject hypothesis, but the relative distribution stays largely the same. Material processes experience a consistent increase from 37.6% in the first element hypothesis to 39.2% in the Subject hypothesis; however, this effect is so small that it may just be due to variance. For all other process types, there was no noticeable trend between the three hypotheses. The distribution of Subject Themes with varying participant roles and Complement Themes remains almost identical as well.

8.1.3 Circumstances

Circumstance Themes are the most common type of marked Themes in English declarative clauses and therefore worth further investigation. As was discussed in Section 5.6, Theme markedness can be analyzed in terms of overall Theme frequency but also in terms of thematic potential. TPot expresses the probability of a constituent to be the Theme of the clause that it is part of. Hence, while certain circumstance types like Place and Time have a high frequency as Themes overall because they are generally used more often than other circumstances, circumstance types like Condition have a higher TPot since they are used as the point of departure more often if they are included in a clause. Depending on the definition of Theme markedness, circumstances with a high TPot are less marked as Theme choice even if their overall occurrence rate is low.

	Overall			FICTION			INSTR			SPEECH			TOU		
Total Clauses	4496			1241			845			1430			980		
Total circumstances	829	2443	33.9%	166	540	30.7%	200	562	35.6%	277	800	34.6%	186	541	34.4%
circumstances per clause	0.54			0.44			0.67			0.56			0.55		
Likelihood of Circ. Theme	18.4			13.4			23.7			19.4			19.0		
Additive	13	18	72.2%	0	0	-	2	3	66.7%	8	8	100%	3	7	42.9%
Behalf	19	50	38.0%	4	10	40%	8	14	57.1%	1	2	50%	6	24	25.0%
Comitative	15	131	11.5%	2	29	6.9%	1	16	6.3%	4	40	10%	8	46	17.4%
Concession	24	39	61.5%	3	11	27.3%	0	2	0%	15	18	83.3%	6	8	75.0%
Condition	195	282	69.1%	20	25	80%	132	187	70.6%	27	50	54.0%	16	20	80%
Comparison	18	51	35.3%	5	29	17.2%	2	4	50%	9	15	60%	2	3	66.7%
Distance	0	6	0%	0	2	0%	0	2	0%	0	1	0%	0	1	0%
Default	0	1	0%	0	0	-	0	1	0%	0	0	-	0	0	-
Degree	0	5	0%	0	1	0%	0	0	-	0	3	0%	0	1	0%
Duration	29	76	38.2%	10	31	32.3%	0	5	0%	11	19	57.9%	8	21	38.1%
Frequency	6	44	13.6%	3	9	33.3%	1	11	9.1%	0	13	0%	2	11	18.2%
Guise	29	61	47.5%	3	5	60%	0	10	0%	12	21	57.1%	14	25	56.0%
Matter	10	21	47.6%	1	5	20%	1	2	50%	7	12	58.3%	1	2	50%
Means	40	128	31.3%	3	11	27.3%	9	56	16.1%	25	52	48.1%	3	9	33.3%
Place	135	520	26.0%	23	105	21.9%	18	107	16.8%	34	115	29.6%	60	193	31.1%
Product	0	9	0%	0	1	0%	0	3	0%	0	2	0%	0	3	0%
Purpose	22	253	8.7%	0	39	0%	11	68	16.2%	10	115	8.7%	1	31	3.2%
Quality	15	184	8.2%	10	58	17.2%	0	29	0%	4	82	4.9%	1	15	6.7%
Reason	27	104	26.0%	6	40	15.0%	4	11	36.4%	6	27	22.2%	11	26	42.3%
Source	8	11	72.7%	0	1	0%	0	0	-	8	10	80%	0	0	-
Time	221	444	49.8%	73	127	57.5%	11	31	35.5%	93	191	48.7%	44	95	46.3%
Viewpoint	3	5	60%	0	1	0%	0	0	-	3	4	75.0%	0	0	-

Table 14 Circumstances and thematic potential in EO

Table 14 shows a detailed account of all 22 circumstance types in EO. The first number in each column represents the absolute number of circumstance Themes followed by the absolute number of occurrences overall. The final number represents the TPot of each circumstance type. As an example, there are 76 cases of circumstances of Duration overall in the corpus, of which 29 are the Theme of their clause. This amounts to a TPot of 38.2%.

The three most common circumstance Themes by far are Time, Condition, and Place. Together, they make up 66.5% of all circumstance Themes in EO. Twelve circumstance types were used as Themes between 10 and 40 times. The remaining seven circumstances have Theme frequencies below 10, with circumstances of Distance, Degree, Default, and Product never being used as Themes.

Around 54% of all English clauses include at least one circumstance, which is used as the point of departure in 33.9% of cases. Although circumstance Themes are generally considered marked Themes in English, it is noticeable that the spread between Subject Themes and circumstance Themes is roughly only two to one if a clause contains both a Subject and a circumstance. As was pointed out earlier, the register with the highest number of circumstance Themes, at 23.7%, is INSTR, while FICTION features the lowest amount of circumstance Themes at 13.4%. Circumstances in INSTR also have the highest overall TPot of 35.6%, but TPot is fairly evenly distributed across the four registers.

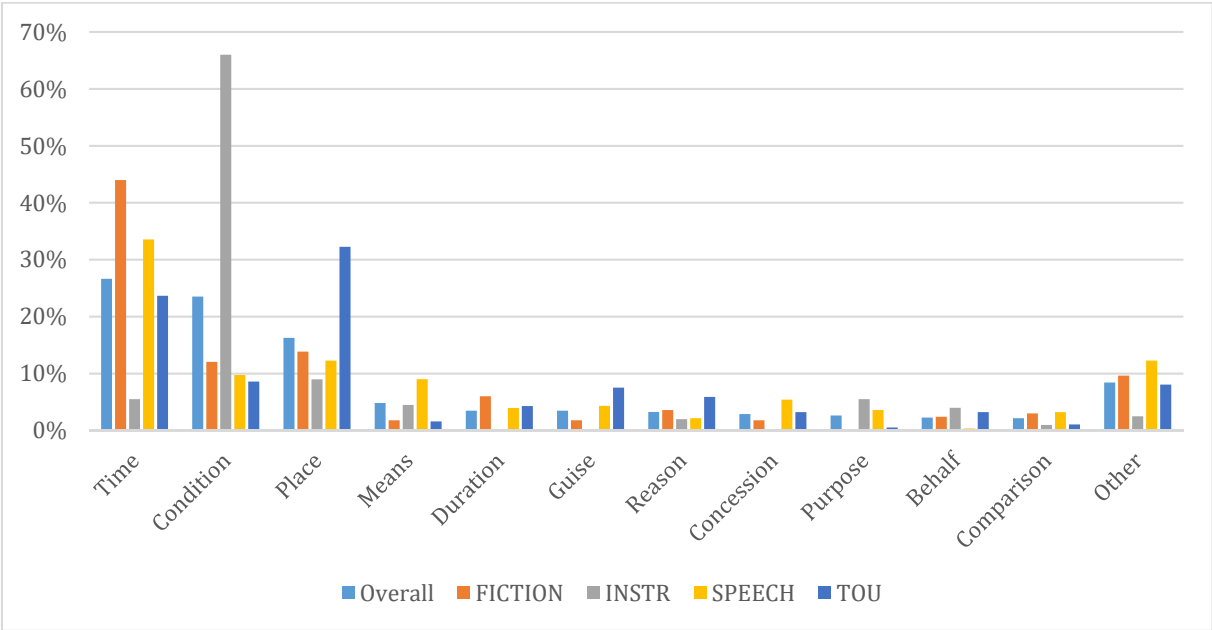


Figure 9 Circumstance Theme distribution in EO

Figure 9 shows the relative frequencies of circumstance Themes per register. For instance, Time Themes make up 23.5% of circumstance Themes overall, but only 8.6% of circumstance Themes in TOU. Since the number of clauses analyzed varied between the registers, relative frequencies are more meaningful than absolute frequencies. Only those circumstances that made up at least 2% of all circumstance Themes were included in the graph. All other circumstances were summarized under *Other*.

Figure 9 reveals noticeable differences in circumstance distributions between the registers. All of the three most common circumstance Themes predominantly come from one register and have a significantly lower occurrence rate in all other registers. The most evenly distributed of the three is Time, which is used relatively frequently in FICTION, SPEECH, and TOU. Time Themes make up 26.7% of circumstance Themes overall, but

44% in FICTION. Compared to that, the number Time Themes in INSTR is surprisingly low at 5.5%. This makes Time Themes in INSTR just as common as Purpose Themes, which are only the ninth most frequent circumstance Theme overall. Even though Condition is the second most common circumstance Theme in total, it is not particularly frequent in FICTION, SPEECH, and TOU. In fact, all three registers have more cases of Place Themes. What increases the overall number of Condition Themes substantially is their high number in INSTR, where they make up a remarkable 66% of circumstance Themes. Notably, this is by far the highest frequency of any circumstance Theme in a register. Place Themes are again more evenly distributed at around 10% in each register, with the exception of TOU, where they occur most often with 32.3%.

Of the circumstance types with medium frequencies, Reason Theme is most evenly distributed. All other circumstances are either characteristic of one register, entirely non-existent in another register, or both. INSTR is the most unique, as it does not include any Duration, Guise, and Concession Themes despite their sizable frequencies in the other registers. At the same time, INSTR include many Purpose Themes, which are less common in the other registers and even non-existent in FICTION. TOU features a lot of Guise and Reason Themes, which are almost as common in that register as Conditions. The category *Other* features a variety of different circumstance types, which explains why they add up to a generally high frequency in all registers. INSTR is again the exception, where peripheral circumstance Themes hardly show up at all. Despite these seemingly substantial differences between the registers, Register is not a significant predictor of circumstance Theme types ($\chi^2 = 5.7401$, $df = 3$, $p\text{-value} = 0.125$).

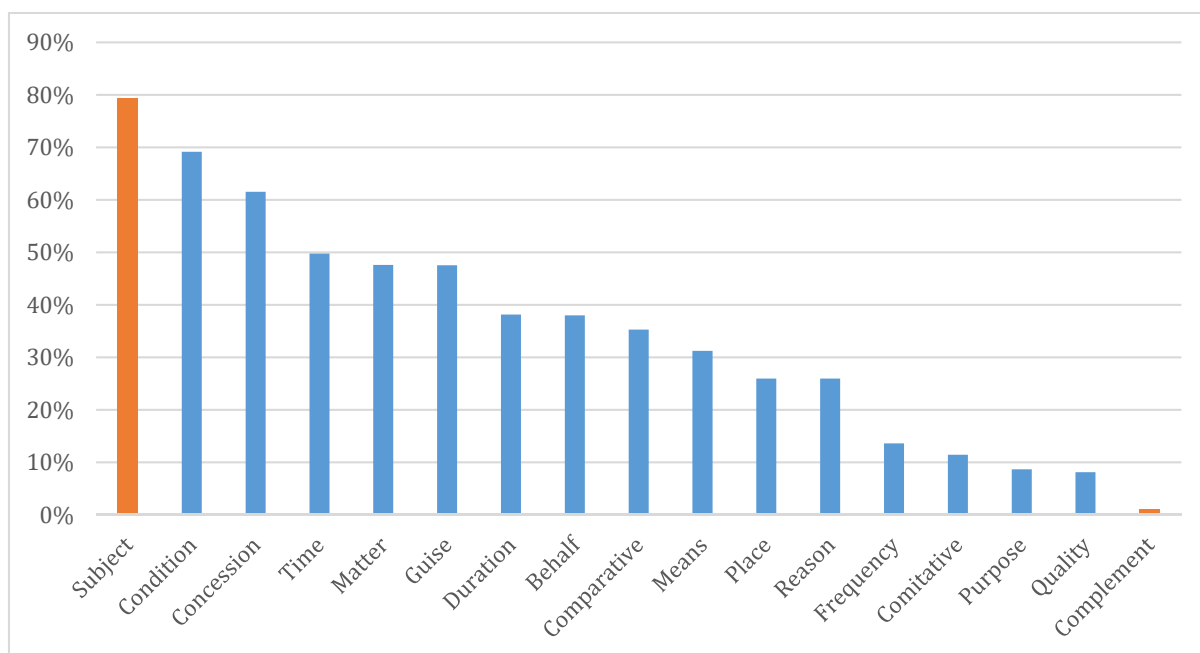


Figure 10 Thematic potential of circumstances in EO overall

Figure 10 represents the thematic potential of circumstance Themes overall. It includes only those circumstances that were found at least 20 times overall. As reference, the TPots of Subjects and Complements are also included.

The two circumstances with the highest TPots are Condition (see example (145) ; circumstance Themes in bold) and circumstances of Concession (see example (146)) with 69.1% and 61.5% respectively. They are also the only two circumstance types that are more likely to be positioned in the Theme than in the Rheme in English.⁸⁰

(145) ***If it does***, you can verify the site's certificate.

[E2G_INSTR_009]

(146) ***Although we tend to focus on our disputes***, most of our economic relationship is conflict-free.

[E2G_SPEECH_014]

Circumstances of Time have the third highest TPot of 49.8%, which means they have an almost even Theme-Rheme distribution in English. Circumstances of Matter and Guise also come close to an even distribution with a TPot of 47.6% and 47.5% respectively. The

⁸⁰ It should be mentioned that there are two circumstances that have an even higher TPot than Condition and Concession, namely Additive and Source circumstances. However, they did not fulfill the minimum requirement of 20 occurrences total.

Subject has the highest TPot out of all constituents, as 79.4% of all Subjects are also the experiential Theme of their clause. However, this number can be deceiving since half of the clauses in English do not even contain a circumstance, making the Subject often the only reasonable choice as Theme. If a clause contains a circumstance of Condition or Concession, the Subject is in fact the more marked choice as point of departure.

The majority of circumstances in English have a TPot between 25% and 40%, which is quite high given their status as marked Theme choices. Circumstance TPot seems generally independent from overall frequencies: for instance, despite their high numbers overall, circumstances of Place have a below average TPot of 26%. This means that just because a circumstance is generally used more often does not mean that it also has a higher probability to be the Theme of its clause. Four circumstance types have a notably low TPot in English, namely Frequency, Comitative, Purpose, and Quality (see examples (147) - (150); circumstance Themes in bold). However, the constituent with the by far lowest potential of becoming the experiential Theme of its clause is the Complement with a TPot of only 1%. This leaves no doubt that Complement Themes are by far the most marked Theme Choice apart from Predicator Themes.

(147) **Every March** *the city celebrates St David, Wales' patron saint, with parades and music.*

[E2G_TOU_008]

(148) **With Javier Solana and Chris Patten**, *the US now has people we can call upon when we want to engage the EU on political and security issues.*

[E2G_SPEECH_009]

(149) **To claim on the guarantee**, *you will need to submit proof of purchase to the seller or an authorised repair agent.*

[E2G_INSTR_008]

(150) **Gently** *Keskarrah rocks Greywing against himself.*

[E2G_FICTION_005]

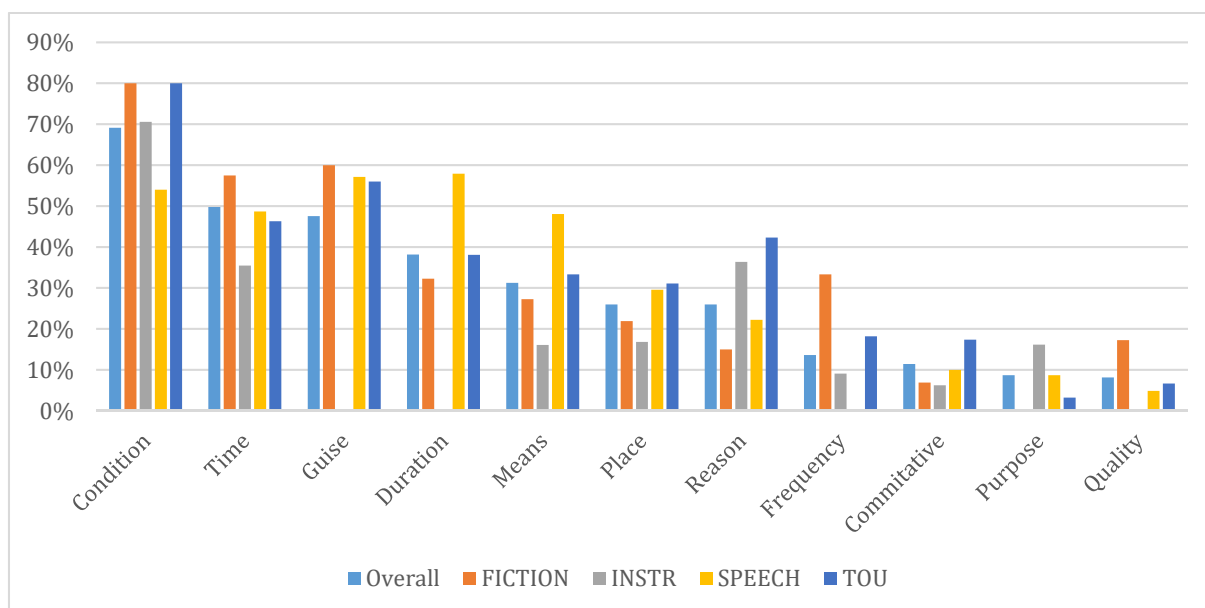


Figure 11 Thematic potential of circumstances in EO overall and per register

Figure 11 shows the same TPot distribution of circumstances, differentiated between the four registers. For this Figure, only those circumstances were included that were found at least ten times in all four registers to ensure a meaningful comparison between the registers. Of the previously shown 15 circumstance types, circumstance of Concession, Matter, Behalf, and Comparison did not meet this requirement, which shows that some of the circumstance TPots are more representative of single registers rather than the language as a whole.

Circumstance TPots vary considerably between the four registers. In many cases, the thematic potential is similar in three of the four registers, with one register having a noticeably higher or lower TPot than the others. The most evenly distributed circumstance Themes in terms of TPot are Condition, Time, and Place. It should be noted that besides Condition, none of the circumstances have a TPot of 50% or higher overall, but three circumstance Themes score over 50% in at least one register, namely Time in FICTION, Guise in FICTION, SPEECH, and TOU, and Duration in SPEECH. This means that these circumstances are in fact less likely to appear in the Rheme than in the Theme, making them a more likely Theme candidate than the Subject in their respective registers. Guise is particularly noteworthy as it has a remarkably high TPot in the aforementioned registers but a TPot of 0% in INSTR, despite the fact that the register includes ten instances of circumstances of Guise overall. While Duration does not have the same high TPot across the registers, INSTR again represents the outlier with a TPot of 0%.

Means TPot scores highest in SPEECH, almost reaching the 50% mark, and lowest in INSTR at 16.1%. Reason Themes are arguably most diverse as they have high thematic potential in two registers, INSTR and TOU, but low thematic potential in FICTION and SPEECH. Frequency Themes are the most likely in FICTION at 33.3% but are not made the Theme a single time in SPEECH even though they are used most often in this register.

A high circumstance frequency overall does not always translate to a high TPot. Even though INSTR has by far the highest number of circumstances of Condition, Conditions have higher TPots in FICTION and TOU. Circumstances of Guise are least common in FICTION, where they also have the highest TPot. SPEECH has the second lowest count but also the highest TPot of circumstances of Duration. On the contrary, circumstances of Comitative, Quality, and Place have the highest TPot in the registers where they also make up the highest relative frequency. A generally consistent relationship between overall use of a circumstance in a register and its thematic potential could thus not be established.

The data on circumstance Themes was based on the first experiential element hypothesis. Between this hypothesis and the Subject hypothesis, the number of analyzed circumstance Themes increased by only 36, which shows that multiple circumstances positioned before the Subject are quite rare. The three circumstance Themes whose absolute frequencies increased the most between the two hypotheses are Place, Reason, and Time. As a consequence, the thematic potential of these three circumstances rises as well. Circumstances of Time in particular are commonly paired with other opening circumstances, mostly with circumstances of Place or other circumstances of Time (see example (151); Time Theme in bold, Place Theme underlined). Such combinations of spatial and temporal circumstances can be argued to establish a circumstantial frame (Brinkmann 1971).

(151) *In Monterrey **last month**, we looked at how to finance development in developing countries.*

[E2G_SPEECH_007]

A larger increase of 98 additional circumstance Themes can be observed between the first element hypothesis and the first experiential element hypothesis. These are circumstances that are preceded by a textual or interpersonal Theme. All circumstances increase at a similar rate of 10-20%. Time again increases the most in total numbers between the two Theme hypotheses, but this is due to their generally high frequency. This goes to

show that there are no circumstances that are particularly likely or unlikely to share the thematic space with a non-experiential Theme.⁸¹ Since the number of total circumstance Themes decreases in the first element hypothesis, circumstance TPot is also lower across all circumstance types.

8.1.4 Subject animacy and sentience

In this section, the relationship between the semantic content of Subject Themes and their relationship to the process will be presented. A particular focus will be put on non-sentient constructions where an inanimate Subject is paired with a process which requires an agentive and/or sentient first participant (see example (152); inanimate Subject in bold).

(152) *And **the Childe's heart** rebelled a little [...].*
[E2G_FICTION_002]

Theoretically, inanimate Subjects do not possess either of these traits, so a combination with such processes should not be plausible. However, due to the relatively strict word order in English, the semantic constraints on Subject and process are less strict to allow more thematic options (Hawkins 1986: 67). Agentive processes include some material, verbal, and behavioral processes. Sentient processes include the same material and verbal processes, since sentience is one aspect of agency, as well as all behavioral and mental processes, which require a sentient first participant by definition.

⁸¹ Two circumstances, Matter and Comparison, increase by 42.9% and 38.5% respectively. However, these are two of the more infrequent circumstance types, so this relatively high increase could just be variance.

	Overall			FICTION			INSTR			SPEECH			TOU		
Total Clauses	4496			1241			845			1430			980		
Relational	1585			371			227			466			521		
Not analyzed	557			135			114			133			175		
Analyzed Subjects	2354			735			504			831			284		
Human	1447			644			205			489			109		
Animal	12	12	100%	12	12	100%	0	0	-	0	0	-	0	0	-
Organization	130	180	72.2%	8	8	100%	1	1	100%	115	156	73.7%	6	15	40%
Machine	46	74	62.2%	0	0	-	46	74	62.2%	0	0	-	0	0	-
Vehicle	2	7	28.6%	2	3	66.7%	0	1	0%	0	0	-	0	3	0%
Concrete Inanimate	40	157	25.5%	13	31	41.9%	13	78	16.7%	4	12	33.3%	10	36	27.8%
Nonconcrete Inanimate	108	323	33.4%	15	31	48.4%	39	108	36.1%	41	146	28.1%	13	38	34.2%
Place	24	82	29.3%	1	1	100%	0	0	-	1	4	25.0%	22	77	28.6%
Time	1	12	8.3%	0	0	-	0	1	0%	0	8	0%	1	3	33.3%
Process	16	60	26.7%	1	5	20%	11	36	30.6%	2	16	12.5%	2	3	66.7%

Table 15 Subject Theme animacy and sentience in EO

Table 15 shows the results of all middle animate and inanimate Subjects and the processes that they were paired with, separated into verbs that require sentience and verbs that do not. The results for each middle animate and inanimate Subject are divided into three numbers. The first number stands for the number of cases in which the Subject type was paired with a sentient verb and thus part of a non-sentient construction. The second number represents the overall occurrence. Note that this does not accurately reflect the frequency of animate and inanimate Subjects in the corpus since relational processes were disregarded entirely. The final number is the relative frequency of non-sentient constructions for that Subject type. As an example, FICTION includes 31 concrete Inanimate Subjects overall, of which 13 are paired with a sentient verb, which amounts to a relative frequency of 41.9%.

Nonconcrete Inanimate Subjects are by far the most frequent non-Human Subject, followed by Organization and concrete Inanimates. Time, Animal, and Vehicle Subjects are used the least. The overall numbers for each Subject do not represent an even spread between the four registers as most of the inanimate Subjects come from one register in particular. All Animal Subjects were found in FICTION and all Machine Subjects in INSTR. The vast majority of Organization Subjects are used in SPEECH and Place Subjects come primarily from TOU. For many Subject types, the overall numbers thus do not represent a meaningful average but rather individual registers. The only two inanimate Subjects that are used consistently throughout the four registers are concrete and nonconcrete

Inanimates. Register is a significant predictor of different Subject Theme types and non-sentient constructions ($\chi^2 = 43.049$, $df = 3$, $p\text{-value} = 2.403e-09$).

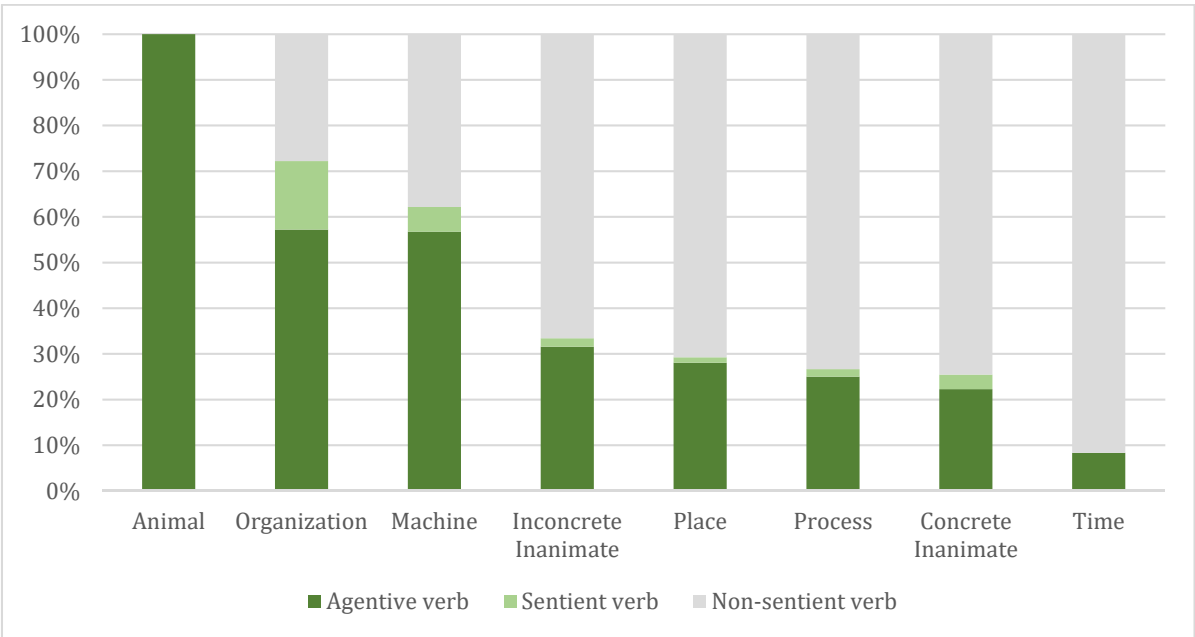


Figure 12 Subject Themes and verb types in EO

Figure 12 illustrates the relative frequencies of the different Subjects being paired with sentient and non-sentient verbs. In this Figure, the verbs are also differentiated according to whether they need an agentive first participant or whether they require only sentience. Most of the processes found are agentive verbs, mostly material processes that require an agent. The only Subject type that is paired with a high number of sentient verbs, mostly mental processes, is Organization (see example (153); Organization Subject in bold). Vehicle Subjects are not included in this Figure due to their low number of occurrences.

(153) **Governments across the Middle East and North Africa** are seeing the need for change.

[E2G_SPEECH_013]

All cases of Animal Subjects were combined with agentive verbs, which is unsurprising since many animals do have sentience and can be the agent of a process. The other middle animate Subject, Organization, also has the second highest rate of non-sentient constructions at 72.2%. Technically, Machines should be categorized as inanimates but Zaenen et al. (2004: 121) argue that they may also be considered of middle animacy given their frequent use with sentient verbs. The data in this project supports this claim as 62.2% of

Machines are paired with sentient verbs, which is considerably higher than for all (other) inanimate Subjects. The division between middle animate and inanimate Subjects is evident given the noteworthy drop in non-sentient construction between Machine Subjects and the next most frequent Subject type, nonconcrete Inanimates. Inanimate Subjects are distributed surprisingly evenly. The frequency of non-sentient constructions ranges from 25.5% to 33.4% across all Subject types. The only exception is Time Subjects, which have the lowest rate at 8.3%.

8.2 Discussion

In this section, the findings of all the Theme-related aspects in EO will be discussed. There are only two Theme aspects which are heavily dependent on the choice of Theme hypothesis, namely multiple Themes and marked Themes. All other measures are affected by the extent of the Theme in absolute terms, but not so much in relative terms. Consequently, the relative distribution of individual circumstance Themes, for instance, stays relatively consistent irrespective of the number of elements considered thematic.

Obviously, the average number of Themes increases the further the Theme can extend. The Subject hypothesis thus has the highest number of multiple Themes, while the first element hypothesis has zero multiple Themes by definition. And yet, it is interesting that in over 70% of cases, the choice of hypothesis makes no difference, as these are the cases where the Subject occupies the very first position of the clause and therefore produces the same Theme zones across all three hypotheses.

With the increase in potential Theme candidates, the absolute and relative distributions of individual types of Themes changes as well. Interestingly, the relative frequencies of experiential Themes stay almost exactly the same between the first element hypothesis and the first experiential element hypothesis. This shows that the selection of non-experiential Themes has no bearing on the choice of experiential Theme, as all of them are similarly likely to be paired together. This supports Halliday and Matthiessen's (2014: 110) claim that these Theme choices are part of different sub-systems.

The relative number of marked Themes decreases between the first experiential element hypothesis and the Subject hypothesis. On the one hand, this was to be expected because in the Subject hypothesis, every clause must contain a Subject Theme by definition, which is not the case for the two other main types of experiential elements, circumstantial Adjuncts and Complements. On the other hand, this shows that it is highly unlikely for a marked Theme to be followed by another marked Theme in the same clause. In fact, there is not a single additional Complement Theme in the Subject hypothesis, meaning that there was no instance in all four registers where a Complement Theme followed a circumstance Theme or another Complement Theme. There are some Themes in the Subject hypothesis that do include multiple circumstances, but such cases, like (154), are quite infrequent (Circumstance Themes in bold).

(154) ***Just yesterday***, | ***at my request***, | *the State Department designated 39 entities as terrorist organizations pursuant to the USA PATRIOT Act.*

[E2G_SPEECH_001]

While the English language does allow experiential Themes other than the Subject, the Subject Theme still needs to come relatively early. This suggests processing effects (Diessel 2005) since the hearer needs to know the order of Subject and Finite to be able to interpret the speech function of the clause and therefore these two elements are rarely delayed. As a matter of fact, circumstance Themes that are followed by another circumstance Theme are shorter than the average of circumstances with 22.2 characters compared to the 31.1-character average. This further supports the assumption that the zone before the Subject is relatively short, most likely for processing reasons.

The differences in non-experiential Themes between the different hypotheses are noteworthy. The relative frequency of textual Themes is highest in the first element hypothesis. This makes sense, as textual elements are typically positioned at the head of the clause. Some textual Themes can even not be positioned anywhere else, which Halliday and Matthiessen (2014: 110) interpret as a development of these elements toward becoming inherently thematic. The decrease of the relative frequencies of textual Themes was expected as a result of the Theme extending further into the clause, since now more Theme elements are considered, and the relative frequency of textual Themes is diluted.

The same effect could not be observed for interpersonal Themes. Their absolute and relative frequencies increased between the three hypotheses. One explanation for this is the natural order of textual and interpersonal Themes. If both elements are used as points of departure, textual Themes typically precede interpersonal Themes (Halliday and Matthiessen 2014: 107). This, however, does not explain the increase in interpersonal Themes from the first experiential element hypothesis to the Subject hypothesis. This effect can be largely attributed to the decision of treating the finite verbal group as an interpersonal, finite-verbal-operator Theme in clauses like (155), where a circumstantial Attribute is used as the Complement Theme (Complement Theme in bold, finite verbal operator underlined).

(155) ***Nearby*** *is* ***St Magnus Cathedral***, *founded in 1137, with some of the finest Norman work in Scotland.*

[E2G_TOU_008]

This also explains the high increase in interpersonal Themes in tourism leaflets, where such relational constructions are particularly common. If these kinds of interpersonal Themes are disregarded, the relative number of interpersonal Themes decreases from 3.3% to 2.9% between the first experiential element hypothesis and the Subject hypothesis. Hence, while some interpersonal Themes, especially comment Adjuncts, can be versatile in their positioning, the majority of late interpersonal Themes either follows textual Themes or are finite verbal operators in circumstantial relational clauses.

The frequency of textual and interpersonal Themes differs between the four registers, which affects the average number of Theme elements in the first experiential element hypothesis. Register differences of this kind were to be expected. It is interesting, however, that these differences in Theme number even out, for the most part, in the Subject hypothesis, where the number of multiple Theme is roughly around 30% across all registers. This is further corroborated by the fact that Register is a significant predictor of multiple Themes in the first experiential element hypothesis but not in the Subject hypothesis. By extension, the kinds of Themes preceding the Subject are very different in the registers, but the likelihood of any element preceding the Subject in a clause is rather similar. While the Theme zone in front of the Subject is limitless in theory, it seems to only tolerate a set number of elements in practice. And this set number is largely independent of register.

As was already mentioned in Section 7.2, Neumann's (2014: 303) Theme analysis of the FICTION register does not correspond to the results reported here, even though the same corpus was used. This discrepancy is primarily due to the inclusion of all mood types in Neumann (2014) as well different Theme categorizations.

Regarding process types and participant Themes, the choice in Theme hypothesis is mostly inconsequential. The small increase in material processes between the first element hypothesis and the Subject hypothesis seems to suggest that this process type is more likely to be accompanied by a circumstance Theme or a non-experiential Theme. However, this effect is so slight that it seems hardly significant. The relative distribution of participant role Themes does not change at all for each of the process types. Regardless of Theme hypothesis, the data solidifies the assumption and previous empirical evidence that Complement Themes are very rare, and that the majority of Subject Themes take on the experiential role of the first participant of the process.

Circumstance Themes are not noticeably affected by the extent of the Theme in English either. In terms of pure frequencies, the data shows that circumstances are a more marked choice for Theme than Subjects. When considering thematic potential, it is surprising that Subjects on average are only twice as likely to be used as Theme over a circumstance if both are included in the clause. With an overall thematic potential of 33.9%, circumstances are common thematic choices and can easily occupy both pre- and post-verbal positions in the clause (Thompson 2014: 149). Diessel (2005: 452) also calculated the distribution between initial and final Adjunct clauses. His results on initial Adjuncts are slightly higher than the results reported here: 32.1% in conversation, 37.6% in fiction and 43.7% in science. This discrepancy may be due to the fact that Diessel (2005) only analyzed Adjunct clauses and disregarded circumstances in the form of phrases. This suggests that circumstance form also has a bearing on thematic potential.

This picture changes again when considering the thematic potential of individual circumstance types. Circumstance Themes such as Purpose and Quality are clearly very marked Theme choices and apparently need a lot of contextual motivation to be used as the point of departure. Other circumstances such as Condition, Concession, and Time are just as likely if not more likely to be used as Theme as the Subject if they occur in the same clause. This is not to say that their choice as Theme does not have to be motivated, but these kinds of motivations seem rather common. It is interesting that the thematic potential of circumstances of Time is so much higher than that of circumstances of Place despite the fact that both circumstances are among the most frequent overall. Apparently, it is more common and arguably more important in English to signal a shift in temporal than in physical space in the Theme. The TPot of Condition and Concession is in line with previous research (for example Downing 1991; Biber et al. 1999; Diessel 2005; Freiwald 2016). Schiffrin (1992: 193) suspects that causal Adjuncts may also have a natural disposition to be used thematically. However, the circumstance types that are part of the superordinate category Cause, namely Reason, Purpose and Behalf, all have average or below-average thematic potentials. In general, it is fair to say that circumstance Themes are more marked than Subject Themes both in terms of general frequency and thematic potential. Nevertheless, the data clearly shows that grammatical function is only one factor and that the experiential meaning is almost equally relevant for clause positioning.

In summary, it is surprising how little the choice in Theme hypothesis mattered for the relative frequencies of Theme-related measures. This interpretation is of course only

meaningful if one considers general statistics on Theme distributions. When analyzing thematic progression of texts, the question whether the content of the Subject is part of the thematic development of a text is highly relevant, as this shapes its interpretation significantly. However, in terms of general distributions, the choice of Theme hypothesis is largely irrelevant. Even the first element hypothesis, which is only rarely used in systemic functional analyses, produced very similar results compared to the other more popular hypotheses. This insight is particularly valuable for corpus linguistics, as the first element of a sentence can be easily retrieved automatically.

In the following, the relationship between Theme and register in EO will be discussed. Thematic patterns are highly affected by register differences, as was confirmed by the chi-square tests. Four different registers were chosen to allow some generalizability to the language system as a whole. However, given the fact that the differences in the selected registers were in part so great, a different set of registers may also produce very different general results. For example, Condition Themes made up a large proportion of the Themes overall. Nonetheless, the vast majority of them comes from INSTR and had different registers been included, Condition Themes would not have been as prevalent. Moreover, all analyzed registers are written registers. Spoken registers most likely possess very different thematic patterns, which are not represented in this sample.

The distribution between Subject, circumstance, and Complement Themes is significantly predicted by the choice of register in all three Theme hypotheses. Register is also a significant predictor of Theme number in the first experiential element hypothesis, which shows that the distribution of textual and interpersonal Themes is also highly register-dependent.

The data of GO shows a consistent inversely proportional relationship between non-experiential Themes and marked Themes (see Section 7.1.1). While this relationship is not as consistent in EO, a similar trend can still be observed, as the registers with fewer non-experiential Themes also have more marked Themes, in particular circumstance Themes. One explanation for this, which is just as true for German as it is for English, is that some textual Themes such as *and* and *then* serve very similar functions as circumstances of Time and some circumstances of Place so that the use of both is often not necessary.

Processing considerations may be another reason for this relationship. In the Subject hypothesis, Register does not remain a useful predictor of Theme number. The types of

elements that are positioned before the Subject differ greatly in the different registers. FICTION contains a lot of textual Themes, TOU uses many interpersonal Themes, INSTR has a high number of circumstance Themes, and SPEECH includes a rather diverse selection of non-Subject Themes. Still, the general likelihood of any element to be placed before the Subject in declaratives is rather even between these four registers. This discovery is quite meaningful for the system of Theme in English. Halliday and Matthiessen (2014: 110) argue that textual and interpersonal Themes do not enter the system of marked Themes, as their inclusion in the Theme does not prevent the addition of circumstance or Complement Themes. This is of course generally true, as textual, and interpersonal Themes can be and are paired with circumstance and Complement Themes. However, while non-experiential Themes do not rule out marked Themes outright, these results do suggest that their inclusion make additional non-Subject Themes less likely. This is most likely tied to processing reasons (Diessel 2005). To allow the hearer to identify mood, the speaker needs to arrive at the Subject and the Finite as quickly as possible. Consequently, if the speaker has already fronted a textual element, for example, they seem to try to avoid adding yet another constituent that is not the Subject to the Theme. Admittedly, it could also be a mere coincidence that the four selected registers all include a similar number of non-Subject Themes, which is why the analysis of further registers would be fruitful.

Process type distribution, participant Themes, and Subject Theme types were also shown to be highly dependent on registers. This is tightly linked to differences in fields of discourse (Halliday and Hasan 1985), which will be discussed below. These results also corroborate the interpretation in Section 7.2 that there is a correlation between experiential content of the Theme and register. This supports Fries (1995a) third hypothesis on the relationship between genre and experiential Themes.

In the visual presentation of circumstance Theme types, the differences between the registers appear substantial and yet the result for the different types of circumstance Themes in relation to the registers are not significant. The number of data points is a lot smaller compared to the other Theme analyses. The results on Theme number or Participant Themes, for example, are based on the entire sample of 4496 clauses, while the chi-square test of circumstance Theme types only takes into consideration 829 data points, which represents the total number of circumstance Themes in the four registers. Besides, circumstance Theme distributions are generally comparable between three of the registers, especially the ones of medium frequency. The one register that consistently deviates

from the rest is INSTR, but since this is also the register with the lowest number of clauses overall, these idiosyncrasies may be counteracted. Nevertheless, the result is rather unexpected and suggests that the choice of circumstances is not as dependent on the register as process types are, for example.

Thematic potential of circumstances also varies noticeably between the registers. It should be noted that some variance is to be expected, especially when dealing with relatively low numbers of cases for some of the circumstance Themes. In fact, the three most common circumstance Themes, Time, Condition, and Place, are also the ones that are distributed most evenly. However, some of the differences cannot simply be explained away by variance. Circumstances of Guise have a TPot of around 60% in FICTION, SPEECH, and TOU, making them a more likely thematic choice than the Subject in the same clause. This is opposed to a TPot of 0% in INSTR despite the fact that the register included ten cases of circumstances of Guise overall. Similarly, Duration has a high TPot of 57.9% in SPEECH and fairly high TPots in FICTION and TOU of around 35% but again a TPot of 0% in INSTR. Such great discrepancies go beyond variance and have to be connected to register effects.

There is no consistent relationship between the TPot of a circumstance type and its relative frequency in a register. Some circumstances, like Place, Comitative, and Quality, have their highest TPot in the registers where they made up a large portion of circumstances overall, while other circumstances have their highest TPot in registers where they were relatively uncommon, such as Condition, Guise, Duration, and Frequency. In other words, the fact that a circumstance is common and thus more expectable in a register does not mean that it is more commonly used as the Theme of a clause, nor is the opposite relationship supported by the data.

The reasons for these varying TPots are difficult to assess. One possible explanation is that texts in different registers develop their thematic progression differently, and that for some registers a particular circumstance as the point of departure is a more likely candidate for contextualizing a clause in the text. This is a reasonable interpretation but looking at some of the specific types of circumstances, it does not become immediately apparent why circumstances of Duration are so much more likely to become the Theme in SPEECH than in INSTR despite Duration being relatively uncommon in both registers.

Another explanation may be related to the form and meaning of these circumstances. With 22 circumstances in total, Halliday and Matthiessen (2014) offer a variety of categories to differentiate between distinct circumstantial meanings. And yet, this can still

only be taken as a rough division. Two Adjuncts may belong to the same circumstance type but express fairly different meanings or be realized in different forms. For example, circumstances of Guise in FICTION, SPEECH, and INSTR often represent a comment on the Subject (see example (156); Guise in bold). Such circumstances are also sometimes referred to as detached predicatives or absolute constructions (Biber et al. 1999: 136). In INSTR, on the other hand, circumstances of Guise typically refer to functions of a product, more specifically as what the product can be used, and their use as Theme would be highly marked (see example (157); Guise in bold). So, while both of these circumstantial meanings can be asked for using the question *as what*, the meanings they express are entirely different and therefore their likelihood of becoming Theme also varies. Above that, some of the circumstance categorizations are purely based on methodological decision (see Section 6.3) and could be distributed very differently.

(156) ***As Minsk group co-chairs***, the U.S. and Russia are actively involved with Azerbaijan and Armenia in finding a solution to the problem of Nagorno-Karabakh.

[E2G_SPEECH_003]

(157) *If you have a hard disk with a System Folder on it, you can use it **as a startup disk**.*

[E2G_TOU_002]

Regarding Duration Themes, it is noteworthy that ten out of eleven Duration Themes in SPEECH refer to a duration in the past, for example for how long a government program has run or a political path has been followed (see example (158); Duration in bold). It often seems to function like a marker for chronological order where past accomplishments are outlined first and then the present state follows, similarly to how Time Themes are often used. Circumstances of Duration in INSTR exclusively refer to durations in either the present or future, specifying for how long something is happening or will be happening (see example (159); Duration in bold). They often appear in clauses with a Condition Theme, where first the Condition is specified, followed by the event, followed by the length of the event. In these constructions, the Duration circumstance functions like an outcome, which is unsurprisingly positioned late in the clause. Such subtle meaning differences can be highly influential for the positioning but cannot accurately be captured by a mostly quantitative study like this. A more qualitative analysis of circumstance positioning, form, and meaning would be fruitful.

(158) ***For more than half a century*** Europe worked hard to make intra-European conflict no more than a memory.

[E2G_SPEECH_012]

(159) When input signal is not detected for 16 seconds, the splash screen will appear ***until a signal is detected.***

[E2G_INSTR_010]

Regarding Subject Theme animacy and sentience, it is generally difficult to assess the effect of registers on non-sentient constructions. The frequencies of these constructions differ between the registers, but this is largely due to the different kinds of Subjects and their semantic meanings that are used in these registers. For example, FICTION has by far the lowest number of non-sentient constructions relative to the overall number of clauses, whereas INSTR has the highest number. However, this grand difference is simply due to the fact that most fictional stories in CroCo deal with people while most instruction manuals focus on machines and other concrete objects. As a consequence, FICTION simply has fewer inanimate Subjects that can be part of non-sentient constructions. However, for the inanimate Subjects that FICTION does have, it has the highest probability of non-sentient constructions out of all the registers. This high probability can be attributed to the fact that fiction allows 'unnatural' or impossible narratives, in a way that instruction manuals never could.

The fact that the types of inanimate Subjects are distributed extremely unevenly poses another challenge. Animal Subjects primarily come from FICTION, Organization from SPEECH, Machines from INSTR, and Place from TOU. It is therefore difficult to compare the registers with each other because they share so few of the same inanimate Subject types. The only two kinds of inanimate Subjects that are relatively common in all registers are concrete and nonconcrete Inanimates. Concrete and nonconcrete Inanimates in FICTION have a substantially higher, while concrete Inanimates in INSTR have a substantially lower likelihood of being the Actor, Senser, or Behavior of an event. Other than that, their numbers are fairly similar across the registers.

As was pointed out in the discussion of GO in Section 7.2, concrete and nonconcrete Inanimates are very broad categories. They include a variety of different Subject meanings, which are again highly dependent on registers. For example, concrete Inanimates in INSTR mostly refer to technical equipment like monitors and batteries or digital displays

like boxes, menus, and buttons. Concrete Inanimates in FICTION refer to natural phenomena like wind and darkness or body parts like eyes and hair. Thus, despite all of these meanings being concrete inanimate, it is of little surprise that there would be differences in their likelihood of being the Subject of a non-sentient construction.

Regardless of register differences, the data clearly shows that there is a difference between middle animate and inanimate Subjects and that this division is meaningful. Animal and Organization are much more likely to function as the Subject in non-sentient constructions than, for example, Place or Vehicle. It is surprising how evenly the probabilities of non-sentient constructions map out for inanimate Subjects. As was pointed out in Section 3.5, English has to be more flexible in the way that semantic meanings map onto grammatical functions to compensate for its rigid word order (Hawkins 1986: 67). This development seems to have settled on a relatively consistent average probability for inanimate Subjects.

As was hypothesized earlier, Machine Subjects behave more like middle animate Subjects and should be considered part of that category based on their capability to be the agent of an event. This reveals an interesting perspective on machines in today's society. Animals are living beings and many of them are sentient, while Organizations are not living themselves, but they are metonymies of living, sentient beings. It makes sense that they would be treated differently than truly inanimate objects, places, and events. However, Machines are just as inanimate and non-sentient as all of the other inanimate Subjects, and yet the English-speaking community still attributes a higher level of agency to Machines than to other Inanimates. This surely has to do with the central role of Machines in our everyday life. In essence, Machines are tools, but they are capable of performances that the average human is often not capable of themselves. This is different from a hammer, for instance, which still requires a skillful user to be effective. Nevertheless, despite their capabilities and functions, Machines have as little sentience, motivation, and intention as any other tool and should in principle not be eligible for constructions that require these traits from their first participant. However, with today's technological advancement, these lines seem to become more and more blurry.

At this point, each register will be discussed individually, and the most noticeable Theme characteristics will be pointed out. Of the four registers, the one that deviates most from the norm for almost every Theme measure is INSTR. It has the lowest numbers of Theme elements and multiple Themes in the first experiential element hypothesis and

the second lowest numbers in the Subject hypothesis. This is due to the low number of textual and interpersonal Themes in the register. At the same time, it has the highest number of marked Themes at 23.7%, all of which are circumstance Themes. The low number of textual Themes is particularly surprising since, as these could organize a manual by outlining steps in which order the products had to be set up. However, as it turns out, most instruction manuals in the corpus do not offer instructions on how to build or install anything but rather list the functions and possible applications of the product in no chronological or causal order. Textual Themes were therefore not common.

INSTR has the highest relative frequency of material processes at 62.2%, which represents the highest percentage of any process type across the registers. Conversely, INSTR also has the lowest numbers of relational, mental, and behavioral processes and the second lowest number of verbal and existential processes. In terms of marked circumstance Themes, INSTR deviates most from the overall average. 66% of all circumstance Themes found in INSTR are Condition Themes, which is more than 5 times as high as for any other register. This is by far the highest relative frequency of any circumstance Theme in a register. Purpose, Means, and Behalf Themes are also relatively high compared to the other registers. Apart from these four, INSTR scores lowest for all other circumstance Theme types. Time Themes, in particular, have a surprisingly low relative frequency at 5.5% compared to the overall average of 26.7%. Despite being the register with the highest number of circumstance Themes overall, INSTR only has the highest circumstance TPot for one of the eleven analyzed circumstance types, namely circumstances of Purpose, while seven of the eleven circumstances have the lowest TPot in the register. INSTR is the only register that includes any Machine Subject Themes and also contributes most of the Process Subject Themes to the overall statistic. Machine Subjects are mostly computers and computer programs, which are actually more likely to be paired with a verb that requires sentience than with a verb that does not (see example (160); Machine Subject in bold).

(160) **Firefox** can handle many types of files.

[E2G_INSTR_006]

This unique thematic structure in INSTR is caused by a unique tenor and mode of discourse. Instruction manuals are primarily about physical events either carried out by the user or by the product. This explains the high number of material processes. At the same

time, the inner workings of the user are mostly irrelevant in this register, which is why mental and behavioral processes score so low. A higher number of relational processes which describe attributes of the product was expected. However, these attributes were mostly realized in a more active sense by describing the possible applications through material processes. What is noteworthy is that the Subject is oftentimes not in thematic focus in instruction manuals. Instead of departing from the product or the user, a large number of clauses open up with a circumstance specifying a condition that needs to be met or the purpose that an action has. Thematically, instruction manuals are not so much focused on providing information but rather on drawing the attention of the reader to the circumstances in which this information is relevant. Circumstances of Condition are by far the most common circumstance Theme, typically realized as an *if*-clause specifying what happens in which situation (see example (161); circumstance Themes in bold). Means and Purpose Themes are also common to explain to the reader how or for what purpose the product can be used. The high number of Behalf Themes is surprising. In most cases, they are used in clauses like (162), where they specify for which operating system the information applies.

(161) ***If you load the tray***, the printer begins printing.

[E2G_INSTR_001]

(162) ***For Macintosh***, types and sizes are all grouped together under the Paper pop-up menu.

[E2G_INSTR_001]

Clauses in INSTR are more formulaic compared to the other three registers. The same thematic and lexico-grammatical patterns are used repeatedly, and some sentences are even repeated almost word-for-word in the same text. Instruction manuals appear to be very specific in describing the way a product works and thus often repeat the same information for every possible setting or application even if it seems obvious to the reader.

The high number of Machine Subject Themes is hardly surprising since most instruction manuals are designed to describe a machine like a computer or other electronic devices. For the same reason, concrete Inanimates are also common. They mostly refer to objects that are generated by computer programs, such as menus, buttons, and icons. Nonconcrete Inanimates also frequently refer to settings and functions more generally.

Process Subject Themes are often used as an alternative to conditional clauses. Instead of specifying what happens if the user performs a certain action, that action is made the Subject and its effect is specified by the rest of the clause (see example (163); Process Subject in bold). Many of these Process Subjects are used in initiating material processes, where the Process Subject allows the user to do something.

(163) ***Saving a file onto your hard drive*** lets you view the page when you aren't connected to the Internet.

[E2G_INSTR_006]

Regarding the tenor of discourse, the reader is often directly addressed in INSTR. The communicative role of the speaker often switches from information provider to demander of goods and services. These demands are usually suggestions by the product designers, talking about what the user can, must or should do if they want to use the product in a certain way. Modal verb constructions were not analyzed quantitatively in this study, but from my impression, they are a lot more common in this register compared to the other three. As an alternative to modal verb constructions, imperatives are also frequently used in INSTR for the same purpose. The frequent switch between active and passive constructions in material processes allows the speaker to draw focus away from what the user has to do to the product itself. Example (164) shows one of the many cases of passives material processes in INSTR.

(164) *Full Screen view is often used for presentations, sometimes with automatic page advancement and transitions.*

[E2G_INSTR_003]

The register with the second most atypical thematic structure is TOU. TOU is similar to INSTR in terms of its low number of non-experiential Themes and high number of marked Themes. Just like with INSTR, a higher number of textual themes in this register was expected, used as a thematic resource to sequence events. Instead, the register employs mostly Place but also Time Themes to establish spatial or temporal sequences. TOU is the only register of the four that has a relatively high number of Complement Themes at 2.4% in the first experiential element hypothesis. Apart from one case, these are all circumstantial Attribute Complements in attributive relational processes. TOU has by far the

highest number of relational processes and existential processes but also the lowest number of material processes. Regarding the other three process types, TOU scores similarly low to INSTR. Place Themes are particularly common at 32.3%, which is the third highest relative frequency out of any register. The frequent use of Place Themes mirrors Neumann's (2003) results on Theme in English tourist guides as well as Ghadessy's (1995) results on guidebooks. Other relatively frequent circumstance Themes are Guise and Reason Themes. All other circumstance types have an average or below average occurrence rate in TOU. In terms of circumstance TPot, TOU does not deviate from the average in too many cases. Circumstances of Comitative and Reason are more likely to be the point of departure in their respective clauses compared to the other registers, but otherwise circumstance TPot is mostly consistent with the overall scores. TOU includes by far the highest number of Place Subject Themes, which was to be expected. Organization Subjects are also fairly common and usually refer to event companies and tourism agencies.

In TOU, the most common experiences shared in the field of discourse are the attributes of touristic places. This is mostly accomplished through relational and existential processes. Such existential processes are often accompanied by a Place Theme, so that the thematic focus is rather on the place itself while its feature is subordinated (see example (165); Place Theme in bold). In relational processes, the feature can be highlighted thematically by making it the Carrier/Token and by having the place follow as a circumstantial Attribute. This order is again reversed in clauses like (166), where the circumstantial Attribute (in bold) is moved to the Theme position. These constructions make up the majority of Complement Themes in this register and overall.

(165) **Here in the South West of England**, there are many opportunities to be active whilst holiday.

[E2G_TOU_006]

(166) **In the centre of the city** stands the 1900 year old Cardiff Castle.

[E2G_TOU_008]

Alternatively, the place can also be made the Subject in a relational process and the feature is realized as the Attribute, oftentimes in possessive relational processes (see example (167); Place Subject in bold). This explains the high number of Place Subject Themes in the register. In this way, the speaker can alternate the thematic focus between place

and feature using a variety of process types and participant sequences. Place Subject Themes are also often used as part of a non-sentient construction. Here, they are presented as sentient doers, offering great attractions or welcoming visitors.

(167) **The town** has a notable Elizabethan parish church [...].

[E2G_TOU_001]

The other typical event in TOU involves the actions that can be carried out by the reader. This experience is realized lexico-grammatically as material processes. These constructions are also often accompanied by circumstances of Place. Circumstances of Place are used thematically to guide the reader through a sequence of spaces, similar to how a tour guide would also lead the participants through different places. Alternatively, TOU also makes use of Reason Themes to focus on why the attractions are interesting or unique (see example (168); Reason Theme in bold). Guise Themes play a similar role since they are used as additional and almost peripheral information about the place realized as the Subject of the clause (see example (169); Guise Theme in bold). Time Themes are also common, though a lot less common than in registers like FICTION and SPEECH. Apart from these, the field of discourse does not cover many other types of circumstantial events.

(168) **Because they are so numerous and capable** many of the choirs travel regularly overseas [...].

[E2G_TOU_007]

(169) **As a peninsula surrounded by the Atlantic Ocean on the north coast and the English Channel on the south coast**, the South West region is ideal for all kinds of watersports.

[E2G_TOU_006]

The social distance between speaker and hearer is rather low, which is demonstrated by the use of direct addresses. And yet, most clauses do not involve the reader at all but rather reference the touristic attractions. In this sense, the reader is less involved in this register than in INSTR. The speaker in TOU also switches between information provider

and demander of goods and services but does so less frequently compared to INSTR. Judging from these lexico-grammatical patterns, tourism leaflets in English come across less as guided tours and more as information brochures.

The thematic structure in FICTION is noteworthy in some respects, but overall, it comes closer to the average for most Theme-related measures than the two previous registers. FICTION has the highest average number of Theme elements due to the frequent occurrence of non-experiential Themes, in particular textual Themes. The vast majority of textual Themes are conjunctions positioned at the beginning of the sentence. *And* and *but* are by far the most frequent, making up almost half of all textual Themes in FICTION. It also has the highest relative frequency of Subject Themes, making it the least marked register in this respect. FICTION has the most even distribution of material processes and relational processes with the latter being slightly more common. FICTION also has the most even spread of process types overall, as mental, verbal, and behavioral processes are all relatively common in the register.

Surprisingly, FICTION does not contain as many circumstance Themes as the other registers. Time Themes were most common in FICTION, and Place Themes occurred relatively frequently as well. But other than that, there are no circumstance types that stand out as being characteristic of the register. Nonetheless, it includes a variety of different circumstance Themes, whose frequencies correspond to the average in almost all cases apart from Means and Purpose. Circumstance TPot in FICTION is close to the average for all of the circumstances that also have a high TPot overall like Condition, Time, and Guise. Where FICTION deviates most is with circumstances that have an overall low TPot. Here, the thematic potentials are either much higher in FICTION, as is the case with Frequency and Quality, or the already low average TPot is even lower, for example for Reason and Purpose. So, while the general frequencies of circumstance Themes, with the exception of Time Themes, are not really unique in FICTION, the likelihood with which these circumstances are made the Theme of their clause is quite characteristic of the register.

Inanimate Subjects are not very common in FICTION. The few Animal Subjects all come from this register, but they make up only a small number of Subjects and most of them come from one text, which happens to involve a circus. I would therefore not consider Animal Subjects as a generally common Subject type in FICTION. Other inanimate Subjects were also quite rare and even the generally common concrete and nonconcrete Inanimate Subject Themes have the lowest relative frequency in FICTION.

The field of discourse in fiction is arguably the most heterogeneous out of all of the registers. While the products in INSTR and the places in TOU change, the kinds of experiences shared in these registers are fairly repetitive. The experiences shared in FICTION, however, depend heavily on the kind of story the author decides to write. For example, one of the stories in the corpus is about the spiritual experiences of a member of an indigenous tribe, while another outlines the experiences of a child in a hospital. It is thus hardly surprising that the experiential, interpersonal, and textual patterns vary heavily between individual texts. Nonetheless, on a very general level, most of the fictional texts portray the story of a main character, who is in a situation of some kind of struggle, which they then have to deal with. This is typically paired with a detailed description of the characters and their surroundings, to enhance the reader's imagination and immersion. These two basic types of experiences are realized experientially as material and relational processes, which explains their very even distribution. Plus, since most of the texts are about human protagonists and their emotions and interactions, FICTION also demonstrates a diverse use of mental, behavioral, and verbal processes.

The same diverse impression continues for circumstance Themes. Unlike the two registers discussed previously, FICTION is not dominated by a single circumstance Theme but rather includes a wide range of relatively frequent circumstance Theme types. So, in a sense, the diverse thematic make-up of FICTION mirrors the diverse field of discourse. The high overall number as well as the high thematic potential of Quality Themes (see example (170); Quality Theme in bold) are also a testament to the focus on human behavior in FICTION.

(170) **Abruptly** *Broadface fell like a tree beside Greenstockings [...].*

[E2G_FICTION_005]

What is consistent in all ten texts in FICTION is that the stories represent chronologically ordered sequences of events. However, unlike in other registers, these events can include large temporal gaps. These are the reasons for the high number of textual Themes and Time Themes. Textual Themes like *and* and *but* are typically employed to order events in quick succession, while circumstances of Time re-orient the reader in the temporal space. These shifts in time are often made the point of departure in FICTION because they help to direct the reader through the unfolding of the story. This temporal focus also carries

over to other time-related circumstances like Duration, which has the highest relative frequency in FICTION, and Frequency, which has the highest TPot in FICTION.

The tenor of discourse can also be quite diverse in this register. From a linguistic point of view, there are generally two different speakers in the texts: the narrator and the characters. The narrator can be a character themselves but does not have to be. There are also two groups of hearers: the reader and other characters. So, while the narrator primarily communicates to themselves and to the reader, the characters only communicate with each other, either through direct or indirect speech. As a consequence of this diverse set of discourse participants, the speech functions are also diverse. The relationship between narrator and reader is very distant, which is illustrated by the fact that the reader is very rarely addressed directly.⁸² The communicative role of the narrator is almost exclusively provider of information. For most texts, the style of the narration is rather informal, which is supported by the high number of interpersonal Themes. In most clauses with interpersonal Themes, like example (171), these comment Adjuncts are used thematically by the narrator to influence the interpretation of the reader.

(171) *Perhaps this story is becoming like the wolf's track often is, it goes farther ahead into where it will happen, on and on, until it leads into beyond, and only then can it circle back to us again.*

[E2G_FICTION_005]

On the other hand, the relationships between the characters in the stories are as diverse as the stories themselves and the communicative roles of speaking characters can cover all of the four basic speech functions. This is also why the number of interrogatives is highest in FICTION, which was also found by Neumann (2014: 163). The fact that many of the analyzed clauses are supposed to represent spoken conversation is another reason for the high number of textual and interpersonal Themes in the register.

SPEECH is the register that has the most average Theme distribution for most of the basic Theme measures. It comes closest to the overall numbers in terms of average Theme number, frequency of multiple Themes, number of marked Themes, number of textual Themes, and number of interpersonal Themes. Like in FICTION, material and relational processes are also fairly evenly distributed with the former being slightly more common.

⁸² This claim can only be made for the ten texts included in this register. Other relationships between narrator, characters and readers are of course also possible in fictional texts.

Also as in FICTION, mental and verbal processes have an above average occurrence rate in SPEECH. That being said, the numbers of existential and behavioral processes are the lowest in this register.

SPEECH also has the least unique distribution of circumstance Themes. Again, like FICTION, SPEECH has a very balanced distribution of circumstance Themes without any noticeably standing out. Time Themes are the most common type in this register but they are also only slightly more frequent than the overall average. The only two types of circumstance Themes that are characteristic of SPEECH are circumstances of Means and Concession. However, their relative frequencies are only higher by small margins. Other than that, the remaining circumstance types are very close to the overall average. SPEECH is also the register with the highest number of *Other* circumstances, which is further testament to its diverse circumstantial space. In terms of circumstance TPot, SPEECH stands out more as it has above average numbers for Duration and Means, whereas Frequency circumstances have a TPot of zero despite the fact that Frequency circumstances are most common in this register.

The one Theme-related measure where SPEECH stands out the most is Subject Themes. Not only does SPEECH have the second highest number of inanimate Subjects, but it also has a varied distribution of middle animate and inanimate Subjects, ranging from Organization to concrete Inanimates to Process Subjects. Organization Subjects mostly come from SPEECH and they generally refer to different governments in the form of country names or to administrative agencies (see example (172); Organization Subject in bold). Also very common are nonconcrete Inanimate Subject Themes (see example (173); nonconcrete Inanimate Subject in bold), which refer to a variety of different entities like policies, decisions, and freedom.

(172) ***The ECB** runs monetary policy, and EU constraints (deficit ceiling) limit the government's ability to stimulate the economy, if necessary.*

[E2G_TOU_002]

(173) *With the No Child Left Behind Act, **education reform** grows up.*

[E2G_SPEECH_010]

What is most noteworthy about SPEECH is how similar it is to FICTION in terms of its thematic patterns, even though field, tenor, and mode of discourse are arguably very different. Nevertheless, while the experience shared in each register differs on a more detailed level, on a general level there are also a lot of parallels. Like in FICTION, there are two main types of experiences that are shared in SPEECH: the state of a country, government or policy, and the plans and actions of politicians and other political leaders. These experiences are mostly realized using relational processes and material processes. As more peripheral experience types, the speakers also address their own feelings towards a matter, realized as mental processes, and share conversations they had with other people, in the form of verbal processes. Such events are remarkably similar to the events covered in FICTION both in terms of experiential meaning and frequency.

In terms of circumstances, SPEECH also primarily uses Time Themes to organize events chronologically. However, unlike FICTION, SPEECH does not arrange a series of events in a temporal order but rather only differentiates between past, present, and future. These three temporal spaces are oftentimes separated with the help of Time Themes (see example (174); Time Themes in bold). Means Themes are primarily used to outline which measures either have already been taken or must be taken in order to achieve a certain goal (see example (175); Means Theme in bold). Concession Themes are an instrument to admit shortcomings in the past, which are then immediately followed by other achievements that relativize these shortcomings (see example (176); Concession Theme in bold).

(174) **Before this program**, only a few options existed for reducing pollutants released from coal, and almost all were expensive. [...] **Today**, because of the clean coal investment, 75 percent of U.S. coal-fired power plants now use, or are installing, low-cost, low-polluting burners to reduce smog-forming nitrogen oxides.

[E2G_SPEECH_002]

(175) **Working with Germany and other countries**, we will seek to advance further on this development agenda during the G-8 Summit in June in Canada [...].

[E2G_SPEECH_005]

(176) **Although our work has not finished**, terrorists now find it much harder to move money and assets around the world.

[E2G_SPEECH_014]

The many thematic similarities between FICTION and SPEECH are striking. The two registers are different in many ways in terms of both content and structure. However, on a more general level, a political speech is very much like a piece of literature, which describes the setting of the events to involve and win over the hearer and which outlines the actions and emotions of the protagonists and antagonists. Similar to fiction, the story that is told in political speeches needs to be convincing and easy to imagine in order to grab the attention of the hearer.

That being said, there are of course also some clear differences between the two registers, which can be detected in the thematic make-up. The relationship between speaker and hearer is quite particular in SPEECH. The speaker basically has two audiences. One is the audience that is present while the speaker is holding their speech, typically cabinet members. These members are present in time and space and can be interacted with during the speech. This is reflected in the high number of interpersonal Themes which include the largest proportion of Vocatives out of all four registers. The other audience group is the general public. Even though there is considerable physical and social distance, the speaker often addresses the general public, typically by using the first-person plural reference. 19% of Subjects in SPEECH are realized as *we* compared to the overall average of 7.2%.⁸³

One other unique feature of SPEECH compared to the other three registers is the mode of discourse. So far, mode of discourse has not been addressed specifically as all four registers are written registers. Political speeches are unique in the sense that they are written texts meant to be spoken. Hence, while they do not include some of the typical features of spoken language such as false starts, they do show traces of the spoken medium, for example in the form of a high number of conjunctions like textual Themes. Vocatives as interpersonal Themes are also a feature of the on-line, face-to-face nature of speeches. This is again similar to FICTION, which also includes parts which are meant to be spoken discourse inside the story.

⁸³ *We* can of course also refer to the political party or the government which the speaker is a part of, so that not all cases of *we* can be considered an address of the audience.

9 Contrastive analysis of Theme in English and German

9.1 Results

In this chapter, the Theme structures of English and German original texts will be compared. The goal of this chapter is to work out the contrastive similarities and differences of Theme in these two languages and assess the relevance of the different Theme hypotheses for contrastive comparisons. In total, three hypotheses for each language were considered. Two hypotheses are based on the same formal criteria in English and German, namely the first experiential element hypothesis and the Subject hypothesis. The Forefield hypothesis for German and the first element hypothesis for English are very close in terms of the extent of the Theme since the Forefield position in German generally only allows a single element. However, in practice, there are many exceptions to this rule, so they cannot be considered formally identical.

As has been shown in the previous chapters, only two Theme aspects are heavily dependent on the Theme hypothesis, namely multiple Themes and marked Themes. For these two measures, four combinations of Theme hypotheses will be compared: Forefield hypothesis in German and first element hypothesis in English, first experiential element hypothesis for both languages, Subject hypothesis for both languages, and lastly the Forefield hypothesis for German and the first experiential element hypothesis for English. The first three comparisons are obvious as they are most equivalent in terms of formal criteria. The final comparison uses two very different formal Theme criteria. However, these are the two most popular Theme hypothesis for each respective language and were also used in the contrastive and translation analysis in Freiwald (2016). Accordingly, this comparison is not only relevant for the state of the art, but it also helps to evaluate the impact of different formal criteria for Theme in a contrastive analysis. All other Theme-related measures are based on the hypotheses that were already used in the intralingual analyses (see Chapters 7 and 8). Tables and figure will mostly be based on the overall results. However, interesting register differences will be commented on the running text. Tables and Figures comparing individual registers are included in the Appendix.

9.1.1 Multiple Themes and Theme Markedness

9.1.1.1 Forefield hypothesis vs. first element hypothesis

	GO		EO	
Total Clauses	4901		4496	
Total Themes	5276		4496	
Single Themes	4544	92.7%	4496	100%
Multiple Themes	357	7.3%	0	0%
Average # of Themes	1.08		1	
Experiential Themes	4475		3955	
Subject Themes	2842	63.5%	3142	79.4%
Circumstance Themes	1255	28.0%	731	18.5%
Complement Themes	364	8.1%	42	1.1%
Predicator Themes	13	0.3%	4	0.1%
Textual Themes	528	10.0%	411	9.1%
Interpersonal Themes	234	4.4%	130	2.9%
Cleft	39	0.7%	36	0.9%

Table 16 Basic Theme distribution in GO (Forefield) and EO (1st element)

Table 16 shows an overview of the basic Theme distribution between GO, based on the Forefield hypothesis and EO, based on the first element hypothesis. As pointed out above, the Forefield hypothesis and the first element hypothesis are formally consistent at first sight since the Forefield in German is typically occupied by only a single element. Some elements are, however, exempt from this rule in German, for example textual Themes in the form of conjunctions or interpersonal Themes like Vocatives. Occasionally, even two or more experiential Themes can share the Forefield space. This explains why the average number of Theme elements in German is higher than one despite the strong Finite-second constraint. Since the number of Theme elements is restricted to one by definition in the first element hypothesis, this numeric difference between the two languages is inevitable and, unsurprisingly, statistically significant ($\chi^2 = 13.218$, $df = 1$, $p\text{-value} < 0.001$).

Generally, more clauses have been analyzed in German than in English. This is true overall and also per register with the sole exception being INSTR, where more clauses in English originals were considered. The same criteria were used for English and German to decide which clauses are analyzed, namely the first T-unit in a clause complex. This difference in number can thus only be explained by either a difference in the number of

clause complexes overall or a difference in interrogative and imperative clauses. The fact that fewer clauses in German instruction manuals were considered is due to the higher number of imperatives in that register.

One contrastive difference is immediately apparent and that is the difference in Theme markedness. At 36.5%, the relative frequency of marked Themes in German is considerably higher compared to the 20.6% in English.⁸⁴ Circumstance Themes are more common in GO than in EO, but the most striking difference is represented by Complement Themes, which are more than seven times more common in German than in English. Predicator Themes are very rare in both languages but again more frequent in GO. The differences in Theme markedness are statistically significant ($\chi^2 = 28.483$, $df = 1$, $p\text{-value} = 9.453e-08$).

The number of non-experiential Themes is quite similar in both languages with German having slightly higher numbers for both textual and interpersonal Themes. However, this similarity can be purely the result of the choice in Theme hypothesis. Since some textual and interpersonal Themes can share the Forefield position with an experiential Theme, the number of non-experiential Themes is diluted compared to the first element hypothesis in English, where an opening textual or interpersonal Theme is never accompanied by an additional Theme element. Clefts and pseudo-clefts are slightly more common in English. The 50.3% higher number of cleft constructions in English, reported by Erdmann (1990b: 71-72), cannot be observed in this data set though.

Considering different registers, some interesting thematic similarities and differences emerge. In general, most of the same trends can be observed in both languages. For example, the registers that have the highest number of textual Themes in German, SPEECH (12.3%) and FICTION (11.6%), also have the highest numbers in English (9.4% and 17.3%), albeit in reversed order. The same holds true for interpersonal Themes, which are also most common in SPEECH and FICTION and least common in INSTR and TOU in both languages.

INSTR has by far the highest relative frequency of circumstance Themes in both languages at 40.0% in GO and 23.3% in EO. However, while clauses in SPEECH are commonly contextualized by a circumstantial Theme in EO, it is the register with the smallest number of circumstance Themes in GO. In fact, the relative frequency of circumstance Themes

⁸⁴ These numbers represent the sum of all non-Subject experiential Themes.

in SPEECH is almost as high in English (19.6%) as it is in German (21.6%), despite the fact that the gap is very wide for all other registers.

Complement Themes in English really only appear in TOU and FICTION, while INSTR and SPEECH are virtually without any Complement Themes. In German, Complement Themes are most frequent in TOU as well, but they are a fairly common Theme choice in all four registers. Complement Themes are thus less register-dependent in German than in English.

9.1.1.2 Forefield hypothesis vs. first experiential element hypothesis

	GO		EO	
Total Clauses	4901		4496	
Total Themes	5276		5097	
Single Themes	4544	92.7%	3955	88.0%
Multiple Themes	357	7.3%	541	12.0%
Average # of Themes	1.08		1.13	
Experiential Themes	4475		4496	
Subject Themes	2842	63.5%	3568	79.4%
Circumstance Themes	1255	28.0%	829	18.4%
Complement Themes	364	8.1%	46	1.0%
Predicator Themes	13	0.3%	4	0.1%
Textual Themes	528	10.0%	431	8.5%
Interpersonal Themes	234	4.4%	170	3.3%
Cleft	39	0.7%	49	1.1%

Table 17 Basic Theme distribution in GO (Forefield) and EO (1st exp. element)

Table 17 contrasts the thematic spaces of Theme in German based on the Forefield hypothesis and Theme in English based on the first experiential element hypothesis. Now that more than one Theme element of the EO can be included, the number of multiple Themes naturally increases and, as predicted, surpasses the numbers in GO. Textual and interpersonal Themes can never be the sole Theme element in the first experiential element hypothesis, which is why the number of multiple Themes is equal to the number of clauses that have at least one non-experiential Theme. In the German Forefield hypothesis, most textual and interpersonal Themes take up the entire thematic space, so whether multiple Forefield elements are allowed depends on the type of non-experiential Theme.

However, while the predicted effect does emerge, it is still noteworthy how relatively small the difference between both languages is. Despite the Finite-second constraint in German, the relative frequency of multiple Themes in English is only 4.7% higher. Word order differences and the position of the Finite in particular are the most significant formal contrasts in comparisons between German and English Themes and yet they are apparently inconsequential for a high number of clauses. Nevertheless, the differences that were found are statistically significant ($\chi^2 = 6.936$, $df = 1$, $p\text{-value} = 0.0084$).

The differences in Theme number vary between registers. Generally, both languages are consistent regarding registers that are more likely to feature multiple Themes. The register with the greatest discrepancy is INSTR, where multiple Themes are more than ten times as likely in EO than in German (5.0% to 0.4%). The register with the smallest difference in both absolute and relative numbers is SPEECH. Here the percentages of multiple Themes are almost even at 11.3% in GO and 12.7% in EO.

Other than multiple Themes, not much else has changed from the previous Theme comparison. The distribution of marked Themes is almost identical and remains statistically significant ($\chi^2 = 16.57$, $df = 1$, $p\text{-value} = 4.689e-05$). The differences between the respective registers have also not changed substantially. The gap in textual Themes has widened further since now single textual Themes in English do not make up the entirety of the Theme. The number of interpersonal Themes has increased in English, thus closing the gap between EO and GO.

9.1.1.3 First experiential element hypothesis in both languages

	GO		EO	
Total Clauses	4901		4496	
Total Themes	6083		5097	
Single Themes	4204	85.8%	3955	88.0%
Multiple Themes	697	14.2%	541	12.0%
Average # of Themes	1.24		1.13	
Avg. # of Themes w/o Finite	1.16		1.13	
Experiential Themes	4863		4496	
Subject Themes	3200	65.8%	3568	79.4%
Circumstance Themes	1259	25.9%	829	18.4%
Complement Themes	387	8.0%	46	1.0%
Predicator Themes	16	0.3%	4	0.1%
Textual Themes	527	8.7%	431	8.5%
Interpersonal Themes	655	10.8%	170	3.3%
Finite Themes	416	63.5%	0	0%
Modal Adjuncts	239	36.5%	170	100%
Cleft	38	0.6%	49	1.1%

Table 18 Basic Theme distribution in GO (1st exp. element) and EO (1st exp. element)

Table 18 represents the Theme based on the first experiential element hypothesis in both languages. What was earlier the reason for the lower number of Theme elements in GO is now also the reason for the higher number in this hypothesis. Due to the Finite-second constraint in German, many experiential Themes are placed after the finite verb in the Midfield. As a consequence, many German clauses, like example (177), gain an additional interpersonal Themes in the form of finite verbal operators (Finite Theme in bold), which cause the higher numbers of Theme elements overall.

(177) *Ausserdem | **kann** | die Elektronik Schaden nehmen, wenn der Rasierer vorzeitig herausgenommen wird und noch nicht trocken ist.*
*'moreover | **can** | the electronics damage take, if the razor prematurely taken.out is and not yet dry is.'*
 [G2E_INSTR_005]

This explains why the general likelihood of a multiple Theme is much more similar between English and German at 12.0% and 14.2% respectively, but the average number of Theme elements is so divergent. If a Theme is multiple in German, it contains on average more Theme elements compared to a multiple Theme in English because of that extra

finite verbal operator. Based on the Theme criteria of the first experiential element hypothesis, this difference in interpersonal elements is a thematic difference between the two languages and has to be presented as such. However, it is caused purely by obligatory clause structure rules and does not represent a deliberate choice by the speakers. Therefore, it should not be over-interpreted as a thematic effect. For this reason, average Theme numbers excluding finite verbal operators have also been added to Table 18 to make the comparison of other non-experiential Themes more meaningful. Contrastive differences in Theme number are statistically significant but only if the sum includes Finite Themes ($\chi^2 = 18.643$, $df = 1$, $p\text{-value} = 1.577e-05$). If Theme numbers are compared excluding Finite Themes, the difference becomes statistically non-significant ($\chi^2 = 1.0478$, $df = 1$, $p\text{-value} = 0.306$).

If Finite Themes are disregarded, the number of Theme elements is remarkably similar in GO and EO. GO still features slightly more Theme elements due to the slightly higher numbers of textual Themes and interpersonal Themes which are not Finites. However, the assumption that the German Theme is considerably more textually focused can be regarded as false, at least from the perspective of this Theme hypothesis comparison.

The average number of Themes and the frequency of multiple Themes is consistently higher in the German registers. Multiple Themes in INSTR previously differed most substantially, with average Theme numbers being much higher in EO. When using the same formal Theme criteria in both languages, this effect is now inverted with German featuring almost twice as many cases of multiple Themes as English (9.5% to 5.0%). The only register that has a higher probability of multiple Themes in English is FICTION. This can be attributed to the higher number of non-experiential Themes in this register.

Due to positioning rules of the German Midfield, the relative frequency of marked Themes has decreased, most notably for circumstance Themes. If a Theme is opened by a textual or interpersonal Theme and the first experiential element is positioned in the Midfield in post-Finite position, that element is most likely the Subject. Circumstance Themes at the onset of the Midfield are rather rare (see Section 7.1.1). Therefore, the differences in marked Themes are now less distinct compared to the previous Theme hypotheses. Nevertheless, there is still a noticeable gap between non-Subject Themes in English and German and this difference remains statistically significant ($\chi^2 = 18.71$, $df = 1$, $p\text{-value} = 1.521e-05$). This small decrease has affected the different registers relatively equally.

9.1.1.4 Subject hypothesis in both languages

	GO		EO	
Total Clauses	4901		4496	
Total Themes	9625		6067	
Single Themes	2693	54.9%	3178	70.7%
Multiple Themes	2208	45.1%	1317	29.3%
Average # of Themes	1.96		1.35	
Avg. # of Themes w/o Finite	1.56		1.35	
Experiential Themes	6729		5413	
Subject Themes	4855	72.2%	4437	82.0%
Circumstance Themes	1370	20.4%	865	16.0%
Complement Themes	425	6.3%	46	0.8%
Predicator Themes	33	0.5%	8	0.1%
Textual Themes	566	5.9%	434	7.2%
Interpersonal Themes	2330	24.2%	220	3.6%
Finite Themes	2001	85.9%	0	0%
Modal Adjuncts	329	14.1%	220	100%
Cleft	57	1.1%	57	1.1%

Table 19 Basic Theme distribution in GO (Subject) and EO (Subject)

Table 19 shows thematic differences between EO and GO based on the Subject hypothesis. For the Subject hypothesis, anything that is positioned before the Subject is automatically part of the Theme. Based on this Theme hypothesis, both languages can have multiple experiential elements in the Theme. In both languages, the most common pattern of multiple experiential Themes involves a circumstance Theme (followed by the Finite in German) followed by the Subject Theme (see example (178); circumstance Themes underlined, Subject Theme in bold).

(178)

EO: In northern California | **hiking along rocky cliffs and visiting secluded beaches to see seals, sea otters or whales are more popular than swimming.**

[E2G_TOU_011]

GO: In Baden-Württemberg | ist | **die Tradition der närrischen fünften Jahreszeit sehr verbreitet.**

'in Baden-Württemberg | is | **the tradition of the foolish fifth season very widespread.**'

[G2E_TOU_005]

Compared to the previous Theme hypothesis, the numbers of multiple Themes and Theme elements increase in both languages, to 29.3% in EO and 45.1% in GO. This increase is evidently much stronger in German than in English and GO now has a higher probability of multiple Themes and higher Theme numbers both with and without considering Finite Themes. This effect is consistent in all registers. The reason for this notable increase in Theme elements in GO comes from the relatively weak notion of thematic markedness in German (Steiner and Teich 2004: 169). In the previous Theme hypotheses, the majority of circumstance Themes constituted simple Themes since the head of the Midfield is rarely occupied by a circumstance. However, in the Subject hypothesis all circumstance Themes are part of multiple Themes now since the Subject needs to be included in the Theme. Thus, the increase of potential Theme candidates affects German much more than English. In the Subject hypothesis, the differences in Theme number between English and German are significant if Finite Themes are included in the count ($\chi^2 = 141.71$, $df = 1$, $p\text{-value} = 2.2e-16$) and also if they are not ($\chi^2 = 48.543$, $df = 1$, $p\text{-value} = 3.232e-12$).

The previously high differences in marked circumstance Themes decrease noticeably in the Subject hypothesis. In SPEECH, English now even surpasses German in the relative frequency of circumstance Themes with 17.1% to 16.5%. The overall decrease is, however, a product of the choice in Theme hypothesis. In the Subject hypothesis, the relative frequency of circumstance Themes inevitably decreases since now every clause that opens up with a circumstantial element also has to have an additional Subject Theme. This dilutes the percentages of circumstance Themes in both languages, but the effect is stronger for German. Most English clauses open with the Subject, so the amount of additional Subject Themes between the first experiential element hypothesis and the Subject hypothesis is relatively small. In German, 36.5% of clauses do not have the Subject as the first experiential element, which is why the number of Subject Themes increased much more in GO than in EO. If the relative frequencies of circumstance Themes were calculated per clause, the difference between EO and GO would still be 28% to 19.2%, which is an even higher difference than for the first experiential element hypothesis. The reason why the divide increases even further is that German is not only more likely to have a clause introduced by a circumstance Theme, but it is also more common to have multiple circumstantial elements in the Theme in German (see example (179); circumstance Themes

in bold). Despite this technicality, the differences in marked Themes remain significant in the Subject hypothesis ($\chi^2 = 15.784$, $df = 1$, $p\text{-value} = 7.102e-05$).

(179) *Wird Ihre Waage über das Maximalgewicht belastet*, | *erscheinen* | *im oberen Teil des Displays* | *fünf waagrechte Striche*.
'*is your scale over the maximum.weight burdened*, | *appear* | *in.the upper part of.the display* | *five horizontal lines*.'

[G2E_INSTR_001]

In summary, the average number of elements per Theme in both languages is – unsurprisingly – highly dependent on the choice of Theme hypothesis. If Theme in German is restricted to the Forefield position, its numbers are smaller compared to English in the First Experiential Element and the Subject hypothesis. If the same formal criteria are applied to both languages, the Theme element count in German outnumbers that of English. In the first experiential element hypothesis, this difference is almost entirely caused by the additional Finite Themes, brought about by the German Finite-second constraint. In the Subject hypothesis, this effect is enhanced by the higher probability of circumstance and Complement Themes in GO. What the data has shown, however, is that textual and interpersonal Themes that are not Finites have a similarly high probability per clause in both languages. It has also been shown that irrespective of Theme hypothesis, marked Themes are always more common in GO than in EO.

9.1.2 Participant Themes and process types

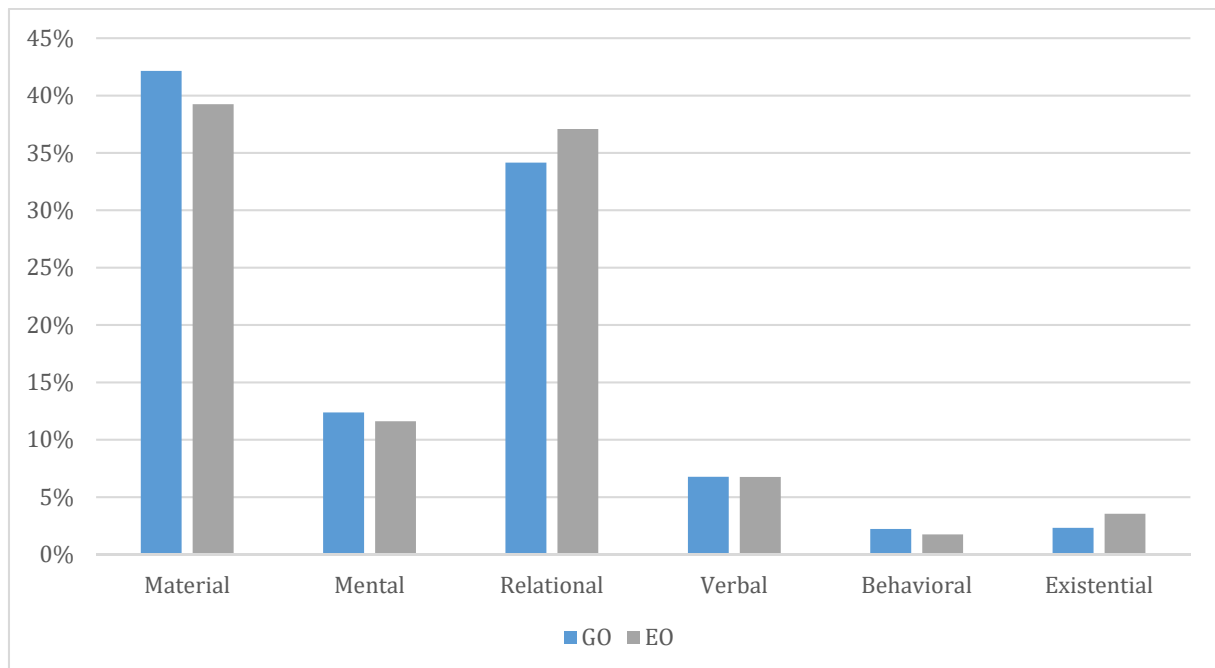


Figure 13 Process type distribution in GO and EO

Figure 13 shows the distribution of process types. Looking at the relative frequencies overall, it is quite remarkable how they are almost identical across all six process types. Relational and existential processes are slightly more common in EO, while GO has higher numbers for the other four types. Nevertheless, these differences are quite marginal and also not statistically significant ($\chi^2 = 0.1326$, $df = 1$, $p\text{-value} = 0.7157$). Thus, it is fair to say that there is no difference in the use of process types in English and German overall based on the registers analyzed in this study.

This picture changes slightly in the analysis of individual registers. SPEECH in EO is more geared towards material processes (EO: 40.5%, GO: 33.8%), whereas SPEECH in GO features more mental processes (EO: 13.0%, GO: 18.1%). The register that is least similar between EO and GO is TOU. Here, EO contains a lot more relational processes (EO: 54.5%, GO: 38.3%) as well as existential processes (EO: 8.7%, GO: 3.1%) while GO includes considerably more material processes (EO: 28.0%, GO: 45.5%). In FICTION, there are higher numbers of mental and existential processes in EO, opposed to a higher frequency of material processes in GO.

Apart from these differences, the process type distribution of the registers is rather similar in English and German. The distribution in INSTR in particular is almost identical across all six process types. And while FICTION does display the previously mentioned

contrasts, its overall distribution is relatively similar. While Theme number and Theme markedness demonstrated clear contrastive differences, the experiential domain of the participant Theme only varies slightly in some registers and is very similar in EO and GO on the whole.

In the majority of cases the experiential Theme represents a conflation of Subject and the first participant role of the process (see Figure 2 in Section 7.1.2 and Figure 8 in Section 8.1.2). However, the use of a different participant role in Theme position as well as the frequency of Complement Themes highly depends on the process type in both English and German. The Theme structure in behavioral processes is identical in both languages since the entirety of behavioral processes have the first participant as the Subject Theme. Existential processes in English only feature Subject Themes in the form of the non-referential Subject *there* in combination with the process. In German, existential processes can also be introduced by an Existent Complement Theme (see example (180); Complement Theme in bold). Such a construction is outright ungrammatical in English, which explains this contrastive difference.

(180) *So viel Wandern gibt es nirgends.*
'so much hiking is there nowhere.'
[G2E_TOU_017]

All other process types allow Complement Themes in both languages. As was already pointed out in the previous section, Complement Themes are considerably more marked in English than in German irrespective of Theme hypothesis. This difference in markedness is not associated with any particular process type but applies to all process types overall. Apart from existential processes, the process types with the highest number of Complement Themes in German are relational (10.8%) and verbal processes (11.5%). Incidentally, these are also the two process types that include at least some cases of Complement Themes in English (2.7% for relational and 0.9% for verbal).

Surprisingly, the number of Subject Themes that are not the first participant is also higher in GO. The contrast is not as distinct as for Complement Themes; and yet, the effect is consistent for the four most frequent process types. The most common lexico-grammatical pattern that has a Subject Theme which is not the first participant of a process is the passive in both languages. Grammatically determined languages like English are said to use passive constructions more frequently (Firbas 1964a) to allow different sequences

of experiential meaning without going against the basic word order. However, a contrastive difference regarding voice could not be demonstrated. Teich (2003: 181-182) found significant differences regarding the use of passives and passive-alternatives between English and German. However, if these two categories are summarized in her data, their numbers are surprisingly similar (225 in GO and 229 in EO). Given that no difference was made in this thesis between passives and passive-alternatives, Teich's results match the results here.

9.1.3 Circumstances

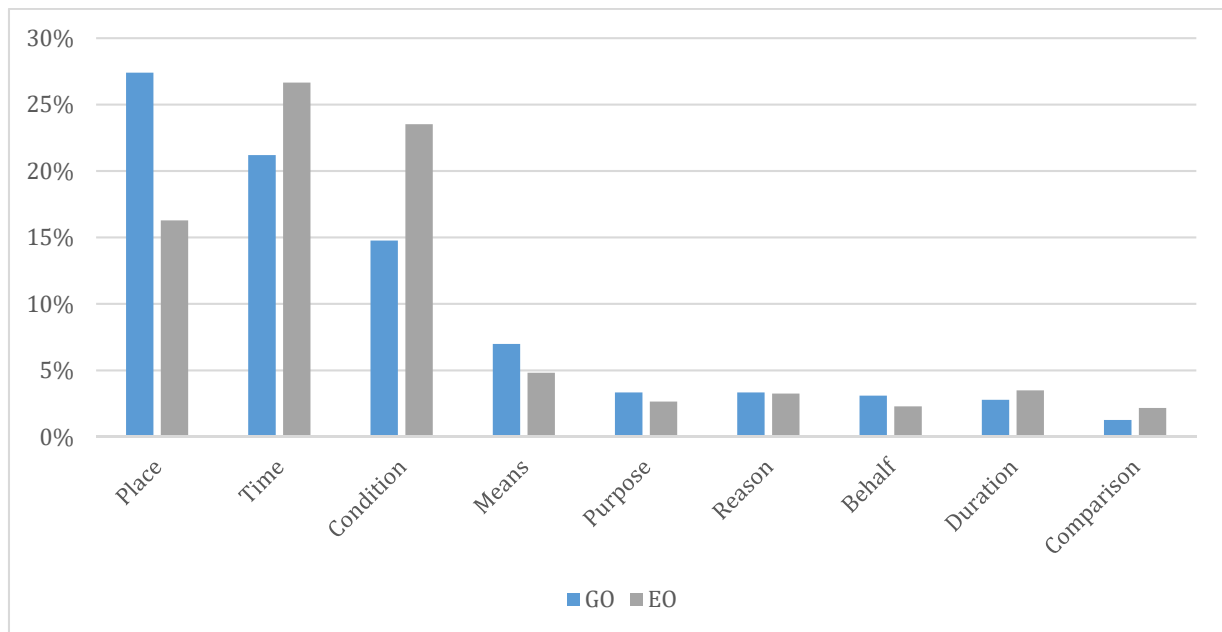


Figure 14 Circumstance Theme distribution in GO and EO

Figure 14 illustrates the general distribution of circumstance Themes in English and German overall. This figure does not reflect the frequencies of circumstance Themes per clause. Instead, it is a division of all circumstance Themes in each language according to their types. So, while there is a higher number of Duration Themes in GO, both in absolute terms and also per clause, their relative frequency is lower since they make up a lower percentage of circumstance Themes overall in German. Figure 14 only takes those circumstance Theme types into account that fulfilled the frequency criteria in both languages (see Sections 7.1.3 and 8.1.3).

Generally, the circumstance type distribution in the Theme is similar in some aspects while different in others. The three most common types are Place, Time, and Condition, which make up the vast majority of circumstance Themes in both languages. The frequency distribution for the remaining circumstances is fairly comparable and, while their order is not perfectly identical in English and German, differences are marginal. Of these remaining circumstances, Means is the only type that differs noticeably between the two languages, being more representative of German circumstance Themes. This is particularly true for INSTR, where they make up 11.8% of circumstances in GO as opposed to only 4.5% in EO. Other than that, the distributions of medium and low frequency circumstance Themes are comparable.

There is, however, a great divide in the frequency distribution of Place, Time, and Condition. Place Themes make up the largest portion of German circumstance Themes followed by Time and Condition. In English, Place Themes only come in third and are heavily outnumbered by Time and Condition. This discrepancy is consistent for the overall distribution as well as per register. Overall, the contrastive differences in circumstance Themes between English and German are significant ($\chi^2 = 4.6027$, $df = 1$, $p\text{-value} = 0.03192$).

In registers where Time and Condition Themes are particularly common, the difference between English and German also becomes larger. Hence, even though INSTR includes the highest percentage of Condition Themes in both EO and GO, the difference between the languages is also the highest in this register. The same holds true for circumstances of Time in FICTION, which has the highest proportion of Time Themes in both languages, and yet the divide between English and German is most evident here as well. Place Themes on the other hand are consistently more common in German, even in TOU, where they are the most common in English.

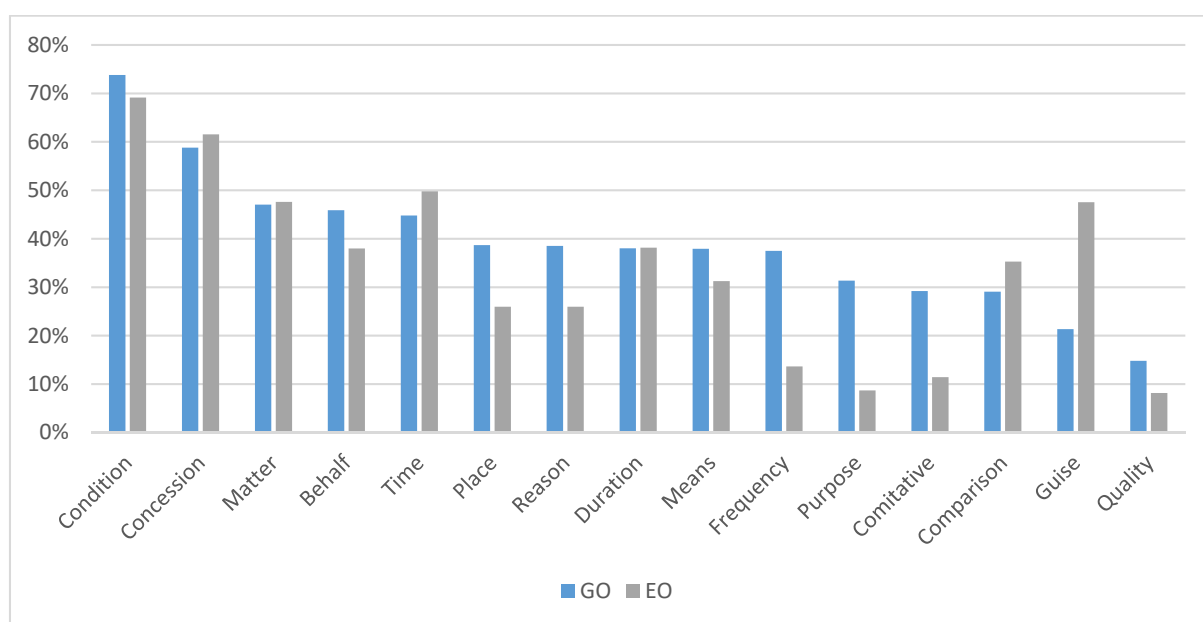


Figure 15 Thematic potential of circumstances in GO and EO

Figure 15 illustrates the differences in thematic potential of the most common circumstance types. Again, only those circumstance types that met the frequency criteria in both languages were included, which eliminated Additives.

As was shown in 9.1.1, marked circumstance Themes are more frequent in GO than in EO by a significant margin. Based on this finding, it is surprising how little the thematic

potential of circumstances differs between the two languages. Taking all types into account, circumstances in German have a likelihood of 38.9% to be made the Theme of their clause compared to the overall thematic potential of 33.9% in English. Accordingly, while the number of circumstance Themes is much higher in German, the general potential of circumstances to become Theme is very comparable.

This similar distribution between English and German is especially apparent for circumstances of high thematic potential like Condition, Concession, and Matter. In the case of Concession and Time, EO even demonstrates higher thematic potentials overall. Where English and German differ the most is in those circumstances that only have a medium or low level of thematic potential. In GO, differences in TPot are evident but the slope is less steep than in EO. German circumstances maintain a fairly even TPot in the medium and lower fields. English, on the other hand, demonstrates drastic differences in TPot across circumstance types. As a consequence, some of the circumstances with an average TPot in German have a considerably lower TPot in English, as is the case with Place, Reason, Quality, Frequency, Purpose, and Comitative. This difference is particularly noteworthy for the latter three types, where German circumstances are roughly three times as likely to be made the Theme of their clauses.

While it is fair to say that German circumstances have a generally higher thematic potential than English circumstances, some of the circumstance types in English also outperform those in German. The TPots of Concession, Time, and Duration are higher in English but the differences are quite marginal. Circumstances of Guise (see example (181) ; circumstance Themes in bold) and, to a smaller degree, circumstances of Comparison (see example (182)) have a substantially higher TPot in English, however. This means that if an English and a German clause both include a circumstance of Guise or Comparison, there is a higher likelihood in English that these circumstances are used thematically.

(181) ***As the largest producer of wine in the USA***, California has established a reputation for fine wines.

[E2G_TOU_011]

(182) ***Like a television set***, the electronics and picture tube give off a certain amount of heat, which is normal and not a cause for concern.

[E2G_INSTR_002]

9.1.4 Subject animacy and sentience

In this section, contrastive differences between English and German Subject Themes are worked out regarding Subject animacy and sentience. Given the more restricted word order in English, the use of non-sentient constructions is one of the strategies in English to allow a more varied semantic sequence (Hawkins 1986: 67). In German, a certain meaning does not have to be mapped on to the Subject to come early and orient the clause in its context. Instead, non-sentient referents can be part of circumstances and Complements and can still be the Theme of the clause. Therefore, German is not as dependent on non-sentient constructions as a resource for the order of semantic meaning. That being said, non-sentient constructions are not ungrammatical in German.

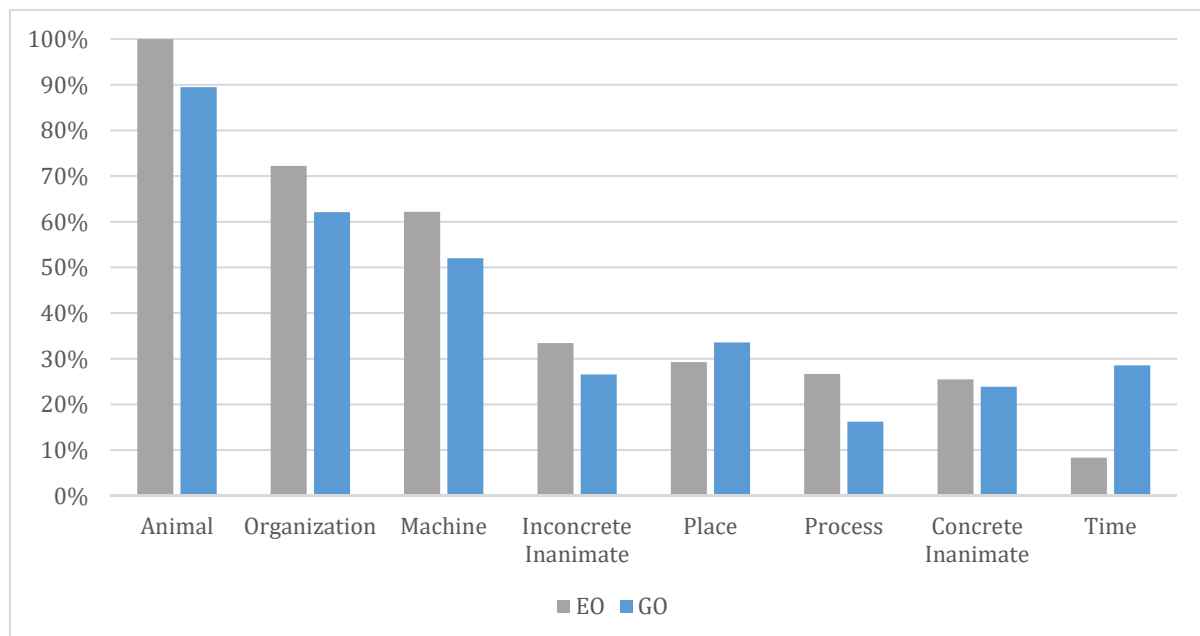


Figure 16 Subject Themes as part of non-sentient constructions in EO and GO

The frequency distribution of non-sentient constructions can be seen in Figure 16. Vehicle Subjects were taken out of the Figure as they were too infrequent in EO. Generally, inanimate Subjects are more likely to be part of non-sentient constructions in English than in German. Overall, 41.8% of English inanimate Subjects were paired with a verb that requires sentience as opposed to 34.7% in German.⁸⁵ This difference is not negligible

⁸⁵ At this point, I would like to emphasize again that these are not the general frequencies of inanimate Subjects and sentient verbs. Relational processes that do not require a sentient first participant are not considered in these numbers. The general probability of non-agentive constructions is thus much lower in both languages and only reflects those cases that were analyzed here.

but also not substantial. Despite the semantic restriction placed on Subjects in German (Hawkins 1986: 57-58), the frequency of non-agentivity is still comparable. Contrary to the original assumption, the distribution of sentient and non-sentient construction is not significantly different between EO and GO ($\chi^2 = 1.979$, $df = 1$, $p\text{-value} = 0.1595$).

Subjects of middle animacy, which includes Machine Subjects, as was argued before (see Section 5.11), are most frequently part of non-sentient constructions in both languages. In all of these categories, EO produced higher frequencies of non-sentient constructions than GO, and the differences are fairly consistently at around 10%. The results for inanimate Subjects are a lot more varied. Apart from Time Subjects, English has a rather even probability for each of the inanimate Subjects. The numbers for German, however, vary a lot more, and their order also does not correspond to that of English. Non-concrete Inanimates and Process Subjects are more likely to be paired with sentient verbs in English, as are concrete Inanimates, albeit by a small margin. Place Subject Themes are more likely to be involved in a non-sentient construction in German, but again the difference is rather small at 4.3%. Time Subjects are the only Subject type where GO has a considerably higher probability of being part of a non-sentient construction than EO at 28.6% to 8.3%.

9.1.5 Subject identifiability

Subject Theme identifiability in German was discussed in Section 7.1.5, to work out the relationship between Subject positioning and its identifiability. Since the positioning of Subjects in English declaratives is not as varied, this Theme analysis was not relevant for the intralingual perspective. From a contrastive perspective, German and English may display differences in Subject Theme identifiability, which will be elaborated here.

		GO		EO	
Subjects in total		4901		4496	
	Identifiable	3438	80.2%	3247	80.3%
	Non-Identifiable	850	19.8%	797	19.7%
Other		613		452	
Pre-verbal Subjects		2900		-	
	Identifiable	2109	81.2%	-	-
	Non-Identifiable	488	18.8%	-	-
Post-verbal Subjects		2001		-	
	Identifiable	1329	78.6%	-	-
	Non-Identifiable	362	21.4%	-	-
Subjects in immediate post-verbal position		1649		-	
	Identifiable	1195	82.1%	-	-
	Non-Identifiable	260	17.9%	-	-

Table 20 Subject Theme identifiability in GO and EO

Table 20 shows the comparison of Subject Theme identifiability for English and German. The German results are separated into pre- and post-verbal Subjects since the position of the Subject relative to the finite verb was assumed to be highly relevant.⁸⁶ While it is also possible for the Subject in English to follow the verb instead of preceding it, such word orders are very rare and not directly linked to the identifiability or non-identifiability of the Subject, which is why they were not considered specifically in this analysis. The most common construction involving such a Subject-Finite inversion are circumstantial attributive constructions with a fronted Complement Theme (see example (183)).

- (183) *To the west lies the Exmoor National Park, bordered by the sea, and home to the wild red deer and the settings of the tragic tale of Lorna Doone.*
[E2G_TOU_006]

⁸⁶ As it turns out, Subject identifiability was much less dependent on position than previously assumed. See Sections 7.1.5 and 7.2 for a discussion.

With a difference of only 0.1% in favor of Subject identifiability in EO, it is evident that there is no meaningful difference in Subject identifiability in English and German overall. Unsurprisingly, this difference is also statistically non-significant ($\chi^2 = 1.6596$, $df = 1$, $p\text{-value} = 0.1977$). Post-verbal Subjects generally have a slightly lower probability of being identifiable in GO, while post-verbal Subjects immediately after the finite verb have a slightly higher one. Other than that, overall numbers are almost identical.

Variation does exist between the registers. With a difference of 0.3%, Subject identifiability in FICTION can be considered equal in English and German. In INSTR and SPEECH, German Subjects are slightly more identifiable than English ones (INSTR: 80.8% to 77.6%; SPEECH: 84.6% to 78.6%). However, the largest difference can be found in TOU, where English Subjects are more likely identifiable (71.2%) than German Subjects (63.3%). On the whole, Subject identifiability seems more varied in German registers because the deviations from the global average are more pronounced. In contrast, Subject identifiability in English is more even across all registers.

9.2 Discussion

In this section, the differences in Theme between German and English originals will be discussed. For both languages, three Theme hypotheses were contrasted in terms of multiple Themes and marked Themes, which is where the choice in hypothesis is most crucial. In total, four combinations of hypotheses were compared to work out all thematic differences from as many different perspectives as possible. This multitude of comparisons is meant to shed light on the true differences in point of departure without being too driven by methodological choices.

The frequencies for Theme number and multiple Themes are most dependent on the formal criteria of Theme. Naturally, the more elements that are potentially considered part of the Theme, the higher the count of Theme elements will be on average. A formal difference between English and German enhances this effect: The finite verb is typically in the second position in German declaratives. This is not a deliberate choice by speakers but a grammatical rule, which would result in ungrammaticality if ignored. Depending on the choice of Theme hypothesis in German, the Finite is either part or not part of the Theme as an interpersonal Theme element. Thus, it comes as no surprise that the average Theme numbers vary considerably between English and German depending on which two hypotheses were contrasted. Higher average Theme numbers in German were to be expected for the comparison of the Finite and the first element hypothesis as well as for the comparisons of the first experiential element hypothesis and the Subject hypothesis. At the same time, higher Theme numbers in English were predictable for the comparison of the Forefield hypothesis in German and the first experiential element hypothesis in English. This finding is consistent with the results in Freiwald (2016) and Niemietz, Neumann, and Freiwald (2017), as well as contrastive studies including English and other Germanic languages, for example Hasselgård (2004).

What is revealing is the comparison of Theme numbers if the difference in Finite positioning and its effect on Theme is disregarded. In this case, the average number of Themes become fairly even, especially in the first experiential element hypothesis at 1.16 in German and 1.13 in English; a small difference that is not statistically significant. This shows that the thematic space in English and German is almost identical in the first experiential element hypothesis regarding multiple Themes if the Finite Theme is not counted.

To be clear, this is not to say that the Finite in German is not a true Theme element if positioned before the first experiential element hypothesis. I do believe that the German

Finite Theme is as much a thematic element as a finite verbal operator in interrogative clauses, as an interpersonal point of departure to signal the mood. Thus, the differences found between EO and GO are meaningful. However, it is important to stress that these differences are almost entirely caused by this contrastive difference in syntax and not due to a difference in textual or (other) interpersonal Themes.

In terms of non-experiential Themes, GO and EO have been shown to be fairly similar. At 0.2%, the difference of textual Themes between German and English is negligible. Interpersonal Themes as modal Adjuncts are more common in German at 4.9% compared to the 3.3% in English in the first experiential element hypothesis. This difference is largely consistent across all registers. The divide is widest in SPEECH, where modal Adjuncts and Vocatives in German outnumber those in English at 7.4% to 3.6%. Only TOU contains more modal Adjunct Themes in English than in German.

It is not immediately clear what the reason for this difference is. In SPEECH, the probability of Vocatives addressing the audience at the beginning of the clause are almost ten times as high in German, which seems to be a contrastive register difference. But, even disregarding these Vocatives, modal Adjuncts as interpersonal Themes are still more common in German in three out of the four registers. Hence, this does not seem to be a result from the choice of registers alone but a consistent difference between English and German. It appears to be generally more common in German for the speaker to use personal evaluation as a point of departure.

The chi-square tests have conclusively shown that non-Subject Themes are more common in German than in English as has previously been described by various authors (for example Neumann 2014; Freiwald 2016; Niemietz, Neumann and Freiwald 2017). Surprisingly, Teich (2003: 185-186) reports more marked Themes in English than in German in her contrastive analysis, which I can only explain as an atypical distribution in her data, given that her finding was also not statistically significant. The higher number of non-Subject Themes in German are consistent across all four Theme comparisons and across all registers. Marked Complement Themes in particular have shown to be a clear contrastive difference, being almost eight times as common in GO across all Theme hypotheses. Complement Themes can be argued to be marked in both languages since they are the least likely candidate for experiential Themes (apart from Predicator Themes). However, while Complement Themes are relatively infrequent in GO as well, they are not a thematic

abnormality. In EO, on the other hand, Complement Themes are extremely rare and must be heavily motivated by context.

Circumstance Themes are also more common in GO than in EO, yet the divide is not as wide. Unlike Complement Themes, the difference in circumstance Themes between English and German is also heavily dependent on the choice of Theme hypothesis. It ranges from 9.6% in the comparison between Forefield hypothesis and first experiential element hypothesis to only 4.4% in the comparison of the Subject hypothesis. This substantial decrease in the Subject hypothesis can be largely attributed to the heavier impact of the additional Subject Themes for the German Theme statistics. Nonetheless, it is indisputable that non-Subject Themes are more common in German than in English, which is the same interpretation that Freiwald (2016) arrives at.

Interestingly enough, the relative frequencies of Complement Themes are not impacted as heavily from this increase of Subject Themes, staying at almost the same difference of eight to one between GO and EO. The reason for this lighter change in ratio is that Complement Themes can be paired with a circumstance Theme and even with a second Complement Theme in the same German Theme. In fact, there are 38 German clauses overall that have a Complement Theme in combination with another marked Theme. For instance, example (184) is opened up by a Complement Theme in the Forefield, followed by the Finite, followed by another Complement Theme at the head of the Midfield before the Subject is introduced (Complement Themes in bold). In English, there is not a single case in all four registers where a Complement Theme is accompanied by another marked Theme.

(184) **Einige Kopfschmerzen** | *bereiteten* | **Kaiser Franz Josef** | *nicht nur Regierungsgeschäfte [...].*
'**some headaches** | *caused* | **emperor Franz Josef** | *not only government.affairs [...].'*
[G2E_TOU_021]

The comparison of process types and participant Themes is very revealing. The overall results of process type distribution are almost even and the small differences that were found are not statistically significant. This finding is in line with Teich (2003: 182-183), who also did not find significant differences in the use of process types between English and German. On the whole, the general types of shared experiences and their frequencies seem to not differ dramatically between the two languages. There is some variation in

individual registers, which will be commented on below. Still, the general experiential space is very comparable. This goes to show that despite linguistic and cultural differences between these two speech communities, human experiences are still rather similar.

There are some notable differences regarding participant Themes, however. Given the lack of word order freedom in English, an increase of passive constructions was expected, resulting in an increase of second participant Subject Themes, such as Goal Themes or Verbiage Themes. This is not the case. In fact, there are more cases of Subject Themes that are not first participants in German. This is consistent for all of the four most common process types.⁸⁷

Most relational processes cannot be passivized, so the majority of second participant Subject Themes for this process type are a result of Value Subjects in identifying relational processes. Phenomena in mental processes do not always lend themselves well to passivization, so the small number of Phenomenon Subjects is not surprising. Also, in German the Phenomenon can assume the role of the Subject and the Senser the role of the Complement without passivization (Steiner and Teich 2004: 153-154). This lexico-grammatical pattern is not available in English, which is reflected in the results. Nevertheless, there is no obvious explanation for why passives in material and verbal processes are not more common in English.

The distribution of circumstance Theme types is similar in some respects yet also different in others. In general, EO and GO agree on which types of circumstances are frequent and infrequent in Theme position. In particular, the group of medium and low frequency circumstances is rather homogenous. The three circumstances with the highest frequencies, Time, Place, and Condition, are also the same yet their relative order differs substantially. Place is the most frequent German circumstance Theme, while Time and Condition are more common types of English circumstance Themes. It should be noted that the high number of circumstances of Conditions is primarily caused by INSTR. In all other registers, Place and Time make up the two most common circumstance Theme types in both languages (though not always in the same order). Nevertheless, even in registers where Condition Themes are not as common, they consistently make up a larger percentage of circumstance Themes in English.

⁸⁷ In both languages, behavioral processes only have first participant Subject Themes and if an existential process begins with a Subject, it is always the empty *there/es* Subject.

The reason for this discrepancy, especially in the case of circumstances of Place, lies partly in the thematic potential. Looking at the circumstances with the highest thematic potential, Condition, Concession, Matter, and Time, the numbers for English and German are almost identical, even higher for English in the case of Time and Concession. Where English and German differ the most is with circumstance types that are less commonly used as Theme. So apparently, there is agreement between the two languages on which circumstances lend themselves naturally to be used as Theme. As was pointed out by Greenberg (1990), Biber et al. (1999) and others, Conditions have a natural disposition to be placed at the beginning of a clause. If the speaker wants to set up a hypothetical situation, this contrast is most easily accomplished by opening the clause with the hypothetical condition so that the Rheme can be interpreted accordingly. A similar argument can be made for Concessions. Concessions often set up a contrast between the circumstantial information and the main experiential material of the clause. Again, it is easier for the speaker to understand and interpret this contrast if the circumstance of Concession is used as the point of departure (see example (185); Concession Themes in bold).

(185)

EO: ***While geothermal heat pumps are typically more expensive to install***, their greater efficiency means the investment may be recouped in three to ten years.

[E2G_SPEECH_002]

GO: ***Trotz vieler, oftmals bitterer Auseinandersetzungen*** haben wir Brücken gebaut, Gegensätze überwunden, Lösungen gefunden.
'***despite many, oftentimes fierce disputes*** have we bridges built, differences overcome, solutions found.'

[G2E_SPEECH_005]

Circumstances of Matter introduce the topic of what a clause or a stretch of text is about. What the clause is about has been rejected as a suitable definition for Theme as it is primarily used to describe the linguistic concept of Topic (see Section 5.1.2). Nonetheless, the Topic is a natural choice as the point of departure as it signals to the hearer in which context the Rheme needs to be interpreted, which is why circumstances of Matter are often fronted if they are included in a clause (see example (186); Matter Themes in bold).

(186)

EO: *As for Jeremy Mohonk, the third principal player in the mortal drama about to unfold, he didn't pay rent, [...].*

[E2G_FICTION_004]

GO: *Was den öffentlichen Dienst angeht, so sind wir derzeit in der Schlichtung. 'what the civil service regards there are we currently in the conciliation.'*

[G2E_TOU_005]

The last circumstance type with an above average thematic potential in English is Time. Unlike the other three types, there is no inherent reason for why the fronted position is a more natural position in the clause. However, circumstances of Time are placed in the Theme in EO exactly half of the time, making it equally likely for Time to be thematic or rhematic. In many registers, texts are temporally ordered, as is the case in FICTION register as well as SPEECH in this sample. It appears that there need not be a very specific context which motivates the choice of a temporal point of departure but rather that the temporal transition from one clause to the next is a common resource of structuring texts, at least in some registers. A similar assumption can be made about circumstances of Place, but as the data shows, a spatial contextualization is a much less probable way of structuring discourse.

What the German-English comparison shows clearly is that English thematizes circumstances that have a natural disposition of being thematic like Condition or Matter and circumstances that require little contextual motivation like Time just as much as German if not more. The more rigid word order of English does not get in the way of thematizing circumstances if there are good reasons to do so. Where English and German do differ are those circumstances which require more contextual motivation: Place, Reason, Means, Frequency, Purpose, Comitative and Quality Themes. Such types also have an average or below-average thematic potential in German. Yet, the relative difference between circumstances of high thematic potential and those of medium and low thematic potential are much larger in English, which accounts for this difference in circumstance Theme frequency.

There are two circumstance types that have a below average thematic potential in GO and a higher potential in EO, namely Comparison and Guise. Circumstances of Comparison are not particularly frequent in either language, so that the difference in thematic potential between 35.5% in English and 29.1% in German may simply be due to variance. The thematic potential of circumstances of Guise, on the other hand, is more than twice

as high in English with a fairly sizeable amount in both languages. Looking at examples of Guise Themes more closely, it appears that this difference is primarily caused by constructions known as detached predicatives (Biber et al. 1999: 136), which were annotated as circumstances of Guise for lack of a better alternative (see Section 6.3). These detached predicatives are usually nominal or adjectival groups which describe the Subject, but which are not formally connected to the Subject as pre- or post-modifiers. Detached predicatives also exist in German but are simply less common, which explains this difference in thematic potential for circumstances of Guise.

In Section 5.11, H_{2.1.3} and H_{2.2.4} were tentatively formulated, stating that those non-Subject Themes that differed noticeably in terms of TPot between EO and GO were strong predictors in translations into German and into English respectively. Now that the original subcorpora have been contrasted, the hypotheses can be phrased more accurately:

H_{2.1.3} Comparison and Guise Themes are significant predictors of Theme change in translations from English to German.

H_{2.2.4} Place, Reason, Means, Frequency, Purpose, Comitative, Quality and Complement Themes are significant predictors of Theme change in translations from German to English.

However, it must be noted that the average thematic potentials of circumstances in EO and GO are not much apart at all. Across all circumstance types, German has a thematic potential of 38.9%, whereas English circumstances have a thematic potential of 33.9% overall. This poses the question why the overall frequency of circumstance Themes is so much greater in GO if the average potential of a circumstance to become Theme is only 5% higher. The answer is that German writing is simply more circumstantial than English writing. The average number of circumstances per clause in German is 0.66 compared to the 0.54 in English. This absolute difference appears small but in relative terms it means that German has 22.2% more circumstances overall. This relative difference is particularly surprising in the case of a circumstance like Place, which is among the three most common types in English but is still used 36.5% less often than in English. Both Neumann (2003: 191-192) and Freiwald (2016: 51) also found a similar discrepancy regarding circumstances of Place in the register of tourist guides and popular scientific texts, respectively.

It is of course the speaker's choice to include or not include a circumstance in a clause. But if the speaker decides against including circumstances, the likelihood of a circumstance Theme is quite obviously zero. The more circumstances are used in a language in the first place, the more the probabilities of a circumstance Theme in that language increase. Accordingly, the average thematic potential of circumstances is almost the same between German and English, but German uses 22.2% more circumstantial Adjuncts and as a consequence more of these Adjuncts become Theme. Neumann (2003: 184) reports a very similar discrepancy between English and German tourist guides, where the German original corpus also included 15.8% more circumstances.

This is an unexpected result. It was assumed that the differences in word order freedom and positional flexibilities are the core reasons for the increased number of marked circumstance Themes in German. And the data on thematic potential clearly shows that these word order differences exist. However, the primary reason for this increase in marked Themes is a difference in circumstantiality. German experience, at least in the four registers analyzed here, is marked by more circumstantial Adjuncts, which has also left its mark on the Theme.

There is no obvious reason why English relies less on circumstantial information than German. One possible explanation might be that information like Place and Time take up grammatical functions other than Adjuncts because of the positional restrictions in English. The Subject is regarded as less semantically restricted in English for exactly this reason (Kast 2012: 148). However, Time Adjuncts already have a higher thematic potential in English than in German, so there seems to be little need for alternative strategies when it comes to temporal information. And in fact, Place Subjects are considerably more common in German. Therefore, it does not seem as if this circumstantial information is expressed in any different way in English, but that English simply does without them.

Regarding inanimate Subjects and agentivity and sentience, the general assumption that English makes use of more non-sentient constructions is not confirmed by the statistical analysis. In terms of descriptive statistics, higher numbers of non-sentient constructions can be found in EO: middle animate and inanimate Subject Themes are paired with agentive or sentient processes in 41.7% in English, compared to the 34.7% in German. However, this difference is not significant. This result contradicts Freiwald (2016: 80), who did find a significant contrastive difference in non-agentivity between German and

English. Additionally, differences in non-sentient constructions between English and German are not consistent across all types of inanimate Subject Themes. Contrary to my expectation, the difference is more pronounced with Subjects of middle animacy, Animals, Organization, and Machines, than with inanimate Subjects. In fact, Place and Time Subjects are more likely to be paired with a verb that requires sentience in GO than in EO. The sample size of Time Subjects is quite small in both languages, so this finding should not be over-interpreted. Nevertheless, Place Subject Themes are relatively common in both English and German and, while the relative difference of 3.4% is marginal, it still shows that Place Subjects are at the very least equally likely to be used as a sentient participant in German. The probabilities for concrete Inanimates are higher in English but, like Place Subjects, the difference between EO and GO is only very small.

What these results suggest is that the more concrete an inanimate Subject is, the more easily it can be conceptualized as an agentive or sentient first participant in German. In English, the concreteness of the inanimate participant seems to play a less important role, resulting in equally high probabilities for nonconcrete Inanimates and Process Subjects. These more abstract Subjects are the two types that have the smallest likelihood of being used in a non-sentient construction in GO. The nonconcrete Inanimates that are used as sentient participants in German primarily come from INSTR and TOU. In INSTR, they most often refer to commands (see example (187); nonconcrete Inanimate in bold), which are not machines themselves but extensions of machines, which is why it is not surprising that they have a higher likelihood of being presented as actively doing something.

(187) **Takt**: *Regelt die horizontale Pixelverteilung zur Optimierung der Bilddarstellung.*
'**tact**: *regulates the horizontal pixel.distribution to.the optimisation of.the image.representation.*'

[G2E_INSTR_002]

The distinction between mental verbs and agentive verbs is only truly relevant for Organization, where mental verbs were used frequently in both languages. Some cases of mental verb construction could be found for almost all Subject types, but they are quite infrequent overall. In practice, it makes little difference whether the annotations rely on sentience or agency since the results are, for the most part, comparable at least for the languages English and German. Since the differences are so small and sentience can be more

easily analyzed I agree with studies like Ackerman and Moore (2001), De Swart (2014), and García García, Primus, and Himmelmann (2018), which abandoned agency in favor of sentience all together.

In summary, it has been shown that non-sentient constructions are more common in EO than in GO in general, so the results are consistent with the state of the art (Kast 2012; Serbina 2015; Freiwald 2016). However, the relative difference found in the results is not as grand as might be expected given the discussed semantic restrictions on the German Subject (König and Gast 2009: 108-109). Besides, it is not consistent with all types of Subjects. As is the case with most of the contrastive Theme differences found so far, it is more a question of tendencies and probabilities rather than of binary categories.

There is no difference between English and German regarding Subject Theme identifiability. The distribution between identifiable and non-identifiable Subject Themes are virtually identical in the overall results. There are some differences within certain registers, but these are also relatively small. The results show that there is a correlation between identifiable or given information and Theme in both languages, which confirms the claim by Fries (1997: 233) and others that the system of INFORMATION parallels that of THEME.

Registers are always dependent on cultural norms given that they describe the different types of interactions that take place in a specific culture. Consequently, just because there are text types in two cultures that are referred to by similar terms, this does not mean that field, tenor, and mode of discourse are the same in both cultures. Nevertheless, there are no obvious cultural differences regarding the four analyzed registers, which is why deviations in thematic patterns were assumed to be due to differences in the language systems themselves rather than register differences. This assumption proved to be accurate for most Theme related measures.

FICTION has a high number of non-experiential Themes in both languages, resulting in an above-average number of Theme elements in all Theme hypotheses. Textual Themes are common in GO but even more so in EO. Most textual Themes are cohesive conjunctions used by the narrator to string events together, but they are also frequently used by the characters in direct speech. Comment Adjuncts as interpersonal Themes are used often, making FICTION the register with the highest number of interpersonal Themes in EO and the second highest in GO, followed by SPEECH. Compared to the other registers, marked circumstance Themes are less common in FICTION.

Process type and participant Theme distributions are very comparable in German and English fictional writing. The relative frequencies of verbal and behavioral processes are almost identical and the deviations in relational and existential processes are small. The only two process types where the numbers differ noticeably are material processes, which are more frequent in GO, and mental processes, which are more common in EO.

These results can be largely attributed to a similar field of discourse in both languages. On a very general level, fictional texts share the experience of their characters, both in terms of their actions and their inner workings. Additionally, characters and settings are often described in detail to allow the reader to visualize the events. Based on the results, the most common types of experiences shared in English and German fiction are the same, yet German writers focus a bit more on the events in the real world, while English writers spend more time on the emotions, thoughts, and desires of their characters, at least in this sample. There are not any fundamental differences in the way material and mental processes are integrated into the fictional writing. However, as was pointed out before, FICTION is likely the register where the field of discourse differs most for individual texts since the kinds of stories told can vary extraordinarily. Therefore, it is difficult to assess whether this contrastive difference is consistent in all of fiction or applies more to the ten analyzed texts respectively. Nevertheless, the data does suggest a more mentally-focused style of English authors in contrast to a more action-focused style of German writers.

Regarding circumstance Themes, FICTION in EO and GO is again very comparable, as almost all types demonstrate similar relative frequencies. Where FICTION differs most contrastively is in terms of the two most frequent circumstance types in this register, namely circumstances of Place and Time. Place Themes make up a higher percentage of circumstance Themes in German while Time Themes are more representative of English circumstance Themes.⁸⁸ Notably, as was pointed out earlier, this is not a unique effect in FICTION but a consistent effect across all registers that Time Themes in English outnumber Time Themes in German and vice versa in the case of Place Themes.

⁸⁸ At this point, it is important to reiterate that these relative frequencies are based on the distribution of all circumstance types. 44.0% of circumstance Themes in EO FICTION are Time Themes compared to the 27.7% in GO. However, that does not mean that the number of Time Themes is higher in English. It rather means that of all circumstance Themes in FICTION, Time makes up a larger percentage in English than in German.

The field of discourse is also similar in that the speakers in fictional stories can switch between the narrator communicating the events to the reader and the characters communicating with each other. Direct speech by the fictional characters is common in both languages and as a consequence, interrogative mood is most common in FICTION.

In summary, the few thematic differences that were found in FICTION are consistent with the contrastive differences overall. Only the differences in the process type distributions and participant Themes appear to be a contrastive register effect and are worth being investigated further.

The register of instruction manuals is the most similar register between EO and GO regarding thematic structure. In both languages, INSTR has the highest number of marked circumstance Themes and the lowest number of marked Complement Themes. There are few textual and interpersonal Themes in INSTR and consequently the average number of Theme elements is among the lowest in all Theme hypotheses apart from the Subject hypothesis in GO. In the Subject hypothesis, INSTR has the highest number of Theme elements in German because of the frequent circumstance Themes.

In terms of participant Themes and process types, INSTR could not be more similar in English and German. The process type distribution is almost identical across all six process types. Relational processes are slightly more common in GO while existential processes are used more often in EO. Oftentimes, both of these process types are used for the same purpose, namely, to describe a feature of the product, so it is not surprising that these effects cancel each other out. The field of discourse is dominated by the same three types of experience: The actions by the user and their effects, the potential performances by the product and the product features. The first two experiences primarily map onto material processes and the latter maps onto relational and existential processes.

In terms of circumstance Themes, it appears as if Condition Themes are a lot more common in English than in German, given that they make up 66.0% of circumstance Themes in English compared to only 42.7% in German. However, this difference is not due to a difference in the use of circumstances of Condition. In fact, in absolute numbers, German and English are almost completely even at 0.158 and 0.157 Condition Themes per clause. The reason why Conditions make up such a large amount of circumstance Themes is due to the fact that besides Conditions no other circumstance type is particularly common in the INSTR Theme in EO. Even the two generally common circumstance

types, Place and Time, have a very low occurrence rate in this register. In German instruction manuals, Condition Themes also make up the vast majority of circumstance Themes, but other Theme types are still relatively common. In particular Place, Time, and Means Themes demonstrate a much larger percentage of circumstance Themes in GO than in EO. INSTR is the only register in which Time Themes are more common in German than in English. Time Themes are used in both languages either to point out potential dangers while the user is doing something or to explain what happens after the user has done something. In English, there are simply fewer of them.

This discrepancy could be caused by a lower use of circumstances of Time overall in English instruction manuals. However, the thematic potential of Time is also the lowest in this register in English, whereas the thematic potential of Time is consistent with the average in German. This is also true for Place and Means. Irrespective of circumstance type, circumstance Themes are marked choices as point of departure in English. INSTR is already the register with the highest number of circumstance Themes precisely because of the high number of circumstances of Condition and their high thematic potential. In order to not enhance this marked thematic structure in INSTR even further, English speakers seem to try to avoid any further unnecessary marked Themes. This then reduces the thematic potential of all other circumstance types. This is not true for German, since circumstantial Adjuncts in Theme position are already a relatively unmarked choice. This is the reason for the more diverse picture of circumstance Themes in German instruction manuals.

INSTR is the most homogenous register in both languages. Lexico-grammatical patterns are repeated constantly, and even entire sentences are often re-used with only slight lexical changes. This homogenous nature of INSTR also reflects in the contrastive comparison. Field, tenor, and mode of discourse are remarkably similar and as a consequence the German and English Themes differ only slightly. Apart from the more varied choice of circumstance Themes in German, this register can be considered contrastively equal in terms of Theme structure.

The most different register is undoubtedly SPEECH. SPEECH contains the lowest number of circumstance Themes in GO but the highest number in EO. In fact, their relative frequencies are almost even at 19.4% and 20.1% in the Subject hypothesis despite the more rigid word order and the fewer circumstances overall in English. Textual Themes

are similarly frequent in both languages, a consequence of the more verbal mode of discourse of political speeches. But interpersonal Themes and in particular Vocatives are a lot more common in German.

This dissimilar picture continues with participant Themes, as SPEECH is among the contrastively more varied registers in terms of process type distribution. Material processes make up a considerably larger number of process types in EO, which is counter-balanced by a larger number of mental processes in GO. These processes are used to describe the same kinds of experiences in English and German. Material processes describe either actions by politicians, governments or government programs, and mental processes are used to express the feelings or thoughts of the speakers or the needs of a country. But apparently English politicians are more focused on material events while German politicians are more focused on mental aspects. Interestingly, this is the exact opposite effect compared to FICTION, where the German authors used more material and English authors more mental processes.

Considering some of the examples of mental clauses, it seems as if German politicians often use them as hedges to present future events as personal desires rather than necessities or obligations. At closer look, it turns out that the German speakers in SPEECH make very little use of modal verbs and instead use clauses like (188), in which mental processes serve similar functions. There is not a single use of the modal verb of obligation *müssen* (*must*) and only 19 cases of modals of recommendation like *sollen* (*should*) in GO. In contrast, SPEECH in EO contains 84 instances of *must*, *have to*, and *need to* (see example (189)) and 45 instances of *should*. In this sense, political speeches in English are more direct in their appeals to the audience. This finding supports Teich's (2003: 109) claim that German prefers to express modality with modal Adjuncts rather than auxiliaries.

(188) *Ich wünsche mir, dass die Politik die Kraft findet, ihre Zuständigkeiten in Bund, Ländern und Gemeinden klar zu trennen und zu ordnen.*

'I wish [refl-1sg] that the politics the strength finds, their responsibilities in confederacy, states and communities clearly to separate and to order.'

[G2E_SPEECH_005]

(189) *We must enlarge this Alliance to complete the task we started in 1948.*

[E2G_SPEECH_008]

Time again makes up a larger proportion of circumstance Themes in EO and circumstances of Place are more representative of the Theme in GO. This is a general contrastive difference and not a register effect. Other than that, the circumstance Theme distribution is similar. Behalf Themes are used more often in GO to specify whom the proposed actions benefit (see example (190); Behalf Theme in bold). In EO, they are virtually non-existent. Instead, Duration Themes are utilized frequently in English political speeches to explain for how long efforts have been made (see example (191); Duration Theme in bold).

(190) ***Gerade für Länder, die sich im Wiederaufbau nach einem Konflikt befinden**, ist eine mittel- bis langfristig konzipierte Unterstützung sehr wichtig.*
'***especially for countries, which themselves in.the reconstruction after a conflict are**, is a medium- to long-term conceived support very important.'*

[G2E_SPEECH_006]

(191) ***For more than half a century** Europe worked hard to make intra-European conflict no more than a memory.*

[E2G_SPEECH_012]

The tenor of discourse in political speeches is insofar the same in English and German in that the speaker has two groups of audiences, namely the attendees of the speech and the general public. What they differ in primarily is the frequency of address. In German political speeches, there are 46 Vocatives addressing the audience during the speech compared to the five Vocatives in English. When the speaker addresses the audience in English, it is always an address to a specific person or group of people (see example (192); Vocative in bold). In German, there are many general addresses in the form of *meine Damen und Herren* (*ladies and gentlemen*), which can refer to both the people present and the public. German political speeches are thus more involving and, in this sense, more interpersonally focused than in English.

(192) ***Mr. Chairman, Members of the Committee**, I am delighted to be here today to review with you the Administration's goals for U.S.-European relations [...].*

[E2G_SPEECH_006]

The thematic differences show that the style of political discourse in English and German is different. Political speeches in English are much more circumstantial than is typical for

English original writing generally. There is a strong focus in English on background information, time frames and purposes. SPEECH also has the most varied distribution of circumstance Themes out of all four registers. Additionally, speeches in English are less personal, as shown by the few direct addresses. In German, political discourse style is more characterized by personal desires of the speaker and fewer demands to the hearer.

TOU is another register that features a number of contrastive thematic differences. In terms of marked Themes in general, English and German tourism leaflets are similar in so far as circumstance Themes have a slightly above-average rate of occurrence and Complement Themes are by far the most numerous. The high number of Complement Themes in both languages are due to the many circumstantial attributive processes, which allow the Attribute of Place to be positioned before the Finite and the Subject. Textual Themes have a below-average occurrence rate in both languages but are still a lot more common in GO. Interpersonal Themes, on the other hand, are more characteristic of TOU in EO. The modal adjuncts *perhaps* and *in fact* are very common in the English Theme, either lessening or strengthening the force of the information.

Process types and participant Themes in TOU are the most different between English and German out of all four registers. English tourism leaflets are dominated by relational and existential processes that are primarily used to describe the features of a place or a sight. In German, there are considerably more material processes describing the activities the reader can enjoy at the place. Both of these types of experiences are communicated regularly in both languages, but their relative distributions are not equal. While the process type distribution of German and English tourist guides in Neumann (2003) does not match the distribution in this thesis perfectly, she also found a higher number of relational processes in the English corpus. Moreover, the number of relational processes in English tourist guides differed significantly from an English reference corpus, while those in German brochures did not (Neumann 2003: 166). This further strengthens the interpretation that English tourism texts are more relationally-focused.

The results for circumstance Themes look more similar. They show the regular discrepancies between Place and Time Themes. Also, Condition and Reason Themes are more frequent in English tourism leaflets than in German ones. Condition Themes in clauses like (193) are often used to narrow down for what kind of reader the information is relevant (Condition Theme in bold). Reason Themes serve the purpose of explaining to

the reader the reason for a place's qualities (see example (194); Reason Theme in bold). Other than that, circumstantial information is provided similarly in TOU in EO and GO.

(193) ***If you are a gourmet***, you will also be spoilt for choice [...].

[E2G_TOU_007]

(194) ***Because so much of the landscape in Wales is of such beauty***, a large proportion of it is protected.

[E2G_TOU_007]

Regarding tenor of discourse, audience involvement is more pronounced in German tourism leaflets. This effect is partially created through the high number of material processes, where the Actor is the reader themselves. The higher number of textual Themes in German also contributes to the appearance of a guided tour, where the reader is led from one place to the next. In fact, many material processes describing reader activities are written in present tense in GO, almost as if the reader is already on the journey while reading the leaflet. TOU in EO, on the other hand, appears more matter-of-fact. Although the register also includes direct addresses to the reader, the majority of clauses describes place features rather than reader activities. The style in German tourism leaflets is more that of a guide, as exemplified by the flowery language in clauses like (195). The style in English tourism leaflets, on the other hand appears to be more of an information brochure.

(195) *Dann taucht der Besucher in langgestreckte unterirdische Hallen mit aktueller Kunst und erreicht zum Schluß die schmucklosen Neubauetagen rund um den eindrucksvollen hohen Lichthof im Ungers-Bau.*

'then dives the visitor in long.stretched underground halls with contemporary art and reaches at.the end the unadorned new.storeys around the impressive high light.well in.the Ungers-building.'

[G2E_FICTION_006]

10 Analysis of Theme in English/German originals and translations

In the final two results chapters, the effects of Theme in English to German and German to English translations will be analyzed and assessed. In Chapter 10, a general analysis of Theme in all four sub-corpora, EO, GO, ET, and GT, will be presented and discussed. This includes a general comparison of Theme distributions in both translation directions. Chapter 11 is concerned with the inferential analysis of the effects of Theme in translations. This includes generalized linear mixed-effects models to work out the most common factors that lead to translation shifts of Theme structure in the two translation directions. Additionally, frequent translation procedures will be outlined and discussed here. Although Chapters 10 and 11 are clearly connected, they are presented separately given their respective lengths.

10.1 Results

This chapter consists of the general analysis of Theme structure in both English to German and German to English translations, followed by a discussion of these results. Both translation directions are presented together to allow a comparison of the translation effect and assess the influence of both source and target language. The presentation of the Theme results will again take into consideration different Theme hypotheses for multiple Themes and marked Themes. For all remaining Theme-related measures, the same Theme hypotheses as in Chapters 7 and 8 will be used. The first element hypothesis in English was excluded in both this chapter and the following chapter. While the very first element in the English clause is clearly thematically meaningful and its analysis in the previous chapters has produced interesting results both intralingually and contrastively, the results did not deviate from the other two hypotheses in a substantial way. Additionally, the first element hypothesis is not commonly used by the state of the art for English Theme analyses, which is why it was excluded from the following two chapters to increase the clarity of the already rather detailed translation analyses. Therefore, the next two chapters will be concerned with three comparisons in total: The Forefield hypothesis for German and the first experiential element hypothesis for English, the first experiential element hypothesis for both languages, and the Subject hypothesis for both languages.

Section 10.1 will only display the overall results and not individual register results. However, register-specific differences will be mentioned in the running text if relevant.

10.1.1 Multiple Themes and Theme Markedness

10.1.1.1 Forefield hypothesis vs. first experiential element hypothesis

	EO		GT		ET		GO	
Total Clauses	4496		4496		4901		4901	
Total Themes	5097		4853		5625		5276	
Single Themes	3955	88.0%	4165	92.6%	4230	86.3%	4544	92.7%
Multiple Themes	541	12.0%	331	7.4%	671	13.7%	357	7.3%
Avg. # of Themes	1.13		1.08		1.15		1.08	
Experiential Themes	4496		4270		4809		4475	
Subject Themes	3568	79.4%	3018	70.7%	3677	76.5%	2842	63.5%
Circumstance Themes	829	18.4%	1058	24.8%	1095	22.8%	1255	28.0%
Complement Themes	46	1.0%	159	3.7%	34	0.7%	364	8.1%
Predicator Themes	4	0.1%	5	0.1%	3	0.1%	13	0.3%
Textual Themes	431	8.5%	425	8.8%	498	8.9%	528	10.0%
Interpersonal Themes	170	3.3%	158	3.3%	226	4.0%	234	4.4%
Cleft	49	1.1%	30	0.7%	92	1.6%	39	0.7%

Table 21 Basic Theme distribution in GO/GT (Forefield) and EO/ET (1st exp. element)

Table 21 shows the basic Theme distribution in terms of multiple Themes, average Theme number, Theme types, and Theme markedness. These Theme statistics are based on the first experiential element hypothesis in EO and ET and the Forefield hypothesis in GO and GT. These are the two most frequently used Theme hypotheses for each language and are also used in Freiwald (2016) to analyze Theme translation shifts in English-German and German-English translations of popular scientific texts.

The Theme in the Forefield hypothesis is generally restricted to a single Theme element unless a textual Theme in the form of a conjunction is included in the Forefield. English, on the other hand, is not restricted by any formal boundary and can theoretically include a variety of textual and interpersonal Themes. Unsurprisingly, the number of mul-

multiple Themes and average Theme elements is higher in EO than in GO, and this distribution is similar for ET and GT. In fact, the average Theme number in GT is virtually identical to that of GO. In ET, the number of Themes is not only higher than in the source language but also higher than in EO. The reason for this increase is the higher frequency of textual and interpersonal Themes in GO, which are largely preserved in ET. The differences in Theme number between EO and GT ($\chi^2 = 5.9841$, $df = 1$, $p\text{-value} = 0.01444$) as well as the difference between GO and ET ($\chi^2 = 11.175$, $df = 1$, $p\text{-value} < 0.001$) are statistically significant.

German has a weak notion of markedness and is more likely to have a non-Subject experiential Theme at the beginning of the clause. In translations into English, this contrastive difference can be a challenge, as the translator has to gauge whether to follow the word order of the source language or to avoid a marked style in the target language. In the other translation direction, there is no word order pattern in the source language that is marked in the target language, as the Subject Theme is the most common experiential Theme in both languages. However, the translator may still want to deviate from the repetitive style in English and make Theme variation more authentic in regards to the target language.

Both of these effects are observable in the Theme structures of English and German translations. The number of marked Themes is lower in ET compared to GO, but at the same time higher than in EO. The ET results thus stand in between both language systems. This difference in Theme markedness between GO and ET is statistically significant ($\chi^2 = 10.304$, $df = 1$, $p\text{-value} = 0.001327$). The same holds true for GT, which has more marked Themes than EO, which also represents a significant difference ($\chi^2 = 29.301$, $df = 1$, $p\text{-value} = 6.195e-08$). At the same time, GT has fewer marked Themes than GO. It is noticeable that this effect is more distinct in GT, where the distribution of marked and unmarked Themes is almost right in the middle between both language systems. The numbers for ET resemble those for EO much more than marked Themes in German. The high number of German Complement Themes in particular is not preserved in the ET, which in fact have fewer Complement Themes than even English original writing.

Textual Themes are more common in GO, which also affects GT and ET, which are again in the middle between both language systems.⁸⁹ Interpersonal Themes in German are

⁸⁹ However, it must be noted that this is also an effect of the Theme hypothesis. Since single textual Themes can take up the entire Forefield in German, they make up a higher relative frequency. This is particularly evident when considering translations from English to German, where the absolute frequency of textual

also more common, and as a consequence ET again demonstrates a mixture of EO and GO. The relative frequency of interpersonal Themes in the GT, however, is identical to that in EO. The number of cleft constructions almost triples in translations from German to English.

10.1.1.2 First experiential element hypothesis in both languages

	EO		GT		ET		GO	
Total Clauses	4496		4496		4901		4901	
Total Themes	5097		5400		5625		6083	
Single Themes	3955	88.0%	3959	88.1%	4230	86.3%	4204	85.8%
Multiple Themes	541	12.0%	537	11.9%	671	13.7%	697	14.2%
Avg. # of Themes	1.13		1.20		1.15		1.24	
Avg. # of Themes w/o Finite	1.13		1.13		1.15		1.16	
Experiential Themes	4496		4496		4809		4863	
Subject Themes	3568	79.4%	3244	72.2%	3677	76.5%	3200	65.8%
Circumstance Themes	829	18.4%	1052	23.4%	1095	22.8%	1259	25.9%
Complement Themes	46	1.0%	167	3.7%	34	0.7%	387	8.0%
Predicator Themes	4	0.1%	6	0.1%	3	0.1%	16	0.3%
Textual Themes	431	8.5%	426	7.9%	498	8.9%	527	8.7%
Interpersonal Themes	170	3.3%	478	8.9%	226	4.0%	655	10.8%
Finite Themes	0	0%	315	65.9%	0	0%	416	63.5%
Modal Adjuncts	170	100%	163	34.1%	226	100%	239	36.5%
Cleft	49	1.1%	27	0.6%	92	1.6%	38	0.6%

Table 22 Basic Theme distribution in GO/GT (1st exp. element) and EO/ET (1st exp. element)

Table 22 represents Theme distributions based on the first experiential element hypothesis in both languages. Now that the Theme can also extend past the Forefield, the number of Theme elements increases noticeably, surpassing the average number of Theme elements of English. The difference in total Theme numbers remains statistically significant in both translation directions, although the differences are now reversed: The average

Themes in fact decreases but the relative frequency increases due to the lower number of Themes in German overall.

Theme number significantly increases in translations into German ($\chi^2 = 8.7474$, $df = 1$, $p\text{-value} = 0.0031$) and significantly decreases in translations into English German ($\chi^2 = 17.921$, $df = 1$, $p\text{-value} = 2.303e-05$). Many of these additional Theme elements are interpersonal Themes in the form of finite verbal operators, which have to be placed in second position to ensure grammaticality in German. While the Finite does belong to the German Theme, its inclusion is not a matter of choice if a non-experiential element is positioned at the beginning of the German clause. For this reason, the average Theme numbers excluding Finite Themes are also presented in this table. These Theme numbers are almost identical across the languages and original and translated writing. The average is slightly higher in GO and ET, again due to the slightly higher numbers of textual and interpersonal Themes. The small differences that can be found between original and translated texts are not statistically significant (EO-GT: $\chi^2 = 0.0141$, $df = 1$, $p\text{-value} = 0.9053$; GO-ET: $\chi^2 = 0.1562$, $df = 1$, $p\text{-value} = 0.6927$). Thus, it is fair to claim that Theme numbers are virtually identical in the first experiential element hypothesis disregarding Finite Themes.

The previously mentioned distribution of marked Themes in ET and GT can also be observed in the first experiential element hypothesis. The number of unmarked Themes has increased in GO and GT, which is due to the fact that the onset of the Midfield favoring Subjects over Complements and circumstances. As a consequence, the distribution in ET resembles the distribution in GO more closely than before. The differences in marked Themes remain statistically significant in English-German translations ($\chi^2 = 34.431$, $df = 1$, $p\text{-value} = 4.417e-09$) as well as German-English translations ($\chi^2 = 15.534$, $df = 1$, $p\text{-value} = 8.103e-05$).

Non-experiential Themes remain more common in GO than in EO and the relative frequency of textual Themes and interpersonal Themes as modal Adjuncts are in fact even higher in ET compared to the source language. Nevertheless, these relative frequencies are misleading since they are calculated against the total number of Theme elements, which are lower in English under this Theme hypothesis. Considering absolute numbers is more meaningful in this case since the number of clauses are identical for each of the translation directions. Textual Themes and interpersonal Themes as Adjuncts decrease slightly in both translation directions but resemble the frequencies of the source language system more than the target language system.

10.1.1.3 Subject hypothesis in both languages

	EO		GT		ET		GO	
Total Clauses	4496		4496		4901		4901	
Total Themes	6067		7896		6858		9625	
Single Themes	3178	70.7%	2818	62.7%	3202	65.3%	2693	54.9%
Multiple Themes	1317	29.3%	1678	37.3%	1699	34.7%	2208	45.1%
Avg. # of Themes	1.35		1.76		1.40		1.96	
Avg. # of Themes w/o Finite	1.35		1.43		1.40		1.56	
Experiential Themes	5413		5816		5958		6729	
Subject Themes	4437	82.0%	4461	76.7%	4789	80.4%	4855	72.2%
Circumstance Themes	865	16.0%	1113	19.1%	1130	19.0%	1370	20.4%
Complement Themes	46	0.8%	188	3.2%	35	0.6%	425	6.3%
Predicator Themes	8	0.1%	19	0.3%	4	0.1%	33	0.5%
Textual Themes	434	7.2%	438	5.5%	504	7.3%	566	5.9%
Interpersonal Themes	220	3.6%	1642	20.8%	284	4.1%	2330	24.2%
Finite Themes	0	0%	1448	88.2%	0	0%	2001	85.9%
Modal Adjuncts	220	100%	194	11.8%	284	100%	329	14.1%
Cleft	57	1.1%	35	0.6%	112	1.6%	46	0.7%

Table 23 Basic Theme distribution in GO/GT (Subject) and EO/ET (Subject)

Table 23 represents the Theme distribution in English and German original and translated language following the Subject hypothesis, where every element is considered thematic up to and including the Subject of the clause. Average Theme numbers in German increase noticeably in this Theme hypothesis and surpass those of English even without Finite Themes. This is a direct consequence of the weaker notion of Theme markedness in German. Since there are more cases where a circumstance or a Complement is positioned before the Subject in German, the number of multiple experiential Themes is also higher leading to an increase of Theme elements overall. In translations from English to German, the differences in Theme number are significant both if the Finite Theme is regarded ($\chi^2 = 239.99$, $df = 1$, $p\text{-value} < 2.2e-16$) and if it is disregarded ($\chi^2 = 11.539$, $df = 1$, $p\text{-value} < 0.001$). The same holds true in the opposite translation direction (with Finite Themes: $\chi^2 = 467.63$, $df = 1$, $p\text{-value} < 2.2e-16$; without Finite Themes: $\chi^2 = 40.845$, $df = 1$, $p\text{-value} = 1.648e-10$).

Regarding Theme markedness, English and German translations remain in between both language systems. Interestingly enough, the number of marked circumstance Themes is now almost identical in both translations at 19.1% and 19.0%. However, ET still follows the distribution of the target language more closely. This is caused entirely by the low number of Complement Themes, which remains smaller than the frequency in GO as well as the frequency in EO. Again, differences in Theme markedness remain statistically significant in both translation directions (EO-GT: $\chi^2 = 28.456$, $df = 1$, $p\text{-value} = 9.583e-08$; GO-ET: $\chi^2 = 10.166$, $df = 1$, $p\text{-value} = 0.00143$).

Looking at the absolute frequencies of textual Themes and modal Adjuncts as interpersonal Themes, the numbers remain almost identical in translations from English to German. In the other translation direction, both of these numbers decrease noticeably.

10.1.2 Participant Themes and process types

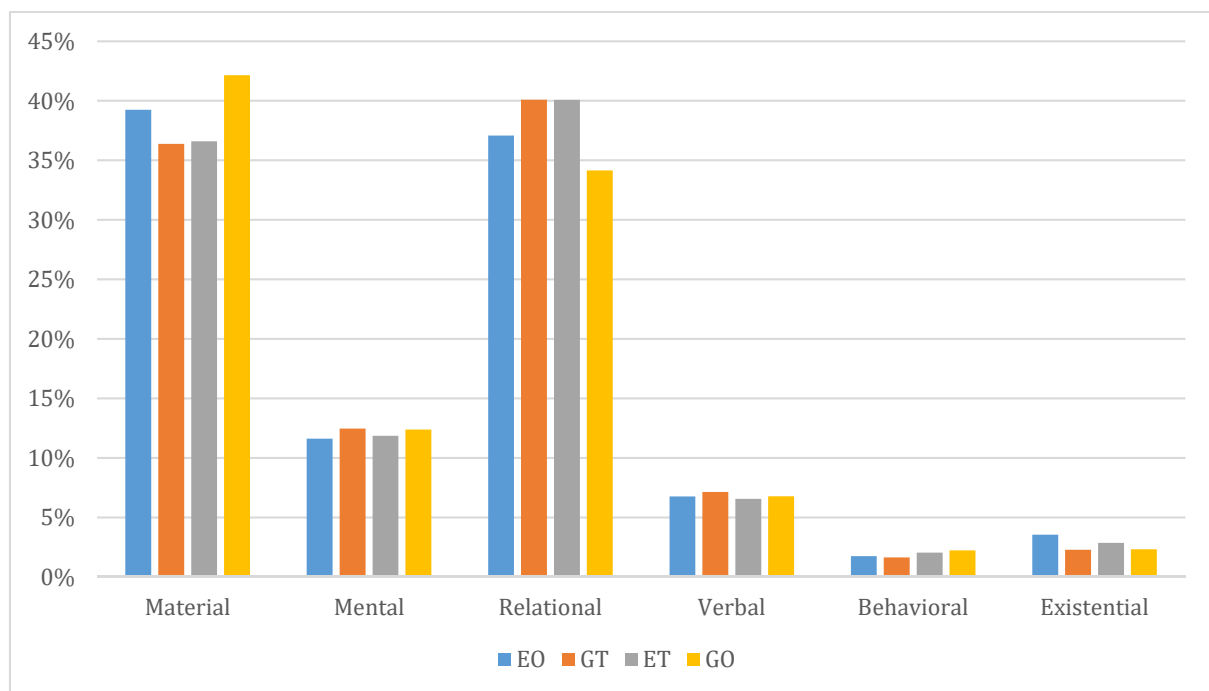


Figure 17 Process type distribution in EO, ET, GO and GT

Figure 17 shows the distribution of process types in English and German originals and translations overall. EO and GO already proved to be very similar in the use of the different process types with only slight differences for material and relational processes (see Section 9.1.2). Therefore, little variation from this distribution could be expected in the

translations, which for the most part turned out to be the case. Nonetheless, in the case of the aforementioned two process types, both ET and GT demonstrate higher relative frequencies of relational processes as well as lower relative frequencies of material processes compared to the original subcorpora. In fact, the use of all six process types is almost identical in ET and GT. That being said, these small differences in the use of process types are not statistically significant, neither in translations from English to German ($\chi^2 = 0.881$, $df = 1$, $p\text{-value} = 0.3479$), nor in translations from German to English ($\chi^2 = 0.0055$, $df = 1$, $p\text{-value} = 0.9408$).

While this is true for the distribution overall, the same effect cannot be observed in individual registers. Here, the distribution differs noticeably between ET and GT. There also seems to be no systematic relationship between process type distribution in the originals and translations. For example, in FICTION, GT has the lowest number of material processes, while ET places in between EO and GO. The exact opposite is the case in political speeches, where ET has the lowest number of material processes and GT is in between EO and GO. What looks to be systematic is that a lower relative frequency of material processes is compensated by a higher relative frequency of relational processes and vice versa.

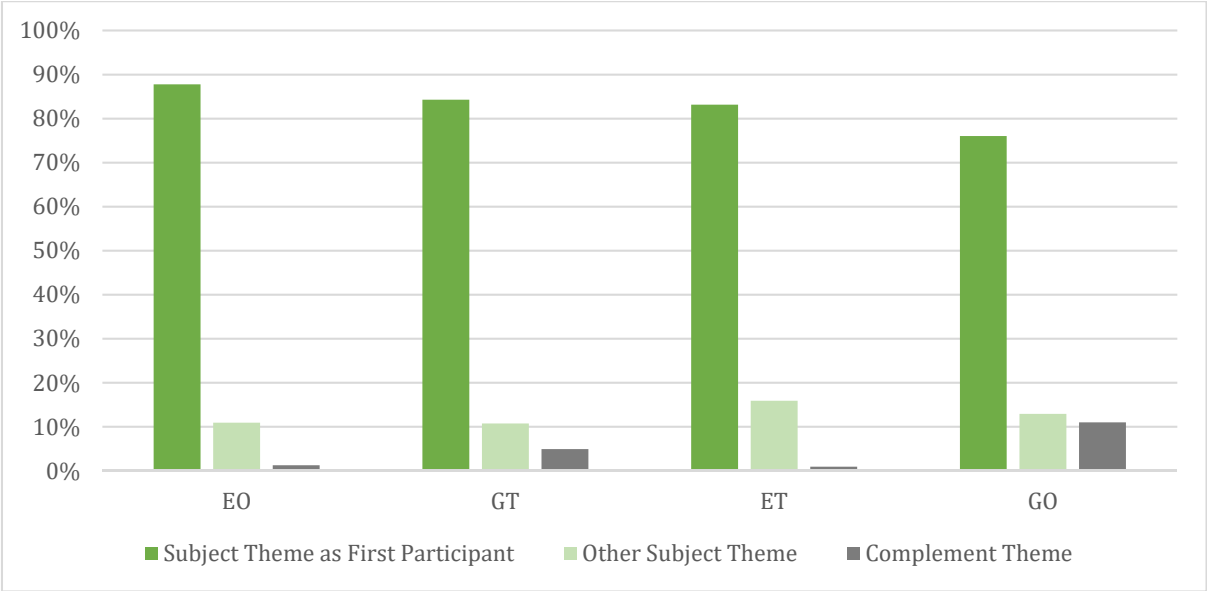


Figure 18 Participant roles and Theme types in EO, ET, GO and GT

Figure 18 shows the distribution of participant roles between the four subcorpora. It was already established that GO has the highest number of Complement Themes and lowest

number of Subject Themes. GT is in between GO and EO regarding Complement Themes, while the frequencies of Complement Themes in EO and ET are similarly low.

Regarding different participant roles in Subject Themes, GT uses almost the same amount of second participant Subject Themes (for example Phenomenon Subject Themes) as EO, granted that the difference between the two original subcorpora is already very small at roughly 2%. ET, however, uses the highest number of second participant Subject Themes out of all the subcorpora, more than both EO and GO at 15.9%. The majority of these Subject Themes are Goal Subjects in material processes and Value Subjects in identifying relational processes. Compared to GO, it is particularly Goal Subject Themes but also Phenomenon Subject Themes that increase in frequency in the translations. The majority of these Themes are part of passive constructions, which supports Doherty's (1996: 637) claim that shifts from active to passive are more readily available in English than in German translations. Teich (2003: 196-197) also reports a significant increase of passives and passive alternatives in translations into English and a significant decrease in translations into German.

10.1.3 Circumstances

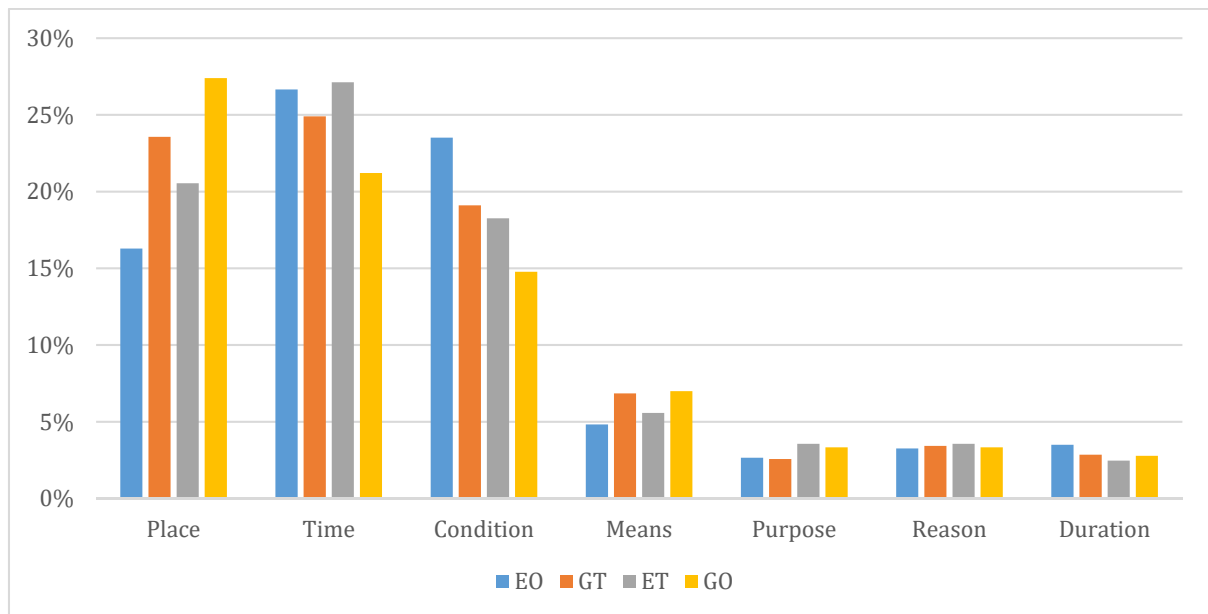


Figure 19 Circumstance Theme distribution in EO, ET, GO and GT

Figure 19 illustrates the general distribution of circumstance Themes in English and German originals and translations. This illustration does not reflect the frequencies of circumstance Themes per clause. Instead, it is a division of all circumstance Themes in each subcorpus according to their types. Figure 19 only takes those circumstance Theme types into account that made up at least 2% of circumstance Themes in each subcorpus.

The three most common circumstance Theme types are Place, Time, and Condition in all subcorpora. But, as has been pointed out in the discussion of the contrastive analysis, GO and EO differ heavily in the distribution of these three common circumstance types. Time and Condition Themes make up a much higher percentage of circumstance Themes in EO, whereas Place Themes are more common in GO. In almost all cases, the translations have a distribution that is in between the two original systems, with the only exception being Time Themes, which make up the highest proportion in ET.

Apart from the three most frequent circumstance Theme types, the differences of the remaining types are only marginal. The frequency of Reason is almost identical across the four categories and the numbers for Duration and Purpose are also very comparable. The frequency of Means Themes in ET is again in the middle of GO, where they are more frequent, and EO. The results for GT, on the other hand, are more similar to those in GO and do not seem to be heavily affected by the circumstance distributions of the source language. All in all, the distributions of circumstance Themes are not significantly different

in either translation direction (EO-GT: $\chi^2 = 0.0275$, $df = 1$, $p\text{-value} = 0.8682$; GO-ET: $\chi^2 = 0.0561$, $df = 1$, $p\text{-value} = 0.8128$)

	EO			GT			ET			GO		
Total Clauses	4496			4496			4901			4901		
Total circumstances	829	2443	33.9%	1052	2751	38.2%	1095	2860	38.3%	1259	3235	38.9%
Circumstances per clause	0.54			0.61			0.58			0.66		
Likelihood of Circ. Theme	18.4			23.40			22.3			25.7		
Additive	13	18	72.2%	13	22	59.1%	15	25	60.0%	15	34	44.1%
Behalf	19	50	38.0%	18	62	29%	33	80	41.3%	39	85	45.9%
Comitative	15	131	11.5%	15	120	12.5%	37	129	28.7%	45	154	29.2%
Concession	24	39	61.5%	24	39	61.5%	20	33	60.6%	20	34	58.8%
Condition	195	282	69.1%	201	274	73.4%	200	270	74.1%	186	252	73.8%
Comparison	18	51	35.3%	19	57	33.3%	18	54	33.3%	16	55	29.1%
Distance	0	6	0%	0	4	0%	0	12	0%	2	16	12.5%
Default	0	1	0%	0	2	0%	0	1	0%	0	1	0%
Degree	0	5	0%	0	5	0%	0	7	0%	1	12	8.3%
Duration	29	76	38.2%	30	78	38.5%	27	91	29.7%	35	92	38.0%
Frequency	6	44	13.6%	9	48	18.8%	18	44	40.9%	18	48	37.5%
Guise	29	61	47.5%	24	57	42.1%	24	78	30.8%	19	89	21.3%
Matter	10	21	47.6%	23	33	69.7%	13	29	44.8%	24	51	47.1%
Means	40	128	31.3%	72	198	36.4%	61	186	32.8%	88	232	37.9%
Place	135	520	26.0%	248	667	37.2%	225	737	30.5%	345	892	38.7%
Product	0	9	0%	0	7	0%	0	11	0%	0	16	0%
Purpose	22	253	8.7%	27	214	12.6%	39	161	24.2%	42	134	31.3%
Quality	15	184	8.2%	19	251	7.6%	8	225	3.6%	46	311	14.8%
Reason	27	104	26.0%	36	118	30.5%	39	104	37.5%	42	109	38.5%
Source	8	11	72.7%	11	12	91.7%	10	12	83.3%	5	13	38.5%
Time	221	444	49.8%	262	479	54.7%	297	559	53.1%	267	596	44.8%
Viewpoint	3	5	60.0%	1	4	25.0%	11	12	91.7%	4	9	44.4%

Table 24 Circumstances and thematic potential in EO, ET, GO and GT

Table 24 shows a complete list of all circumstance types and their positions in the clause. The first cell always represents the frequency of a circumstance type to be used as the Theme of its clause in absolute numbers. The second cell is the absolute number of circumstances used in the entire subcorpus. The third cell represents the thematic potential of each type of circumstance.

As was pointed out earlier, the number of circumstance Themes differs significantly between EO and GO even though the average TPot per circumstance is only slightly lower in English. The reason for this discrepancy between TPot and actual thematic use is a

difference in the use of circumstances generally. GO simply contains more circumstances overall, irrespective of position, and more circumstances end up being used as the point of departure as a consequence. The average thematic potential in GT and ET is almost identical to that of GO. Nevertheless, the frequencies of circumstance are considerably lower in both subcorpora compared to GO. In the case of GT, this makes sense since EO contains fewer circumstances. Thus, if the overall thematic potential of circumstances is the same in GT and GO, the number of circumstance Themes in GT will be smaller compared to GO. However, this explanation is not sufficient for English translations, as they are based on the German source texts. If ET and GO have an equally high circumstance TPot, the overall number of circumstance Themes should also be the same and yet it is not.

This discrepancy is again due to the number of circumstances overall. While GO contains 3235 circumstances in total, only 2860 circumstances can be found in ET, which is a decrease of 11.6%. This means that 11.6% of all circumstances used in the original texts are not included in their English translations. Conversely, the number of total circumstances used in English original texts increases by 12.6% in the German translations. In other words, 12.6% of all circumstances found in GT do not have an equivalent in the source texts.

This effect is not consistent across all circumstance types. For the most part, the number of circumstances used overall is very comparable between the source language and the target language. In the case of Condition, the absolute numbers even decrease in GT and increase in ET. However, there are three circumstance types in particular where this discrepancy is astonishingly large, namely circumstances of Means, Place, and Quality. The total number of circumstances of Means, regardless of position, increases by 55.7% in GT and decreases by 19.8% in ET. Circumstances of Place increase by 28.3% in GT and decrease by 18.4% in ET. And finally, circumstances of Quality increase by 36.4% in GT and decrease by 27.7% in ET.⁹⁰ This shows that the use of circumstances in the original language system influences translations to the effect that circumstantial information is apparently added or omitted in the translated texts. The actual number of circumstances inside and outside the Theme of course vary between registers, but the effect of adding and omitting particular circumstance types is consistent in all four registers.

⁹⁰ Further noticeable is Matter, which increases by 57.1% in GT and decreases by 56.7% in ET respectively. However, circumstances of Matter are overall rather infrequent, which is why this discrepancy can purely be variation.

10.1.4 Subject animacy and sentence

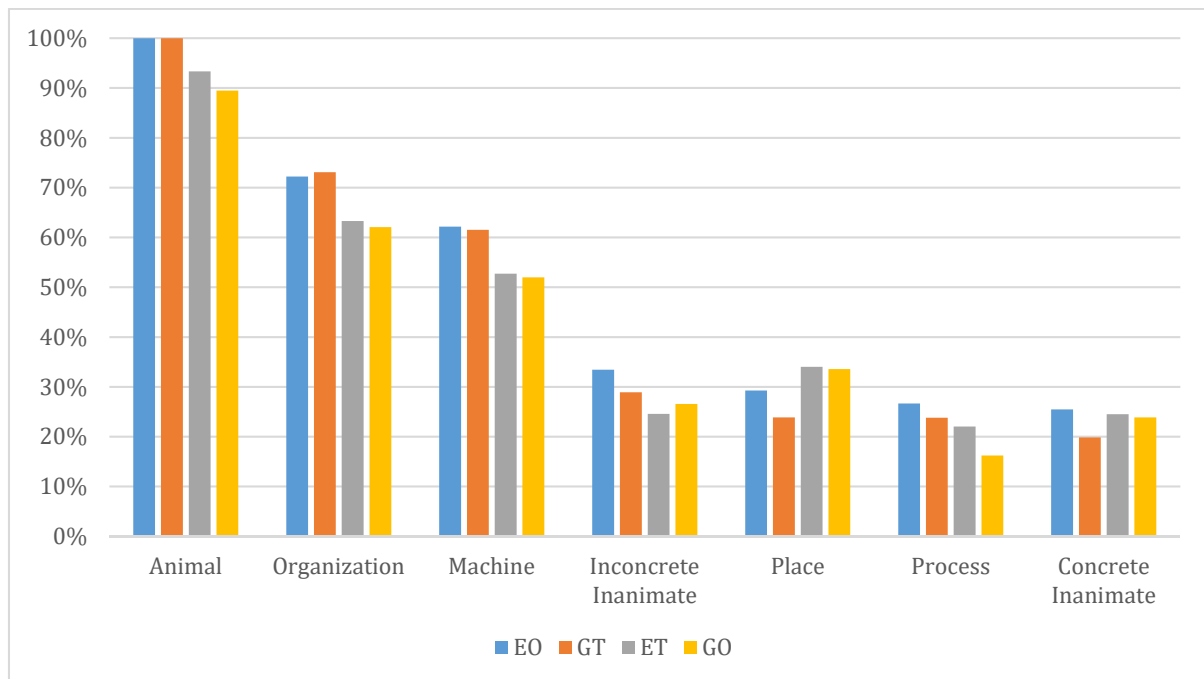


Figure 20 Subject Themes as part of non-sentient constructions in EO, ET, GO and GT

Subject Theme sentence is a contrastive difference between English and German and as such potentially poses a challenge in translations. Figure 20 shows the distribution of the different inanimate Subjects and their likelihood of being part of a non-sentient construction. For this presentation, the difference between Subject agency and Subject sentence is not made explicit as it has not proved to be a helpful distinction in practice (see discussion in 7.2 and 8.2). Also, Vehicle and Time have been disregarded, given their generally low number of occurrences in all subcorpora.

Non-sentient constructions are more common in English, regardless of the animacy level, with the only exception being Place Subjects. In translations, however, the animacy status of the Subject Theme does make a difference, particularly in GT. Subjects of middle animacy, including Animal, Organization, and also Machine, are used in non-sentient constructions at the same or even higher rate in GT compared to EO. Inanimate Subjects, like concrete and nonconcrete Inanimates, are considerably less likely to be used in a non-sentient construction compared to EO and even sometimes compared to GO. That being said the differences in the use of non-sentient constructions between EO and GT are not statistically significant, neither for middle animate Subjects ($\chi^2 = 1.0375$, $df = 1$, $p\text{-value} = 0.3084$) nor for inanimate Subjects ($\chi^2 = 2.273$, $df = 1$, $p\text{-value} = 0.1316$).

For most Theme-related variables up until now, ET seemed to be less affected by the source language and more affected by the target language system compared to GT. In the case of Subject Theme sentience, however, this trend does not continue. The frequencies of non-sentient constructions in ET are very similar to those of GO, regardless of animacy hierarchy. Even in the case of Place Subjects, which is the only inanimate Subject type that is more likely to be used as a sentient participant in German than in English, the frequency in ET is on par with that of GO. Only for Process Subjects, which is the least likely Subject type to be part of a non-sentient construction in German, are the numbers in ET noticeably higher and more in between EO and GO. In example (196), for instance the Human Subject in GO is changed to a Process Subject in ET (Subjects in bold), which is a common translation procedure, especially in the register of instruction manuals. However, given the close results for all other Subject types, it is unsurprising that the small differences between non-sentient constructions are not statistically significant, neither for middle animate Subjects ($\chi^2 = 0.2319$, $df = 1$, $p\text{-value} = 0.6301$) nor for inanimate Subjects ($\chi^2 = 1.4857$, $df = 1$, $p\text{-value} = 0.2229$).

(196)

GO: *Beim Anschluss an ein Fernsehgerät mit Euro-AV oder DIN-AV Buchse, erzielen **Sie** die beste Bild- und Tonqualität bei der Wiedergabe.*

*'With the connection to a television with Euro-AV or DIN-AV socket, achieve **you** the best picture- and sound quality during the playback.'*

ET: ***Connecting a TV set with Euro AV or DIN AV Socket** ensures best picture and sound quality during playback.*

[G2E_INSTR_008]

10.1.5 Subject identifiability

	EO		GT		ET		GO	
Subjects in total	4496		4496		4901		4901	
Identifiable	3247	80.3%	3406	83.5%	3426	78.8%	3438	80.2%
Non-Identifiable	797	19.7%	672	16.5%	921	21.2%	850	19.8%
Other	452		417		554		613	
Pre-verbal Subjects	4496		3048		4901		2900	
Identifiable	3247	80.3%	2339	84.4%	3426	78.8%	2109	81.2%
Non-Identifiable	797	19.7%	432	15.6%	921	21.2%	488	18.8%
Post-verbal Subjects	-		1448		-		2001	
Identifiable	-	-	1067	81.6%	-	-	1329	78.6%
Non-Identifiable	-	-	240	18.4%	-	-	362	21.4%
Subjects in immediate post-verbal position	-		1311		-		1649	
Identifiable	-	-	1025	85.8%	-	-	1195	82.1%
Non-Identifiable	-	-	170	14.2%	-	-	260	17.9%

Table 25 Subject Theme identifiability in EO, ET, GO and GT

Table 25 shows the comparison of Subject Theme identifiability between English and German originals and translations. The German results are separated into pre- and post-verbal Subjects since the position of the Subject relative to the finite verb is relevant in German. While it is also possible for the Subject in English to follow the verb instead of preceding it, such word orders are extremely rare and not directly linked to the identifiability or non-identifiability of the Subject. This is why Table 25 does not include frequencies of post-verbal Subjects in English.

The distribution of identifiable and non-identifiable Subjects is identical between EO and GO and yet the number of identifiable Subjects decreases in ET and increases in GT. What is also noticeable is that the relative frequency of pre-verbal Subject Themes is higher in GT (67.8%) than in GO (59.2%). If the Subject Theme is moved to a post-Finite position in German, the likelihood of that Subject to immediately follow the finite verb is also higher in GT (90.5%) than in GO (82.4%). In other words, Subject Themes in GT are much more likely to be positioned either in the Forefield or the onset of the Midfield, while Subject Themes in GO are much more flexible in their positioning in Fore- and Midfield. In fact, only 3.0% of Subjects assume a late Midfield position in German translations, compared to 7.2% in GO. Interestingly enough, the differences in Subject Theme identifiability are statistically significant in translations from English to German ($\chi^2 = 14.7$, $df =$

1, p-value < 0.001), yet not significant in translations from German to English ($\chi^2 = 2.6502$, df = 1, p-value = 0.1035).

Regarding the relationship between identifiability and position, the same effect that was found for GO (see Section 7.1.5) is also observable in GT, which is that pre-verbal Subjects are more likely to be identifiable than non-identifiable even though there is no restriction of identifiability on a Subject in the Forefield. Non-identifiable Subjects are in fact more likely to be part of the Midfield. Nevertheless, these types of Subjects are generally placed in a late Midfield position, especially if other, identifiable referents are also part of the Midfield. This is exemplified by the increased number of identifiable Subjects that immediately follow the finite verb. If an identifiable Subject enters the Midfield, it is more likely to stay close to the rest of the Theme and to not be preceded by other elements. This distribution is consistent between GO and GT.

10.2 Discussion

What follows is a discussion of the general analysis of Theme differences between original and translated language. In Section 5.11, twelve hypotheses were put forward, which will be assessed in the following as well. For reference, here are again the hypotheses for Theme differences in translations:

a. English-German translations

- H_{1.1.1} In the Forefield hypothesis, the average number of Theme elements decreases significantly in German translations due to the Finite-second constraint.
- H_{1.1.2} In the first experiential element hypothesis and the Subject hypothesis, the average number of Theme elements increases significantly in German translations due to the Finite-second constraint.
- H_{1.1.3} Non-Subject Themes are significantly more frequent in German translations across all Theme hypotheses.
- H_{1.1.4} Inanimate Subject Themes in combination with sentient verbs are significantly less frequent in German translations.
- H_{1.1.5} The frequency of middle animate Subject Themes in combination with sentient verbs in German translations do not deviate significantly from English originals. Machine Subjects behave like middle animate Subject Themes.
- H_{1.1.6} The number of identifiable Subject Themes significantly increases in German translations.

b. German-English translations

- H_{1.2.1} In the Forefield hypothesis, the average number of Theme elements increases significantly in English translations due to the Finite-second constraint.
- H_{1.2.2} In the first experiential element hypothesis, the average number of Theme elements decreases significantly in English translations due to the Finite-second constraint.
- H_{1.2.3} In the Subject hypothesis, the average number of Theme elements decreases significantly in English translations due to the Finite-second constraint and the higher number of non-Subject Themes.
- H_{1.2.4} Non-Subject Themes are significantly less frequent in English translations across all Theme hypotheses.
- H_{1.2.5} The number of identifiable Subject Themes significantly increases in English translations.

As was expected and previously shown in the analysis of the contrastive data (see Section 9.1.1), the choice in Theme hypothesis made a significant difference in the average Theme number in originals and translations. If the Theme in German can extend past the Forefield, it often contains an additional Finite, which is not part of the Theme in English. The other way around, if the German Theme is restricted to the Forefield, English can include additional Theme elements, which cannot all occupy the pre-Finite position in German and thus have to be moved to the Midfield. Thus, unsurprisingly, average Theme numbers in the two German sub-corpora were significantly lower than those of the English sub-corpora, when comparing Forefield and first experiential element hypothesis. This finding is consistent with the results in Freiwald (2016) and confirms H_{1.1.1} and H_{1.2.1}. In fact, Theme numbers are identical in GO and GT, while ET even includes more Theme elements than originals. This higher number of Theme elements in ET can be explained by the increased number in textual and interpersonal elements, likely a shining through (Teich 2003: 145) effect from the source language.

In terms of marked Themes, the circumstance Theme frequencies in GT as well as ET are in between English and German original writings. As is often the case, the feature distributions in translations are on the one hand subject to the shining through effect from the source text but at the same time also affected by a normalization effect (Baker 1996: 176-177) from the target language system. Throughout this discussion, these two simultaneous effects will be visible for many of the Theme-related features. For Complement Themes, this effect can also be observed in GT, where Complement Themes have a frequency of 3.7%, which is in the middle of the EO and GO frequencies. However, there is no shining through in ET. Instead, the use of Complement Themes is fully normalized in ET and even falls below the already low number of Complement Themes in EO. Complement Themes are more than 100 times (!) more common in GO than in ET. It seems as if English translators are well aware of this contrastive difference between English and German and try to avoid this highly marked constructions in the target language to the point where they over-correct. In most cases, Complement Themes in translations from German to English are moved out of the Theme into the Rheme (see example (197); Complements in bold). Another rather common translation procedure is changing voice from active to passive (see example (198); Subjects in bold, Goals underlined).

(197)

GO: *Fenster mag ich die starre Verglasung des zehnten Stockwerks nicht nennen [...].*
'**windows** want I the rigid glazing of the tenth floor not call [...].'

ET: *I wouldn't describe the rigid glazing on the tenth floor **as windows** [...].*

[G2E_FICTION_004]

(198)

GO: *Das entsprechende Anschlusskabel erhalten Sie bei Ihrem Fachhändler.*
'the corresponding connection.cable get **you** from your dealer.'

ET: *Corresponding connection cables can be obtained from your dealer.*

[G2E_INSTR_001]

In Neumann's (2003: 205) Theme analysis of tourist guides, she found that German translations resemble German originals more closely in terms of their thematic patterns. This impression is not corroborated by the results here. The frequency of circumstance Themes in GT is closer to GO than to EO, but apart from that all other Theme numbers in GT are more similar to the numbers in EO. One clear thematic difference in Neumann (2003) was the use of Predicator Themes, which were much higher in English tourist guides compared to German originals as well as German translations. This points towards more imperative clauses in the English subcorpus, which were not analyzed in this study.

If an experiential Theme in EO is paired with a textual or interpersonal Theme, the likelihood of it being changed in GT slightly increases from 12.7% to 14.5%. The same effect cannot be found in translations from German to English, where the likelihood remains exactly 20.0% regardless of whether additional Themes are present or not. This result supports Hasselgård's (2000: 36) claim that German translators are more likely to preserve the very first thematic element regardless to which metafunctional domain it belongs. That being said, the increase is relatively small, which is why this is only a tentative interpretation.

Interpersonal Themes remain the same and textual Themes increase between EO and GT in terms of relative frequencies. However, this is misleading since the relative frequencies are based on Theme elements overall, which GT have fewer of. In reality, the number of textual and interpersonal Themes both decrease in German translations, though the decrease of textual Themes is only marginal. It is not surprising that German translators would opt to move or omit certain non-experiential Themes because they often compete with other Theme elements, which is not the case in the EO. This also explains why this

effect is stronger for interpersonal Themes, since many textual Themes can share the Forefield position in German. In most cases, like (199), the non-experiential Theme is just moved behind the verb (textual element in bold). The same decrease in non-experiential Themes can also be observed in the opposite translation direction. Here, it is likely a case of normalization again, as English makes less use of these Theme elements than German.

(199)

EO: ***For example**, you might want to have large gray letters reading "draft" or "confidential" placed diagonally across the first page or all pages of a document.*

GT: *Sie können **beispielsweise** "Entwurf" oder "Vertraulich" in grossen grauen Buchstaben diagonal über die erste Seite oder alle Seiten eines Dokuments drucken.*

*'you can **for.example** "draft" or "confidential" in big gray letters diagonally over the first page or all pages of a document print.'*

[E2G_INSTR_001]

Cleft constructions almost triple in ET and are also more common compared to EO. Cleft constructions are primarily used in translations of German instruction manuals and political speeches. In INSTR, translators often choose a cleft construction when the Subject in German is a Process Subject formally realized as a nominal group. This Process Subject becomes part of the postponed clause in the cleft construction (see example (200); Process Subject in bold).

(200)

GO: *Leider ist aber **das Auslesen eines aktiven Dateisystems** streng genommen nicht möglich.*

*'unfortunately is however **the reading of an active file.system** strictly speaking not possible.'*

ET: *Strictly speaking it is not possible to read an active file system because the data might change during reading.*

[G2E_INSTR_007]

In SPEECH, German Complement Themes are oftentimes changed to a predicated Theme (Thompson 2014: 155-156) so that the thematic focus of the participant of the original Complement would not be lost while avoiding the marked Theme (see example (201); Complement Theme in bold).

(201)

GO: **Wahr** ist aber auch, dass das Umfeld für Terroristen häufig unter dem Eindruck der Chancenlosigkeit und einer unabwendbar erscheinenden Marginalisierung steht.

'**true** is but also that the environment for terrorists often under the impression of lack.of.prospects and an unavoidable seeming marginalization stands.'

ET: Yet it is also true that perceived marginalization and lack of prospects may favour a climate that is grist to the terrorists' mill.

[G2E_SPEECH_006]

When using the first experiential element hypothesis for both languages, the Theme element count in German surpasses that of English in both directions, which represents a significant difference in both translation directions. As was pointed out multiple times, this is primarily caused by the obligatory Finite Theme element that has to be included in the German Theme if a non-experiential element occupies the Forefield. If this additional Theme is disregarded from the calculation, the Theme numbers are virtually identical in both translation directions and the minute differences between the sub-corpora become statistically non-significant. Example (202) shows one of many cases where the Theme count in GO is one higher because of the Finite Theme (in bold). This confirms the first part of H_{1.1.2} as well as H_{1.2.2} and also supports the claim that this difference is purely caused by the extra Finite Themes in German. The most common translation procedure of such multiple Themes is simply to move the finite verb into its standard position in the target language.

(202)

GO: Gleichzeitig | **informiert** | eine neue, 68 Seiten starke Wellness-Broschüre über das Wellness-Ziel Mecklenburg-Vorpommern [...].

'at.the.same.time | **informs** | a new, 68 pages strong wellness-brochure over the wellness-destination Mecklenburg-Vorpommern [...].'

ET: At the same time | a new 68-page wellness brochure provides information about the wellness destination Mecklenburg-Western Pomerania [...].

[G2E_TOU_005]

Other than that, the same Theme distributions can also be observed in this combination of Theme hypotheses. Circumstance Themes in both translations are in between the two language systems and Complement Themes remain a rarity in ET. The number of interpersonal Themes increase tremendously in both German subcorpora, simply because Finite Themes were analyzed as interpersonal finite verbal operator Themes.

Under the Subject Theme hypothesis, Theme numbers again diverge between English and German. Now, the German subcorpora have more Theme elements both with and without considering finite verbal operators. This effect is observable in both translation directions but more pronounced in the translations from German to English. This discrepancy is a direct effect from the higher number of marked Themes in German, which are now always part of a multiple Theme construction. In H_{1.2.3}, this exact effect was predicted for translations into English and the hypothesis is verified by the results. The second part of H_{1.1.2} can also be considered confirmed. However, here only the additional Finite Themes were assumed to cause the increase in Theme elements in German translations. The additional marked Themes in GT clearly contributed to the significant Theme difference as well, so while this part of H_{1.1.2} is generally accurate, the proposed explanation is incomplete.

Regarding marked Themes, the relative frequencies appear to be more similar for both translation directions in the Subject hypothesis. However, this is again just a statistical effect from the way the relative frequencies are calculated. Now that every Theme in English and German must include a Subject Theme, the number of Subject Themes naturally increases and as a consequence the relative frequencies of all other experiential Themes are diluted. In reality, the difference in marked circumstance Themes is greatest for this Theme hypothesis, as ET only contains 82.5% of circumstance Themes used in the original (compared to the 87.0% in the previous comparison), while circumstance Themes in GT have increased to 128.7% (compared to the previous 126.9%). Even though circumstances do not frequently assume the head position of the Midfield in German, this pattern still seems to be more common than multiple circumstance Themes in front of the Subject Theme in English (see example (203); circumstances in bold). If a German Theme does include multiple circumstance Themes, one of the circumstances is usually moved into the Rheme in English translations.

(203)

GO: *Bei jedem Drücken der Taste TRANSMIT wird automatisch das nächste Fernsehprogramm gesucht.*
'with every push of the button TRANSMIT is automatically the next TV-broadcaster searched.'

ET: *Each time you press the TRANSMIT Button, the next station will be searched automatically.*

[G2E_INSTR_008]

Across all Theme hypotheses the observation regarding Theme markedness is the same, which is that both circumstance and Complement Themes significantly increase in English to German translations and significantly decrease in German to English translations. This is hardly surprising given the results from other studies like Neumann (2014), Freiwald (2016), and Niemietz, Neumann, and Freiwald (2017). Thus, H_{1.1.3} as well as H_{1.2.4} are clearly confirmed by the data.

Process type distribution is relatively similar between the four sub-corpora. Both ET and GT have a higher number of relational processes, yet a lower number of material processes compared to the two original subcorpora. However, these small differences are not statistically significant overall. Teich's (2003: 197-198) finding that translations include more material processes than originals cannot be corroborated by the results of this thesis. This strongly suggests that process changes in translations are fairly register-specific.

The distribution in individual registers is a lot less consistent. In FICTION, for instance, GT has the lowest number of material processes even though GO has the highest. Similarly, in SPEECH, ET has the fewest material processes, while EO has the highest. In other registers, the process type frequencies fall in between both original subcorpora, as is the case for example with material processes in TOU, where ET is in between GO and EO.

Process type distribution in individual registers appears rather unsystematic, with no dominant translation effect at play. What is consistent across all registers, however, is that neither translation subcorpus ever has the highest number of material processes or the lowest number of relational processes. GT and ET even score the highest for relational processes in three of four registers. These results do suggest that relational processes are slightly more favored than material processes in translations. In fact, the most common kind of process type translation shift is from a material process in the original to a relational process in the translation.

In the translation direction from English to German, many of the material clauses that are turned into relational clauses include an inanimate Subject. As was hypothesized, non-sentient constructions represent a challenge for German translators, since such constructions are less common in the German system. A change from material to relational seems to be a common translation procedure to deal with a combination of inanimate Subject and material process (see example (204); Subjects in bold, Verbs underlined).

(204)

EO: *Judging by the abundance of high-spirited customers, **the playhouse** was screening a gem of a show.*

GT: *Dem Zustrom erwartungsfroher Zuschauer nach zu schliessen musste **das Programm** ausgezeichnet sein.*

*'the abundance of high-spirited customers to judge must **the programme** excellent be.'*

[E2G_FICTION_010]

Interestingly enough, this change also occurs regularly with material verbs that do not require agency according to FrameNet (Fillmore, Johnson, and Petruck 2003). In these cases, the meaning of the material process is often expressed through the Attribute in the translation (see example (205)). It seems as if the combination of an inanimate Subject with any process that describes an action is more commonly expressed as a state rather than an action in German.

(205)

EO: *Tax cuts have helped cushion the blow of the slowdown [...].*

GT: *Steuersenkungen waren behilflich, die negativen Auswirkungen des Abschwungs abzufangen [...].*

'tax.cuts were helpful, the negative effects of.the slowdown to cushion [...].'

[E2G_SPEECH_005]

In translation from German to English, relational processes are sometimes used to avoid Complement Themes. For instance, in example (206), the Complement Goal (in bold) is turned into a Subject Carrier (underlined) in the English translation. If possible, the original Actor is then turned into a circumstance.

(206)

GO: ***Coolen Stahl und edelste Tischkultur** zeigt Ostovics.*

'cool steel and noble menu.sets shows Ostovics.'

ET: *Cool steel and noble menu sets are on show at Ostovics.*

[G2E_TOU_022]

Similar to the other translation direction, material processes in German are also translated as a relational process in English by expressing the meaning of the process through the Attribute (see example (207)). This translation procedure appears to be independent

of source and target language. However, to make such a claim confidently, further language pairs would have to be investigated. It is also very clear that the translation of process types is highly dependent on the choice of register and had a different set of registers been considered, the outcome might have been quite different.

(207)

GO: *Das Scheitern des EU-Gipfels im vergangenen Jahr hat Europa hierbei leider ein Stück zurück geworfen.*

'the failure of the EU-summit in the past year has Europe here unfortunately a piece back thrown.'

ET: *Sadly, the failure of the European Summit last year has been something of a setback for Europe in this respect.*

[G2E_SPEECH_010]

The number of second participant Themes in GT is identical to that of EO, which hints at shining through. The number of Actor participants decreases in both ET and GT due to the general decrease of material processes. For that same reason, Carrier, Token, and Value Themes all increase in frequency. Second participant Themes are noticeably more common in ET, compared to both EO and GO. A change from first to second participant Subject Theme occurs frequently when the original clause includes a Complement Theme. This way, the Goal participant can stay in the same position in the clause and the verb is changed to passive. In most cases, the original Actor is omitted in the translation, which is why this translation shift can often be found in clauses like (208), where the original Subject has little communicative value (Subjects in bold, Goals underlined).

(208)

GO: *Die angezeigten Meldungen können **Sie** entweder speichern, drucken oder verwerfen.*
*'the shown messages can **you** either save, print or discard.'*

ET: ***The messages shown** can either be saved, printed or discarded.*

[G2E_INSTR_007]

One of the most astonishing translation effects of Theme involves circumstance Themes. There is serious discrepancy between the thematic potential of circumstances and their actual use between English and German. Individual circumstances have almost the same likelihood on average to be used as the Theme in English and German. However, German

has a higher number of circumstances per clause and as a consequence more circumstances are also used thematically. This contrastive difference also transfers to translations. The total number of circumstances increases in German translations in relation to English originals and, conversely, the circumstance number decreases in English translations in relation to German originals. Neumann (2003: 196) also found this increase in circumstance use in German translations of English tourist guides.

In the translation direction from English to German, the TPot of circumstances is on average identical, yet the number of circumstance Themes decreases since the overall number of circumstances decreases as well. Three circumstance types in particular were identified as being noticeably less frequent in ET, namely Means, Place, and Quality, which are also among the circumstance types that demonstrate the highest relative difference between EO and GO (see Section 9.1.3).

This result is quite astonishing. Influence from the target language system is a common effect in translations and normalization has been demonstrated for many of the Theme-related measures so far. That being said, I did not anticipate that the influence from the English target language was so strong that entire pieces of circumstantial information are omitted in English translations. One would assume that circumstances play an important role in contributing to the experiential meaning of the clause and that an omission would severely alter the meaning of the clause as a representation.

A qualitative analysis of such sentence pairs reveals that a total omission of the circumstantial information is not very common. In the case of Place, the likelihood for the information of place to be entirely lost is fairly low. This happens if the information of place has little communicative value (see example (209)). However, in most cases, the circumstance of Place is changed into a different type of circumstance with similar meaning (see example (210)) or the Place becomes the Subject of the English translation (see example (211); circumstances of Place in bold). This is a common translation procedure since English allows a wider semantic mapping onto clause elements (Hawkins 1986: 65).

(209)

GO: **Hier** liegen wir etwa auf der Hälfte.
'**here** lay we about on the half.'

ET: *We have attained about half that figure.*

[G2E_SPEECH_015]

(210)

GO: **Auf einer abendlichen Autofahrt zurück nach Beirut** lehnt er träumerisch-geniesse-
risch-melancholisch den Kopf über die Lehne [...].

'**on an evening car.drive back to Beirut** leans he dreamily-appreciatively-melan-
cholic the head over the seat [...].'

ET: *Driving back to Beirut one evening he leans his head back over the seat, dreamily [...].*

[G2E_FICTION_005]

(211)

GO: **In diesem Zeitfenster** wird die Uhrzeit der angezeigten SFI-Tafel angezeigt.
'**in this time.window** is the time of the shown SFI-table displayed.'

ET: *This window will show the time of the SFI table displayed.*

[G2E_INSTR_003]

Circumstances of Means are also frequently turned into Subjects, especially in INSTR, where many circumstances of Means are introduced by a preposition followed by a gerund. In the English translations, these gerunds are instead made the Subject (see example (212); circumstance of Means in bold, Subjects underlined). Cases where circumstances of Means are omitted entirely are not incredibly frequent, yet they do exist as well. This is again common if the circumstance does not carry much communicative value or if it refers back to a previously mentioned Means. Some translators also choose to translate Means as a textual element. Still, there are also many circumstances of Means that represent new information but are not included in the translations, often if they are inferable from context or can be expressed through other clause elements (see example (213); circumstance of Means in bold).

(212)

GO: **Durch Drücken der Pfeiltaste links** können Sie die Tafeln entsprechend wieder zurückblättern.

'**by pushing of.the arrow.button left** can you the tables correspondingly again backwards.page.'

ET: Pressing the arrow key left will correspondingly allow you to page backward through the tables.

[G2E_INSTR_003]

(213)

GO: **Mit einem Ruck** zog sie ihren Gürtel fest.

'**with a jerk** drew she her belt tight.'

ET: She jerked her belt tight.

[G2E_FICTION_006]

Circumstances of Quality often receive slight changes to their form in English translations, so that the translated circumstances were analyzed as a different type, in many cases as a Comitative. However, out of all the circumstances discussed, Quality is undoubtedly the one type that is left out of the translations the most. Particularly in FICTION and TOU, circumstances of Quality are often missing entirely from the translation (see example (214); circumstance of Quality in bold).

(214)

GO: **Kraftvoll** stampfen die Hufe des Warmbluts durch den feinen Sand.

'**powerfully** stomp the hooves of the warm.blood through the fine sand.'

ET: The hooves of the warm-blooded stamp through the fine sand.

[G2E_TOU_011]

In summary, the meaning of most of the circumstances that are not included in the translation is not lost entirely. Instead, the original circumstances are changed to a different clause element or a different type of circumstance. Yet, while the cases of circumstances that are omitted are the minority, they are nevertheless not a complete rarity.

In the opposite translation direction, the exact opposite effect can be observed. Circumstances of Place, Means, and Quality on the whole increase in German translations of English. For Place, the reversed translation procedures can be observed in translations into German. A high number of additional circumstances of Place stem from inanimate

Subjects in EO. Some of these Subjects are part of non-sentient constructions (see example (215); Place Subject in bold, circumstance of Place underlined) but many are also paired with processes that do not require sentience or agency. Apparently, inanimate Subjects alone are prone to translation shifts, which is particularly common in INSTR.

(215)

EO: **His May 1 speech** made clear his vision [...].

GT: In seiner Rede vom 1. Mai verdeutlichte er seine Vision [...].
'in his speech from.the 1st of.May clarified he his vision [...].'

[E2G_SPEECH_003]

Besides that, German translators also add circumstances of Place that do not have an original counterpart. Such circumstances of Place in clauses like (216) are usually deictic and refer back to a place that was previously mentioned (Circumstance of Place in bold). Just looking at this translation direction, one might assume that this is a case of explicitation (Baker 1996). However, given that the exact opposite effect in the other translation direction, it is also possible that it is a sign of target language influence.

(216)

EO: *The meeting was industry-wide [...].*

GT: **Da** versammelten sich Leute aus der ganzen Branche, [...].
'**there** gathered [refl-3sg] people from the whole industry, [...].'

[E2G_FICTION_003]

The increased number of circumstances of Means in GT comes almost entirely from INSTR. These added circumstances of Means are again a common translation procedure to resolve inanimate Subjects. Many of these Subjects are, like in the other translation direction, Gerund Subjects, which are changed to circumstances by adding a preposition (see example (217); circumstance of Means in bold, Subjects underlined). The Subject of the translated clause is usually the Complement from the original and the voice is changed to passive. This same procedure is also used for concrete and nonconcrete Inanimate Subjects. There was not a single case in the data where a circumstance of Means is simply added to the translated clause without any equivalence in the original text.

(217)

EO: *Connecting the monitor and the computer through the Apple Desktop Bus provides a path for information [...].*

GT: ***Durch die Verbindung von Monitor und Computer über den Apple Desktop Bus können Informationen [...] an den Computer übertragen werden.***
'through the connection of monitor and computer through the Apple Desktop Bus can information [...] to the computer transmitted be.'

[E2G_INSTR_002]

Most of the additional circumstances of Quality in German come from either other kinds of circumstances in the original, which are altered slightly in form or meaning, from parts of the process or from an Attribute. Circumstances of Quality that are simply added to the translation are rare, but they do exist. Most of these are again deictic and can also be argued to be textual elements instead (see example (218); circumstance of Quality in bold).

(218)

EO: *And she touched off a revolution of freedom across the American South.*

GT: ***So löste sie im Süden der Vereinigten Staaten eine Revolution der Freiheit aus.***
'this.way triggered she in.the south of.the United States a revolution of freedom.'

[E2G_SPEECH_011]

Translations of inanimate Subject Theme demonstrate the same two translation effects that have been observed for most other features so far, which is a combination of shining through and normalization. For the first time so far, the shining through effect is much more evident in ET than in GT. The general frequency of non-sentient constructions in English translations mirrors the frequency of GO much more than that of EO. Given the relatively high number of non-sentient constructions in English originals, one might assume that English translators include additional non-sentient constructions in their translations, similar to German translators adding additional Complement Themes in their translations of English texts. However, if a German original clause includes an inanimate Subject paired with a verb that does not require sentience, it is very uncommon for the English translator to change that to a non-sentient construction. The only Subject type which ET does pair with sentient verbs more commonly is Process Subjects. As was pointed out above, changing an original Means Theme into a Process Subject is a common translation procedure to avoid a marked Theme in English translation. In this case, the newly introduced Process Subject is paired with whatever process was used in the source

clause, which can require sentience or not. In example (219), the original Means Theme (in bold) is changed into Process Subject (underlined) in ET, while the original sentient verb remains the same. This means that the increase of Process Subjects as part of non-sentient constructions is not a result of changing sentience requirements but a result of newly added Process Subjects. In fact, there is not a single case in which the sentience requirements of a construction involving Process Subject Themes in the original was changed in the translations.

(219)

GO: ***Durch erneutes Drücken der Taste Videotext*** schalten Sie den Mixbetrieb [...] ein.
'***by repeated pressing of the button Videotext*** switch you the mixed.mode [...] on.'

ET: *Pressing the key teletext again* will switch the screen display to mixed mode [...].

[G2E_INSTR_003]

The relative frequency of non-sentient constructions in GT does not consistently correspond to the source or target language frequencies. In the case of middle animate Subjects, the frequencies are almost identical to those of EO and consistently higher compared to GO. Even though the relative frequency of non-sentient constructions differs contrastively between English and German to some degree, it appears as if this combination is not heavily marked in the German system. As a consequence, the higher the level of animacy, the less pressure on the translator there is to make a change to the original clause. This result confirms H_{1.1.5}.

In the case of inanimate Subjects, the numbers of non-sentient constructions are consistently lower in GT compared to EO. However, only in the case of nonconcrete Inanimate and Process Subjects do the relative frequencies line up in between source and target language. In the case of concrete Inanimates, the likelihood of a non-sentient construction decreases noticeably even though the contrastive difference is almost negligible. This is a case of normalization where the translators appear to over-correct the markedness of the original construction. For Place Subject Themes, the frequencies decrease again in GT despite the fact that non-sentient constructions including Place Subjects are more common in GO than in EO. For instance, in example (220), the original Place Subject (in bold) is changed into a circumstance of Place (underlined) and a Human Subject is introduced instead. This outcome is difficult to explain since there is apparently no influence from either the source or the target language. Potentially, translators are generally aware

of the markedness of non-sentient constructions in German and apply it broadly to all inanimate Subject types even in cases where such construction are common in the target language.

(220)

EO: *Located just east of Palm Springs, **this park** preserves the curious flora and fauna of the Mojave desert [...].*

GT: *Östlich von Palm Springs, finden Sie in diesem Park die ungewöhnliche Flora und Fauna der Mojave Wüste [...].*

'east of Palm Springs, find you in this park the curious flora and fauna of the Mojave Desert.'

[E2G_TOU_011]

The translation procedures used for inanimate Subject Themes and processes requiring sentience are diverse and differ between registers. There are two dominant procedures that are used by German translators to resolve this marked construction. Either the process is changed so that the inanimate Subject can stay in Subject Theme position but is not paired with a process that requires sentience (see example (221); inanimate Subjects in bold), or the original Subject is changed into a circumstantial Adjunct. In the second case, the new Subject is almost always an original Complement with a change in voice or a direct address of the reader (see example (222); Subjects in bold, circumstance Theme underlined).

(221)

EO: ***The barren peak of Slieve Donard, climbing steeply to 2,796 ft,** dominates the mysterious blue distance of the landscape.*

GT: ***Die kahle Spitze des Slieve Donard, der sich steil auf 852 m erhebt,** ragt über die geheimnisvoll blaue Landschaft empor.*

*'**the barren peak of Slieve Donard, which [refl-3sg] steeply to 852 m rises,** towers over the mysterious blue landscape.'*

[E2G_TOU_001]

(222)

EO: ***Continuous layout** arranges the pages in a continuous vertical column.*

GT: *Im Layout "Fortlaufend" werden **die Seiten** in einer fortlaufenden vertikalen Kolonne angeordnet.*

*'in the layout "continuous" are **the pages** in a continuous vertical column ordered.'*

[E2G_INSTR_003]

The first procedure is very common in SPEECH and INSTR, the second can frequently be found in TOU. In FICTION, no one dominant translation procedure could be detected. The fact that these two translation shifts are the most common shows that translators are more likely to preserve the semantic sequence in the Theme rather than the order of clause constituents since in both cases the inanimate entity stays in Theme position. A replacement of the inanimate Subject by an animate one is not very common and can only be found if the context allows for this change (see example (223); Subjects in bold).

(223)

EO: **Joyce's eyes** moved about evasively [...].

GT: **Joyce** wich ihrem Blick aus [...].

Joyce dodged her gaze [...].'

[E2G_FICTION_009]

While the differences in non-sentient constructions are noticeably different if the Subject is inanimate rather than middle animate, the number of non-sentient constructions does not differ significantly between EO and GT. This is consistent with the contrastive results regarding non-sentient constructions, which were also not significant, contrary to previous belief. I do believe that a small difference in sentience requirements and Subject-Verb pairings is systematic between English and German, both contrastively and in terms of translations. However, the semantic restrictions on the German Subject appear to be softening up and non-sentient construction seem to become increasingly common in German, which is potentially a result of the influence of English (König and Gast 2009: 108-109). A study of non-sentient constructions in German in a historical corpus may be very illuminating. Nevertheless, H_{1.1.4} needs to be rejected.

In terms of Subject Theme identifiability, ET and GT demonstrate directly opposite effects. Even though the likelihood of a Subject to be identifiable is almost identical in EO and GO, the number of identifiable Subject Themes decreases in ET and increases in GT. The decrease in identifiability of Subjects in ET cannot be easily explained with any of the translation effects discussed so far. If the Subject Theme is changed from identifiable to non-identifiable, in the vast majority of cases, the content of the Subject is not actually changed but the definite article is either changed into an indefinite article (see example (224); Subjects in bold) or lost entirely (see example (225)). This happens very frequently and across all registers.

(224)

GO: *Die verschiedensten politischen Vereine und Gruppen in Wien* arbeiten aktiv und federführend an der Beseitigung der Ungleichbehandlungen auf rechtlicher und gesellschaftlicher Ebene.

'the most diverse political associations and groups in Vienna are working actively and in charge on the abolishment of the inequality on the legal and social level.'

ET: *A diverse array of political associations and groups in Vienna* are working actively to lead the way in wiping out inequality on the legal and social levels.

[G2E_TOU_021]

(225)

GO: *Der Handel mit den neuen Mitgliedsstaaten* wächst dynamischer als der deutsche Aussenhandel insgesamt.

'the trade with the new member states grows more dynamically than the German foreign trade as a whole.'

ET: *Trade with the new member states* is growing more dynamically than German foreign trade as a whole.

[G2E_SPEECH_014]

Looking at these results, one might assume that more nominal groups in English are introduced by indefinite articles, yet the contrastive analysis does not support this assumption. Apparently, English translations are less specific and vaguer than the original texts. That being said, the differences in Subject Theme identifiability are not significant, so these results might not replicate using a different data set. It was assumed that the number of identifiable Subjects would increase in English translations as an effect of explicitation (Baker 1996) in the translations. This assumption was clearly incorrect and H_{1.2.5} needs to be rejected.

In GT, the identifiability of Subject Themes increases. It was originally hypothesized that this would happen. If the English clause has a multiple Theme including an indefinite Subject Theme and another Theme element, the German translator is faced with a difficult challenge. If they want to preserve the original Theme order, the finite verb would have to be inserted between the first Theme element and the Subject Theme. In this case the Subject Theme would move to the first Midfield position. However, this is an atypical position for non-identifiable information. For this reason, it was assumed that translators occasionally change the identifiability of the original Subject to avoid particularly marked information structures. While the general hypothesis of H_{1.1.6} regarding an increase in

identifiability is confirmed by the data, the proposed explanation for this effect is inadequate. There is no restriction on identifiability in the Forefield, so both identifiable and non-identifiable Subjects can be and are positioned there frequently. I predicted an increase in identifiability in GT due to the identifiability restrictions in the German Midfield. Neumann, Freiwald, and Heilmann (forthcoming) also analyzed Subject Theme identifiability in translations from English to German in popular scientific texts and they found a similar increase of identifiability in the German translations as well as increased behavioral measures during the translation of non-identifiable English Subjects. They also attributed these effects to the more restricted sequencing rules in the German Midfield and the consideration of different translation choices.

However, if the Midfield restrictions were the only cause for this increase in identifiability, there would be no reason for the numbers of identifiable Subjects to increase in the Forefield and yet they do. The number of identifiable Subject Themes is higher in GT regardless of position. The reasons for this discrepancy are difficult to make out and depend on the register. There is no obvious pattern in the registers of FICTION and TOU. In fact, in FICTION, the change from non-identifiable to identifiable is just as common as the reverse change. In INSTR, changes in identifiability appear to be a byproduct of other kinds of changes, as most of the changed non-identifiable Subjects are also inanimate Subjects. As was pointed out above, the most common way to translate inanimate Subject in INSTR is to make the Subject a circumstantial Adjunct and use either a Complement or a direct address of the reader in the form of a pronoun as the Subject of the translation (see example (226); Subjects in bold, circumstance Theme underlined). Pronouns are identifiable, so all of these animacy changes are also accompanied by identifiability changes even though the former is arguably the true source of the shift.

(226)

EO: ***Any additional arrow down and up arrow*** move the selection through the menu commands.

GT: *Durch jede weitere Betätigung der Auf- oder Abwärtspfeiltaste bewegen Sie die Auswahl durch die Menübefehle.*
'through every further push of the up- or down.arrow.button move ***you*** the selection through the menu.commands.'

[E2G_INSTR_005]

The majority of identifiability shifts occurs in SPEECH. These shifts are quite interesting as most of these Subjects are nominal groups that originally include an indefinite article or no article at all and that are changed to the same Subject, simply including a definite article in the German translation. This is the same type of change as in translations into English, only reversed. This could just be a register-related effect. Even though the average identifiability frequencies in GO and EO level out overall, Subjects in German political speeches are more likely to be identifiable than in English political speeches. Words like *U.S. foreign policy* or *freedom* are more likely to be accompanied by a definite article in German originals, which carries over to the German translations (see example (227); Subjects in bold). This does not explain why the overall frequency of identifiable Subject increases in GT, even though it is the same in EO and GO but at least the discrepancy in SPEECH can be explained by this contrastive difference.

(227)

EO: **Freedom** defines our opportunity and our challenge.

GT: **Die Freiheit** bestimmt unsere Chancen und unsere Herausforderungen.
'**the freedom** defines our chances and our challenges.'

[E2G_SPEECH_011]

11 Regression analysis of Theme in translations between English and German

11.1 Results: English to German translations

One of the main aims of the dissertation project is to gauge which Theme related factors affect the likelihood of translation shifts. To answer this question, logistic regression analyses were conducted. The response variable used in all logistic regressions is Change, which is represented both as a binary and as a numerical category. The predictor variables include all previously discussed Theme attributes, such as Theme number, Theme type, and Subject Theme Sentience, as well as the variable Register. Including all levels of the predictor variables, the regressions consider 57 variables. For each super-category, one variable was selected against which all other variables of that same super-category were compared. The models illustrate the relationship between each variable and Change in comparison to the relationship between Change and the reference variable in order to assess whether the differences are statistically significant. Text ID was used as a random factor (see Section 6.4 for a more detailed discussion of the models). It is important to keep in mind that a positive or negative effect on Change does not necessarily mean that the Theme category itself underwent change. For instance, existential processes in translations from English to German have a positive, significant effect on Theme Change in most regression models. This, however, does not necessarily mean that the existential process itself was changed. Instead, marked Themes could have frequently been added by the German translators, resulting in a positive effect on Theme Change.

Like in the previous chapter, three combinations of Theme hypotheses were investigated for both translation directions: First experiential element Hypothesis (English)/Forefield hypothesis (German), first experiential element hypothesis (both languages), and Subject hypothesis (both languages). For each of these combinations, two regression analyses were performed, one with Change as a binary and the other with Change as a numerical category. This totals six logistic regressions for each translation direction.

In the following sections, the results of all twelve models are summarized in two tables, in which all variables that were consistently as well as occasionally significant are included. These results are subsequently discussed and linked to the hypotheses. The individual results of each of the logistic regressions are added to the Appendix. Presenting

and interpreting the results of the regressions is challenging. As will be shown below, many of the measures do not consistently produce statistically significant or non-significant results across all six regressions. This can be due to a multitude of reasons. The distinction between binary and numerical Change can have an influence on the significance of a variable. For instance, a variable may be associated with a single type of change, which is why it has a positive, significant effect on binary Change. Yet, in terms of numerical Change that deviation from the reference category is only slight. In some cases, the choice of Theme hypothesis can have a direct effect on a variable, as is the case with Theme number. In other cases, particular Theme hypotheses factor in more data points, for example in the case of the Subject Theme hypothesis and all Subject Theme related measures, which results in a more accurate evaluation of the variable. And, of course, testing for the same variables multiple times can also result in type I and II errors. To avoid an over-interpretation of inconsistently significant variables and ensure readability and clarity of the analyses, I will focus on those variables that were statistically significant repeatedly. Above that, I will attach more importance to results of Theme hypotheses that include more data points for a given variable.

		FirstExp --> Forefield				FirstExp --> FirstExp				Subject --> Subject			
		Binary		Numerical		Binary		Numerical		Binary		Numerical	
Length of the exp. Theme Reference: Middle	Short	0.336	0.0007	0.136	0.0026	-0.191	0.0509	0.119	0.0106	0.437	0.0001	0.249	1.32E-08
	Long	-0.008	0.9311	-0.052	0.2465	0.191	0.0372	0.005	0.9202	0.496	1.28E-06	0.119	0.0017
	Very long	-0.143	0.3665	-0.097	0.2140	-0.053	0.7366	-0.035	0.6592	0.744	9.92E-07	0.134	0.0146
Marked Themes Reference: Subject	Additive	0.299	0.6152	-0.062	0.8711	-0.273	0.6608	-0.365	0.4178	3.445	0.0031	0.684	0.0025
	Behalf	1.493	0.0219	0.214	0.3097	1.003	0.0888	0.182	0.3971	1.780	0.0303	0.290	0.1019
	Concession	0.445	0.3274	0.297	0.1728	-0.365	0.4331	-0.332	0.2606	3.082	0.0006	0.797	3.17E-06
	Condition	-0.383	0.0442	-0.475	1.69E-05	-0.702	0.0003	-0.482	1.27E-05	5.370	6.62E-10	0.551	1.21E-07
	Guise	1.027	0.0115	0.264	0.1908	0.651	0.1020	0.148	0.4986	18.446	0.27159	0.873	3.14E-08
	Means	0.188	0.5915	-0.300	0.1996	0.053	0.8785	-0.153	0.4828	4.767	2.39E-05	0.857	3.22E-09
	Place	0.167	0.3970	-0.053	0.6111	0.034	0.8636	-0.098	0.3698	3.307	3.83E-08	0.705	1.31E-10
	Purpose	0.722	0.1219	0.161	0.5046	-0.341	0.4816	0.213	0.3658	4.496	9.28E-05	0.797	1.30E-05
	Reason	0.021	0.9624	-0.142	0.5165	-0.335	0.4478	-0.324	0.1781	4.221	0.0003	0.406	0.0136
Time	-0.051	0.7610	-0.232	0.0113	-0.432	0.0127	-0.327	0.0009	4.900	1.88E-12	0.599	2.42E-08	
Participant Role Reference: Carrier	Actor	-0.570	3.38E-06	-0.206	0.0007	-0.331	0.0070	-0.207	0.0010	-0.605	1.28E-06	-0.337	8.48E-12
	Senser	-0.368	0.0108	-0.083	0.2353	-0.196	0.1789	-0.060	0.4055	-0.494	0.0008	-0.164	0.0045
	Sayer	-0.958	3.32E-07	-0.536	7.45E-08	-0.842	1.03E-05	-0.630	5.22E-09	-0.939	3.63E-07	-0.562	9.68E-13
	Behaver	-0.059	0.8399	-0.154	0.2644	-0.283	0.3400	-0.323	0.0336	-0.498	0.0912	-0.360	0.0041
	Existential	2.080	0.0002	0.304	0.0038	0.900	0.0125	0.301	0.0043	-0.695	0.0170	0.246	0.0056
	Phenomenon	0.732	0.0518	0.357	0.0141	0.490	0.1716	0.598	8.56E-06	0.455	0.2171	0.393	0.0002
	Empty	-1.884	0.0002	-0.338	0.1003	-1.353	0.0068	-0.247	0.2166	-1.433	0.0036	-0.339	0.0414
Non-Exp. Themes Reference: No Textual/No Interpersonal Theme	Textual Yes	-12.065	0.3328	0.771	4.57E-08	1.497	0.0310	0.791	1.52E-07	2.317	7.27E-06	0.432	1.69E-05
	Interpersonal Yes	-10.984	0.3780	0.876	2.48E-09	2.223	0.0025	0.876	2.37E-08	3.302	3.59E-08	0.641	1.67E-09
Subject Theme Identifiability Reference: Identifiable	Non-Identifiable	0.879	< 2e-16	0.369	1.73E-15	0.655	1.92E-10	0.365	1.09E-14	0.870	2.38E-16	0.363	< 2e-16
	Other	1.390	3.52E-11	0.339	2.13E-05	1.380	5.64E-12	0.425	4.95E-08	1.075	1.28E-07	0.301	6.80E-07

Subject Theme Sentience Reference: Human	Organization No Sentience	1.021	0.0073	0.312	0.0981	0.573	0.1320	0.331	0.0896	0.657	0.0826	0.372	0.0101
	Organization Sentience	0.795	0.0011	0.247	0.0621	0.953	8.10E-05	0.306	0.0218	0.884	0.0003	0.443	3.29E-06
	Machine No Sentience	1.798	0.0090	0.155	0.4553	0.423	0.4437	0.066	0.7523	1.236	0.0427	0.347	0.0164
	Concrete In- animate No Sentience	1.154	2.52E-05	0.474	8.24E-06	0.787	0.0034	0.486	7.09E-06	1.172	2.82E-05	0.512	7.58E-10
	Concrete In- animate Sentience	2.052	2.48E-05	0.677	1.53E-05	1.758	0.0003	0.761	6.25E-07	2.097	5.28E-05	0.749	1.31E-10
	Nonconcrete Inanimate No Sentience	0.910	5.88E-06	0.424	3.09E-06	0.701	0.0004	0.401	1.88E-05	0.725	0.0003	0.486	6.53E-11
	Nonconcrete Inanimate Sentience	1.624	6.53E-10	0.396	0.0005	1.072	1.88E-05	0.387	0.0009	1.039	4.00E-05	0.638	7.35E-13
	Place No Sentience	0.614	0.0715	0.239	0.1184	0.433	0.1991	0.277	0.0821	0.754	0.0333	0.412	0.0007
	Place Sentience	0.935	0.0656	0.515	0.0097	1.112	0.0330	0.637	0.0015	1.623	0.0036	0.698	1.15E-05
	process No Sentience	0.880	0.0290	0.317	0.0483	0.924	0.0242	0.556	0.0001	2.025	8.58E-05	0.760	5.41E-11
	process Sentience	1.330	0.0571	0.546	0.0240	1.328	0.0631	0.511	0.0352	1.997	0.0139	1.010	2.88E-10
Register Reference: FIC- TION	INSTR	0.401	0.0504	0.223	0.0349	0.395	0.0559	0.259	0.0252	0.471	0.0697	0.242	0.0349
	SPEECH	-0.438	0.0166	-0.223	0.0203	-0.377	0.0418	-0.246	0.0204	-0.250	0.2847	-0.282	0.0079
	TOU	0.455	0.0207	0.278	0.0061	0.331	0.0949	0.209	0.0621	0.221	0.3797	0.099	0.3729

Table 26 Summary of regression analyses 1-6

Table 26 shows a summary of all measures that produced statistically significant results, which are colored in green. It only includes those variables that were significant in at least two regression models. This excludes *Theme number*, the marked Theme types *Comitative*, *Comparison*, *Complement*, *Duration*, and *Matter*, the Participant Roles *Attribute*, *Token*, *Value*, *Initiator*, *Receiver*, and *Other* and the Subject Sentience types *Animal Sentient*, *Machine Sentient*, and *Other*. These categories either produced significant results in only one of the regressions or no significant results at all.

Regarding the length of the experiential Theme, short Themes are almost always associated with more translation shifts compared to Themes of medium length. In the Subject hypothesis, long and very long Themes are also changed significantly more often than medium sized Themes, but the same effect cannot be demonstrated for the other two Theme hypotheses.

The effects of marked Themes on translation shifts turn out very differently between the four models based on the first experiential element hypothesis as well as the Forefield hypothesis and the two models based on the Subject hypothesis. Marked Themes in the first four regression models are generally not more likely to undergo change compared to Subject Themes, with the exception of circumstances of Guise and Behalf. Circumstance Themes of Condition and Time are even significantly less likely to be changed than Subjects in translations into German. In the Subject hypothesis, all marked Themes have a positive effect on Change with most of these results being significant.

Regarding participant roles, most of the significant results have a negative effect on translation shifts. In other words, these are participant roles that are less likely to be changed, compared to the reference category of Carrier Themes. They include Actor, Sensor, Sayer, Behaver, and Empty Subject Themes. Empty Themes are usually *it* Subjects, which are non-referential but also not part of a cleft construction. The vast majority of these Themes are found in FICTION and are usually translated by a non-referential *es* Subject in German (see example (228); Subjects in bold). Existential processes and Phenomena are the only participant roles that have a significant, positive effect on Change.

(228)

EO: **It** was twelve, late for these parts.

GT: **Es** war zwölf Uhr, spät für diese Gegend.
'it was twelve o'clock, late for this part.'

[E2G_FICTION_009]

The use of textual and interpersonal Themes both have a positive effect on translation shifts, which is significant in five of the six regressions.

Non-identifiable Subject Themes are universally more likely to be changed than identifiable Subject Themes. The same is true for Subjects that were not analyzed in terms of their identifiability. A large number of these kinds of Subjects includes *there* Subjects used in existential processes, which have already been shown to be a significant effect on Change. Thus, these two categories largely overlap. Other examples of Subject Themes that cannot be analyzed in terms of identifiability are rank-shifted clauses (Halliday and Matthiessen 2014: 382) functioning as Subject.

The results for Subject Theme Sentience are not uniform at all. Subject Themes of middle animacy, Organization, and Machine, are more prone to change than Human Subject Themes in some of the regression analyses, even if the process does not require a sentient first participant. Concrete and nonconcrete Inanimate Subjects universally have a positive effect on Change, which is again independent of the semantic requirements stemming from the process. The effects of Place Subjects are positive without exception and mostly significant if the Place Subject is part of a non-sentient construction. In the case of Process Subject Themes, it is the other way around, where most of the positive, significant effects occur in cases where the verb does not require sentience.

The regressions also show that the amount of Theme translation shifts depends on the register. INSTR and TOU generally have a positive effect on Change in relation to FICTION, but this effect is not universally significant. SPEECH has a negative effect on Theme Change, which is statistically significant in five of six regression models.

11.2 Discussion: English to German translations

As was pointed out in the previous section, a discussion of six regression models based on similar data sets but with different results is a challenge. It is particularly problematic if some results confirm a hypothesis while others do not. For these reasons, the discussion will be focused on largely consistent results and the regression results that are based on the largest data sizes. If any percentages of change are referenced in the following paragraphs, they are based on the Theme hypotheses that are most meaningful for the feature in question. These are the same Theme hypotheses that were already argued to be most appropriate in the intralingual and contrastive chapters (see Sections 7.1 and 8.1).

To substantiate the discussion in this section as well as Section 11.4, each of the significant variables will be assessed regarding their most common types of change. For example, Organization Subjects in non-sentient constructions are a significant, positive predictor of Change in translations from English to German in four of six models. However, it is not actually Subject Theme sentience that is changed most frequently in the case of Organization Themes: The likelihood of a sentience shift in translations from English to German is on average 14.8%. In the case of a non-sentient construction including an Organization Subject that average in fact drops to 13.1%. This means that Organization Subjects are less likely to undergo a change of Subject sentience compared to the overall average. What does change more noticeably regarding Organization Themes is Theme number. The Theme number changes in 34.9% of cases considering the entire data set. However, if the Subject Theme is Organization, changes to Theme number increase to 40.0%. These measures are not interaction effects as they do not illustrate whether Organization Subjects correlate with a particular Theme number in EO. Instead, they demonstrate whether Organization Themes are associated with a change in Theme number in GT, regardless of what that Theme number is in the original. It is important to mention that these deviations between the global average and individual Theme types are not tested for statistical significance as this would have required another hundreds of significance tests. These comparisons are meant to enrich the discussion and illuminate the actual reasons for the high or low estimates in the regressions. However, given the lack of statistical testing, their meaningfulness should not be overstated.

For reference, these are the relevant hypotheses that were postulated in Section 5.11:

- H2.1.1 High Theme numbers are a significant predictor of Change.
- H2.1.2 Experiential Themes that are not of medium length are significant predictors of Change.
- H2.1.3 Comparison and Guise Themes are significant predictors of Change in translations from English to German.
- H2.1.4 Inanimate Subject Themes in combination with sentient verbs are significant predictors of Change.
- H2.1.5 Middle animate Subject Themes, including Machine Subjects, in combination with sentient verbs are not significant predictors of Change.
- H2.1.6 Non-identifiable Subject Themes are significant predictors of Change.

Regarding experiential Theme length, only short Themes deviate consistently from the reference category of medium Themes as they are more prone to change. Short Themes in English are between one and seven characters long and are usually personal pronouns or very short nominal groups. The most common type of change is also a change in length with many short Themes being translated as medium or even longer sized Themes in German. However, in the Forefield and the first experiential element hypothesis, many of the changes in length do not actually concern the same lexico-grammatical elements. Oftentimes, when the English original uses a short experiential Subject Theme in the form of a pronoun, the translator moves this Subject to the first Midfield position and fronts another experiential element like an Adjunct or a Complement instead (see example (229); Subjects in bold). This is why the second most common change involving short Themes is a change in participant role. This is particularly common in FICTION, where many Subject Themes in form of personal pronouns occur.

(229)

EO: **He** might see meadows or fountains [...].

GT: *Wiesen und Quellen werde **er** erblicken [...].*
'meadows and fountains might **he** see [...].'

[E2G_FICTION_002]

In the Subject hypothesis, a change in Theme length is not one of the more common types of change in response to short Themes in the English original. Occasionally, the Theme length of the Subject Theme is increased in the translations, which typically includes a personal pronoun that is turned into a nominal group with the same referent. This can be interpreted as a case of explicitation (Baker 1996). However, a much more common type of change is an increase in Theme number for the same reasons as outlined above. Short Subjects themselves are not actually changed frequently but moved to a later position in the German Theme instead, resulting in an increase in Theme elements in the translations. So, while the results technically confirm H_{2.1.2}, they do so for other reasons than originally assumed. Short Themes were hypothesized to be frequently turned into longer, more explicit version of themselves. Such cases like (230) do exist (Subjects in bold), but they are not the primary reason for the statistically significant results. This shows, however, that short identifiable Subject Themes can be moved out of the pre-verbal position more easily and make room for other clause elements in the Forefield, since they fulfill all the requirements of an early Midfield constituent.

(230)

EO: **It** *is relatively unknown [...].*

GT: **Dieser Park** *ist nicht sehr bekannt [...].*

this park *is not very well.known [...].'*

[E2G_TOU_011]

In the Subject hypothesis, long Subject Themes are also more prone to translation shifts compared to medium Subject Themes. In these cases, it is again Theme length that is changed most often. In the majority of cases, long Themes are turned into medium or even short Themes. Very rarely does a long Theme increase in size and turn into a very long Theme. In INSTR and TOU, many inanimate Subjects that are long are changed into circumstances, and a Human Subject is added to the clause (see example (231); Subjects in bold, circumstance Theme underlined). This is often a pronoun and thus short. Occasionally, modifiers are left out of translated nominal groups, resulting in reduced Theme length.

(231)

EO: *Security features help to prevent people from gaining access to information [...].*

GT: *Mit Hilfe von Sicherheitsfunktionen können Sie den Zugriff von Unbefugten auf kritische Informationen [...] unterbinden.*

'with the help of security features can you the access of unauthorized people to critical information [...] prevent.'

[E2G_INSTR_009]

However, in truth, there are a lot of observed changes of long Themes to medium Themes, where the medium Theme in German is a literal translation of the source but just a couple of characters shorter. It is even possible to have a word-for-word translation of a proper noun as Subject Theme and find a change in Theme length, since German Subjects are longer on average. In example (232), the Subject *General Electric* was used unaltered in the German translation. *General Electric* consists of 15 characters, which is right at the threshold of long Themes in English, but still within the range of medium Themes in German. This looks like an isolated incident, but such cases are numerous. This finding calls into question whether Theme length based on characters is in fact the most appropriate measurement.

(232)

EO: *For example, General Electric is among the ten largest investors in Poland, [...].*

GT: *General Electric zählt beispielsweise zu den zehn größten Investoren in Polen, [...].*

'General Electric belongs for example to the ten largest investors in Poland, [...].'

[E2G_SPEECH_014]

Very long Subject Themes are also primarily changed in terms of length. Given that *very long* is the longest category for length, this change is always a reduction. In INSTR, the same kinds of examples involving inanimate Subjects can be found for very long Themes. The data also includes similar 'incorrect' hits, where the most literal translation in German just happens to not be very long. Nevertheless, there are also cases where a very long Subject Theme is intentionally reduced in size in the translations. This is often the case if the head of the nominal group is followed by a long relative clause or an apposition. Such modifiers are typically not omitted in the translations. Instead, the rank-shifted clauses or nominal groups are frequently changed into ranking clauses, resulting in sentence splitting (see example (233)).

(233)

EO: *Little more than 100 miles west from London are the twin cities of Bath, made popular by the Romans 2000 years ago and revived by the Georgians splendours of the 18th century [...].*

GT: *Gute 150 km westlich von London liegen die Zwillingsstädte Bath und Bristol. Bath stand bereits vor 2000 Jahren bei den Römern hoch im Kurs und wurde im 18. Jahrhundert in gregorianischer Pracht zu neuem Leben erweckt.*

'little.more-than 150 km west of London are the twin.cities Bath and Bristol. Bath was already ago 2000 years among the Romans popular and was in.the 18th century in Gregorian's splendour to new life brought.back.'

[E2G_TOU_006]

In summary, H_{2.1.2}, which hypothesized that all experiential Themes that are not of medium length are significant predictors of Change, is generally confirmed by the data. However, these predicted effects are not due to the predicted causes, simplification and explicitation. While short experiential Themes do have a generally positive, significant effect on Change, an increase in length is rarely due to a more explicit translation of the same element, but instead caused by other factors like German word order flexibility. The same holds true for long Themes, which are only a statistically significant predictor variable in the Subject hypothesis. Their reduction in length in the German translations is often a byproduct of Subject Theme sentience. Solely the change of very long Subject Themes suggests an explicitation and simplification effect in GT. If the original long Subject is very complex and can also be turned into a ranking clause, the translators often decide to express this relationship more explicitly, usually by splitting up the sentence.

The results of marked Themes are among the least consistent. Almost every marked Theme has a significant, positive effect on Change in the Subject hypothesis, while in the Forefield and first experiential element hypothesis the estimates and the significance vary between the different kinds of marked Themes. However, there is an obvious reason for this inconsistency. If the Theme in English contains a marked Theme in the Subject hypothesis, it must have at least two if not more experiential Themes. Two experiential elements almost never share the Forefield position in German, which means the Finite must be included in the translated Theme, if both are meant to be kept in the translated Theme. This addition of the Finite increases the Theme number and is considered a type of change. If the translator instead decides to omit or move the marked Theme so as not to increase the Theme number, the experiential Theme types are different, which is also a type of change. Accordingly, by default, every use of a marked Theme in English must

be associated with some kind of change in the German translations in the Subject hypothesis, either in the form of an additional Theme element or a change of experiential Themes. This also explains the large effects in the binary model but the much smaller effects in the numerical model, since most changes only include an increase in Theme number and no further kinds of changes.

To accurately account for the effects of non-Subject Themes in English to German translations, the Forefield hypothesis and the first experiential element hypothesis need to be considered. Here, the results are generally consistent, as most marked Themes do not result in statistically significant, higher rates of change compared to Subject Themes. Two circumstance Theme types, Condition and Time, are even significantly less likely to be changed in GT. This result is not surprising in the case of Condition Themes since circumstances of Condition have the highest thematic potential at 74% in GO. This means that if they are used in a clause, they are more likely than any other experiential element, including the Subject, to be the Theme. In the case of Time Themes, the results were less predictable. Circumstances of Time have a generally high thematic potential in German at 44.8%, so a significant, positive effect was not to be expected. However, there are three other circumstance types that have an even higher thematic potential in German, namely Concession, Matter, and Behalf, so the fact that Time Themes in particular have a significant negative effect on Change is unexpected.

Two marked Themes have a significant, positive effect on Change, namely Behalf and Guise. However, they are only statistically significant in the regression analysis based on the Forefield Hypothesis and a binary distinction of Change. The effect of Behalf Themes on GT is very surprising given that Circumstances of Behalf are among the circumstances with the highest thematic potential in GO. The higher rate of change stems primarily from INSTR, which includes a number of circumstances that were analyzed as Behalf in English but as Place in German (see example (234); circumstance Themes in bold).

(234)

EO: **For Macintosh**, types and sizes are all grouped together under the Paper pop-up menu.

GT: **Auf Macintosh-Computern** sind sowohl die Papiertypen als auch die Papierformate im Popup-Menü Papier aufgeführt.

'on Macintosh-computers are both the paper.types and the paper.formats in.the popup-menu Papier listed.'

[E2G_INSTR_001]

The significant, positive effect from Guise Themes was expected. Circumstances of Guise have the second lowest thematic potential in German and at the same time a fairly high potential in English. They are one of the few circumstances that are more likely to be thematic in English than in German. For this reason, they were included in the revised version of H_{2.1.3} alongside Comparison Themes as a hypothesized strong predictor of Change in English to German translations. However, Guise Themes only have a significant effect in one of the four relevant regression models and that model is based on the Forefield Hypothesis in German, which considers fewer data points than the model using the first experiential element Hypothesis. Additionally, Comparison Themes are not statistically significant in any of the four relevant regression models. For this reason, I do not consider H_{2.1.3} to be confirmed by the data. Even though there are some discrepancies in thematic potential between EO and GO, it seems as if the German word order is more forgiving even in the case of rather marked thematic elements. Generally, the results give the impression that the German translators tried to adhere to the sentence structure in the source text as much as possible, even if the result is slightly marked.

These results on marked Themes do not mirror those by Freiwald (2016). In his analysis of Theme shifts in the popular scientific register, he employed a very similar logistic regression analysis to gauge the effects of non-Subject Themes on translations into German. He found that the only two circumstance Themes that had a significant, positive estimate on Change were Concession and Comparison (Freiwald 2016: 67). The results on Comparison match the assumptions of H_{2.1.3} but they are not corroborated by the results of this thesis. The fact that Concession Themes were among the more marked Themes in German translation is quite astounding given their high thematic potential in both German and English. That being said, Freiwald (2016) did not use the Subject but rather Time Themes as his reference category, which make the results less comparable. Also, he only analyzed a single register, and his data size was therefore limited. This is also the reason why Guise Themes, for example, did not enter Freiwald's (2016) regression analysis because their frequency was too low. The comparison does show though that Theme markedness does depend on register.

No significant predictors were hypothesized with regard to participant roles of experiential Themes simply because neither the state of the art nor the results in this thesis warranted such predictions. In this category, Carrier was used as the reference variable because it is the second most frequent participant role after Actor. The frequencies of relational processes increase in German as well as English translations and the relational

process type is the least affected by Subject Sentience differences. In spite of these facts, the first participant roles in four of the other five process types have a significant, negative effect on Change in translations. Hence, as it turns out, Themes including a Carrier Theme are more likely to undergo change compared to most other participant roles and process types.

In the analyses based on the Forefield and first experiential element hypothesis in German, Themes including a Carrier are slightly more likely to have a change of participant role. In terms of absolute frequencies, the most common change is from Subject Carrier to Subject Actor, which is not surprising given that material processes and Actor are the most common process type and participant role in GO and EO. The second most common participant change is from Carrier to Senser, both as Subject Senser and Complement Senser in the German translations. 16.7% of all participant changes of Carrier involve a Senser, even though their overall frequency in GT is only 8.1%. The majority of these changes occur in FICTION, where most of the mental processes are used. In English, many events revolving around feelings are expressed as relational processes, where the Attribute specifies the emotional or cognitive state of the Carrier. In GT, these processes are often changed to mental processes (see example (235)).

(235)

EO: *I am already disturbed by the ash falling from the blue distance down on these streets.*

GT: *Mich beunruhigt schon die Asche, die aus der blauen Ferne auf diese Strassen fällt.
'me[dat] worries already the ash with in the blue distance on the streets falls.'*

[E2G_FICTION_001]

In the Subject hypothesis, the likelihood of a change in participant role is actually below the overall average for Carrier Themes. The same is true for changes in Theme number, marked Themes, and Subject Theme sentience. The only Change category in which Carrier Themes deviate noticeably from the global average is Theme length. Apparently, Subject Carriers in EO are a lot more likely to be long or very long compared to the Subject length average. Of all the different Subjects in EO, 24.9% are long and 8.0% are very long. However, of the Subjects that are Carriers, 31.2% and 10.3% of Carriers are long and very long respectively. As was discussed earlier, long and very long experiential Themes are reliable predictors of change in the Subject hypothesis. Most of the long and very long Carrier Themes come from TOU and include a detailed description of a place or sight to

which an Attribute is ascribed. In translations, some of these detailed descriptions are omitted or moved (see above) or the reference is changed entirely, presumably to avoid unnecessarily long Themes. Nonetheless, it also seems to be the case that some of these inanimate Subjects are changed in the translations because the Attribute that is ascribed to them sounds atypical in German (see example (236); Subjects in bold, Attribute underlined). The analysis of Subject Theme Sentience was neglected in relational processes, assuming that any type of Subject can be used as the Carrier. And yet, the meaning of the Attribute may also be relevant regarding Subject Themes and semantic mappings.

(236)

EO: **Your visit to Grampian Highlands and Aberdeen** can be as energetic or relaxed as you want.

GT: **Sie** können Ihren Aufenthalt in den Grampian Highlands und Aberdeen so aktiv oder so erholsam gestalten, wie Sie wollen.

'**you** can your stay in the Grampian Highlands and Aberdeen as actively or as relaxing organize as you like.'

[E2G_TOU_004]

The two participant Themes that have a significant, positive effect on Change in relation to Carrier Themes are existential processes and Phenomenon Themes.⁹¹ The most common change involving existential process Themes is a change in participant role and process type. In 49.3% of participant role changes, the translator turns the existential process into a relational process. The majority of these cases, again, comes from TOU and they usually involve a circumstance of Place, either in Theme position as a marked Theme or in the Rheme. In the German translations, the circumstance of Place is either turned into the Subject and the original Existent becomes the Attribute (see example (237); circumstances of Place in bold) or the circumstance of Place itself becomes the Attribute and the Existent is turned into the Carrier (see example (238)).

⁹¹ Interestingly, the positive effects of Phenomena are only significant in the three regression models that treat change as a numerical category. This suggests that Phenomenon Themes are similarly likely to be changed compared to Carrier Themes. But if a Phenomenon is changed, it is oftentimes changed in numerous thematic aspects.

(237)

EO: *In the **Mid-Atlantic**, there are more wooded hills than factory chimneys, [...].*

GT: *Die mittleren Atlantikstaaten haben mehr bewaldete Hügel als Fabrikschornsteine, [...].
'the middle Atlantic.states have more wooded hills than factory.chimneys, [...].'*

[E2G_TOU_003]

(238)

EO: *There is an interpretation centre **at Lyness on Hoy**.*

GT: *In Lyness auf Hoy befindet sich ein Informationszentrum.
'In Lyness on Hoy is [refl.-3sg] an information.centre.'*

[E2G_TOU_002]

Phenomenon Themes are also mostly changed in terms of their participant role. Aside from a few exceptions, Phenomenon Themes are Subject Themes, and they can occur in both active and passive constructions. If the Phenomenon is part of a passive construction, it is commonly changed to an active clause, and the Phenomenon either stays in Theme position in form of a Complement (see example (239); Phenomena in bold), or the Sensor Subject is moved to the front (see example (240)).

(239)

EO: ***The result** can be seen in the generated HTML code after saving the source document.*

GT: ***Das Ergebnis** sehen Sie nach dem Speichern des Quelldokuments in dem generierten HTML-Code.
'**the result** see you after the saving of.the source.document in the generated HTML-code.'*

[E2G_INSTR_005]

(240)

EO: ***Caws Cenarth, which is delicious local cheese**, can be seen being made here.*

GT: *Man kann zuschauen, **wie Caws Cenarth, ein köstlicher örtlicher Käse hergestellt wird**.
'you can watch **how Caws Cenarth a delicious local cheese made is**.'*

[E2G_TOU_009]

Active clauses involving a Subject Phenomenon in EO usually stay active in the translations, but either the verb is changed slightly so that the Phenomenon Subject is turned

into a Senser Subject (see example (241); Phenomenon in bold) or the process is changed entirely.

(241)

EO: **Free trade** benefits us as consumers and our families, [...].

GT: *Wir als Verbraucher und unsere Familien profitieren vom **Freihandel**, [...].*
'we as consumers and our families benefit **from.the free.trade** [...].'

[E2G_SPEECH_005]

If the Theme in EO includes a textual or an interpersonal Theme, the likelihood of a change in the translations increases significantly. However, in the case of textual Themes, a change to the textual Theme itself is not particularly likely. 76.8% of textual Themes are also translated as Themes. The most common change involving textual Themes is a change in Theme number. In the Forefield hypothesis, Theme numbers are usually decreased, while in the first experiential element hypothesis and the Subject hypothesis they are usually increased. This has to do with the Finite-second constraint, which often prevents multiple Theme elements in the Forefield. If the English original includes a textual Theme (that is not a conjunction) and an experiential Theme, the translator has to make a choice which Theme element to keep. In the first experiential element hypothesis as well as the Subject hypothesis, the translator does not necessarily have to make that choice and the finite verb can enter the Theme as an additional interpersonal Theme, resulting in a higher Theme number (see example (242)).

(242)

EO: *Then | he got a thrashing for getting in the way [...].*

GT: *Dann | wurde | er verprügelt, weil er im Weg war, [...].*
'then | was | he beaten.up because he in.the way was [...].'

[E2G_FICTION_007]

In the case of interpersonal Themes, it is somewhat different. In the translations, you can find the same instances of Theme number increases and decreases depending on the German Theme hypothesis. These cases are also brought about by the Finite-second constraint. However, unlike textual Themes, interpersonal Themes are often not translated as Theme units. Only 45.4% of interpersonal Themes in the original are still interpersonal Themes in the translations. This low number of translated interpersonal Themes can also

be tightly linked to the question of Theme number. If the English original Theme includes an interpersonal Theme alongside an experiential Theme, the interpersonal element can only remain Theme if it is used as the very first element of the German clause. This is fairly independent of Theme hypothesis since it is uncommon for interpersonal elements to assume any other Theme position but the Forefield position. So, the 45.4% may represent the number of interpersonal Themes that are kept in the Forefield, while the rest is moved to the Rheme. There are plenty of examples like (243), where the interpersonal element (in bold) is moved from the Theme to the Rheme in the translation. However, there are also numerous cases where a thematic interpersonal Theme is not translated at all (see example (244)).

(243)

EO: Now, **unfortunately**, she was in and out of hospital.

GT: *Inzwischen mußte sie **leider** immer wieder für längere Zeit ins Krankenhaus.
'now must she **unfortunately** time and again for longer periods to the hospital.'*

[E2G_FICTION_009]

(244)

EO: **In fact**, a selection of properties in the care of the National Trust for Scotland remains open, [...].

GT: *Etliche im Besitz des National Trust for Scotland befindliche Grundstücke sind ganzjährig zugänglich [...].
'several in the possession of the National Trust for Scotland being properties are year-round accessible [...].'*

[E2G_TOU_002]

This distinguishes interpersonal Themes from textual Themes, which, for the most part, are still included in the translated clause even if not in Theme position. Whether or not the interpersonal element is omitted in the translations is very much dependent on its type. Interpersonal Themes of desirability such as *unfortunately* or probability such as *probably* are rarely excluded in the translations. On the contrary, interpersonal Themes of counterexpectancy like *indeed* or *in fact* (Halliday and Matthiessen 2014: 189-191) are omitted repeatedly.

Even though no hypotheses were formulated regarding the preservation or omission of textual and interpersonal Themes, these results still come as a surprise to me. If any-

thing, I would have assumed that it is more likely to leave out a textual Theme if the cohesive relation was already well established. Interpersonal elements reveal the personal stance of the author, and their omission arguably changes the interpretation of a text. That being said, the kinds of interpersonal Adjuncts that are left out frequently like *indeed* carry comparatively little personal stance.

Tightly linked to textual and interpersonal Themes is the category of Theme number. In H_{2.1.1}, it was hypothesized that high Theme numbers in EO are a strong predictor of Theme change in all Theme hypotheses. The premise of this hypothesis was that multiple Themes could generally not remain unchanged in GT regardless of the Theme hypothesis. In the Forefield hypothesis, most cases of multiple Themes in the source clause need to be reduced to one Theme element in the target clause because the Forefield can, for the most part, only include one element. In the first experiential element hypothesis and the Subject Hypothesis, the number of Theme elements in multiple Themes should increase because the additional Finite Theme is often added in GT to keep all original Theme elements. While the estimates of high Theme numbers are consistently positive, none of these results are significant. Theme number was expected to be one of the most reliable and strongest predictors of Change in German translations given previous results by Freiwald (2016) and the hard rule of the Finite-second constraint.

This unexpected result can only be explained by assuming that the category Theme number and textual/interpersonal Themes largely predict the same effect and therefore take away from their respective explanatory power. In the first experiential element hypothesis in English a multiple Theme is only possible if the English Theme includes at least one textual or interpersonal element. Hence, the combination of the variables Textual Theme and Interpersonal Theme are identical to the variable Theme number. In the Subject hypothesis, it is possible to have a multiple Theme without a textual or interpersonal Theme by combining a marked experiential Theme with a Subject Theme. But given the lower number of marked Themes in EO, textual and interpersonal Themes still make up the majority of multiple Themes in the English Subject hypothesis. If the predictor variables Textual Theme and Interpersonal Theme are excluded from the regression models, all of the results of Theme number become significant. All predictor variables were tested in terms of multicollinearity with the help of the `vif.mer` function in R (Frank 2014) and it was exactly these three variables that repeatedly had elevated levels. Nevertheless, I consider H_{2.1.1} falsified, even though I do not deem this result to accurately reflect the relationship between Theme number and Change.

Subject Theme identifiability has been demonstrated to be a very reliable predictor of translation changes in English to German translations. The category *Other* has a consistently significant effect on Change; however, this category is very heterogeneous as the Subjects that are part of this group have very little in common with each other, other than that their identifiability could not be determined. Also, many of the cases that are included in this category are existential processes, which were already discussed above.

Apart from *Other*, the data also clearly shows that non-identifiable Subjects are significantly more likely to be changed in some form than identifiable Subjects. Unsurprisingly, the most likely change for non-identifiable Subject Themes is a change of their identifiability: 30.1% of non-identifiable Subjects are changed to identifiable Subjects, as opposed to only 3.5% of identifiable Subjects that are translated as non-identifiable Subjects. The reasons for these changes are manifold. A higher number of identifiable Subjects in translations can be a sign of explicitation, where the translator tries to avoid any ambiguity by specifying the concrete reference of the Subject Theme. In translations from English to German, contrastive differences are likely also responsible for the increase of identifiable Subjects. The comparison between EO and GO has shown that there is no significant difference regarding Subject Theme identifiability overall. However, the comparison also showed that the onset of the Midfield in German does have higher numbers of identifiable Subjects due to the part-of-speech and definiteness sequencing rules of the Midfield. In GO, an author can include non-identifiable Subjects in an unmarked position in a clause by either using the Subject Theme as its Forefield element or positioning it late in the Midfield. However, if a non-identifiable Subject Theme is accompanied by an additional Theme element in the English original, like an interpersonal or marked experiential Theme, the translator has to make a decision:

- 1 They can keep the original word order largely intact by only moving the finite element to second position but accept the non-identifiable Subject Theme in the first Midfield position, which is a marked position for non-identifiable referents in German (see example (245); all Subjects in bold).
- 2 They can move the additional Theme element to the Rheme and position the non-identifiable Subject in the unmarked Forefield position (see example (246)).
- 3 They can move other, identifiable nominal groups to the onset of the Midfield and have the non-identifiable Subject assume a later position in the clause (see example (247)).

- 4 They can keep the original word order largely intact and avoid a marked Subject Theme position in the Midfield by changing the identifiability of the Subject Theme from non-identifiable to identifiable (see example (248)).
- 5 Or they can choose a new, identifiable Subject Theme in the translations (see example (249)).

All five translation procedures can be found in the data, but the two most frequent translation procedures are the first, where the translator accepts the marked Subject position in German, and the fourth, where the identifiability of the Subject Theme is changed. This fourth procedure is particularly common in INSTR and TOU, where many of the indefinite nominal groups can be translated by simply turning the indefinite or zero article to a definite article without much change in meaning. This procedure is the primary reason for the high number of identifiability changes of non-identifiable Subject Themes. These results clearly confirm H_{2.1.6}.

The results for Subject Theme sentience represent the most surprising results of all. It was originally hypothesized that the combination of inanimate Subject Themes and verbs requiring sentient first participants would be a strong predictor of Change. And this general assumption has, for the most part, been confirmed by the data. Yet, surprisingly, inanimate Subject Themes without sentient verbs and even middle animate Subjects like Organization and Machine deviate significantly from Human Subjects as well. This finding is not consistent across all logistic regressions. Middle animate Subjects do not have a significant effect in some of the models based on the Forefield and first experiential element hypotheses. However, in the Subject hypothesis, which takes into account animacy and sentience of all Subjects and verbs, the combination of Organization and sentient verb as well as Machine and non-sentient verb have a consistent significant and positive effect on Change. The only Subject Theme type that unequivocally does not deviate significantly from Human Subjects is Animal Subjects. Other than that, all Subject-verb combinations have shown significant effects in some of the regression models.

(245)

EO: *Thirty-six years ago, **another Texas President, Lyndon Johnson**, signed the first Elementary and Secondary Education Act [...].*

GT: *Vor 36 Jahren unterzeichnete **ein anderer texanischer Präsident, Lyndon Johnson**, das erste Grund- und Sekundärschulgesetz [...].*

*'ago 36 years signed **another Texan president, Lyndon Johnson**, the first Elementary-and Secondary.education.act [...].'*

[E2G_SPEECH_004]

(246)

EO: *In the front row, **critics from London** were taking their seats, [...].*

GT: ***Mehrere Kritiker aus London** nahmen ihre Plätze in der ersten Reihe ein, [...].*

'several critics from London took their seats in the first row [pfx lex. verb] [...].'

[E2G_FICTION_009]

(247)

EO: *In 1164 **Cistercian monks** first settled here.*

GT: *Im Jahre 1164 siedelten sich hier zuerst **Zisterziensermönche** an.*

*'in.the year 1164 settled [refl-3sg] here first **Cistercian.monks** [pfx lex. verb].'*

[E2G_TOU_009]

(248)

EO: *In Full Screen view, **PDF pages** fill the entire screen [...].*

GT: *Im Vollbildmodus füllen **die PDF-Seiten** den gesamten Bildschirm [...].*

*'in.the Full.screen.view fill **the PDF-pages** the entire screen [...].'*

[E2G_INSTR_003]

(249)

EO: *Before we knew it, **a teeming concourse** had gathered to debate the directions.*

GT: *Ehe wir uns versahen, waren **wir** von einer Menge Menschen umringt, die lebhaft über die richtige Richtung debattierten.*

*'before we [refl-3sg] knew were **we** by a crowd of.people surrounded who lively the right direction discussed.'*

[E2G_FICTION_010]

These results are entirely unexpected. It is debatable whether Machine Subjects should be considered part of the middle animate category of the animacy hierarchy, so it was not inconceivable that they would turn out to be a strong predictor of Change, even though H_{2.1.5} hypothesized that that is not the case. But the fact that even Organization Subjects,

which are undoubtedly in the middle animate Subject category, are changed significantly more often than Human Subjects is entirely unexpected and is also not consistent with the findings in Serbina (2015) and Freiwald (2016).

That being said, Organization Subjects are not changed frequently in terms of their sentence category. The vast majority of Organization Subjects stay Organization Subjects in the German translations.⁹² Organization Subjects score below the global average in all Change categories except one, which is Theme number. In other words, even though Organization Subjects deviate significantly from Human Subjects, this is not because of their status as middle animate Subjects but because they are frequently part of Themes whose Theme number is changed in the German translations. And the reason Theme number is frequently changed is because Organization Subjects are much more likely to be paired with circumstance Themes in English. 29.4% of Themes with an Organization Subject also include a circumstance Theme, compared to the 15.5% global average. And while marked Themes are not frequently changed themselves in GT, they are almost always the cause of an increase in Theme elements in the Subject Hypothesis (see above). This is the true reason for the significant results. It also explains why these results deviate from the results by Freiwald (2016), because his analyses only took into consideration whether the Subject sentence itself was changed.

This leaves the question why Organization Subject Themes are frequently paired with circumstance Themes. The vast majority of Organization Themes comes from SPEECH and they are usually accompanied by Time Themes. This is not saying much since Time Themes are the most frequent circumstance Theme type in EO overall and in SPEECH. That being said, the number of Time Themes in combination with Organization still exceeds the already high average number of Time Themes. From the data, it is not immediately clear why such Time-Organization combinations are so popular, especially in SPEECH, but there are numerous clauses that start with a reference to a point in time or a time period, followed by an Organization as Subject Theme, followed by their actions or accomplishments (see example (250); Time Theme underlined, Subject in bold). It seems to be very common to highlight either the point in time in the future when an initiative takes effect or to stress the longevity of, for instance, a government program, to either praise its usefulness or point out its ineffectiveness.

⁹² In total, out of 180 cases of Organization Subjects, only five are changed to a Human Subject and one to a nonconcrete Inanimate Subject. The rest either stay Organization Subjects, or Subject animacy was not analyzed (usually because the process was changed to relational).

(250)

EO: In the days and months ahead | **the Coalition** will work with you to provide security, justice and prosperity for all Iraqis.

GT: In den kommenden Tagen und Monaten | wird | **die Koalition** mit Ihnen zusammenarbeiten, um Sicherheit, Recht und Wohlstand für alle Iraker zu gewährleisten.
'in the up.coming days and months | will | **the Coalition** with you work.together to security, justice and prosperity for all Iraqis provide.'

[E2G_SPEECH_004]

In the case of Machine Subjects, the results look quite different. To be clear, all the Machine Subjects that were changed in some form in the translations come from INSTR, so any common translation procedures discussed here may also be purely register-dependent. In INSTR, Machine Subjects are changed frequently regarding their animacy. This is not only not dependent on the sentience requirements of the original verb, but Machine Subjects and non-sentient verbs are even more likely to be changed in the translations. There are two common procedures to translate Machine Subject Themes in German: Either the Subject is changed to a Human Subject and the Machine Subject is occasionally turned into a circumstance (see example (251); Subjects in bold, circumstance underlined) or the active clause is changed to a passive, which involves a change of participant role, usually from Actor to Goal (see example (252); Subjects in bold). The second translation procedure is by far the most common.

(251)

EO: **Profile Assistant** can save you from having to enter the same information, [...].

GT: Mit Hilfe des Profil-Assistenten können **Sie** vermeiden, immer wieder die gleichen Informationen eingeben zu müssen, [...].
'with.the help of.the profile-assistant can **you** prevent again and.again the same information enter to have [...].'

[E2G_INSTR_009]

(252)

EO: **The printer driver** opens.

GT: Daraufhin wird **der Druckertreiber** geöffnet.
'thereupon is **the printer.driver** opened.'

[E2G_INSTR_001]

Inanimate Subject Themes are consistently significant Change predictors, irrespective of verb requirements. This finding is again very surprising since it was assumed that inanimate Subject Themes only pose a problem if they are paired with a verb that has sentience or agency requirements. Instead, any inanimate Subject appears to be prone to change in GT. In the case of concrete and nonconcrete Inanimates, the kinds of changes are comparable to Machine Subjects, where the inanimate Subject is either changed to a Human Subject or turned into a Goal in a passive construction. For some nonconcrete Inanimates that are part of a non-sentient construction, the verb is changed so that the semantic requirements are not as restrictive (see example (253)).

(253)

EO: *Today's hearing follows on the heels of the President's meeting with President Putin in Genoa.*

GT: *Die heutige Anhörung findet direkt im Anschluss an das Treffen des Präsidenten mit Präsident Putin in Genua statt.*

'the present-day hearing takes directly after the meeting of the President with President Putin in Genoa place.'

[E2G_SPEECH_003]

Place Subjects occur predominantly in TOU, so their changes may again be register-specific. One possible translation procedure for Place Subjects is to turn the Place into a circumstance in the translation and introduce a new Subject (see example (254); Subjects in bold). However, while such cases do exist, they are surprisingly rare. A more common translation procedure again involves a change of the verb, typically turning the participant role of the Place Subject into a Carrier (see example (255); Subjects in bold).

(254)

EO: **Poole** mixes ancient with modern [...].

GT: *In Poole verbindet sich **die Vergangenheit mit der Moderne** [...].*
*'in Poole combines [refl-3sg] **the past with the modernity** [...].'*

[E2G_TOU_006]

(255)

EO: **Grimsbj** celebrates its history in the National Fishing Heritage Centre.

GT: **Grimsbj** verfügt über ein einzigartiges Erbe.
'Grimsbj possesses a unique heritage.'

[E2G_TOU_010]

Process Subjects are changed in a variety of ways. Regarding Theme length, many Process Subjects are either long or very long, especially in INSTR. These are typically rank-shifted clauses, which are either turned to nominal groups of reduced length or turned into a circumstance of Means in the German translations (see example (256); Process Subject in bold, Means Theme underlined). These formal changes of rank-shifted clauses functioning as Process Subject Themes are also often associated with changes in identifiability. Embedded clauses were not analyzed according to their identifiability, given that they can be neither definite nor indefinite. However, if they are translated as nominal groups in German, their identifiability automatically changes to either identifiable or non-identifiable. And lastly, Process Subjects are also often translated as circumstances of other types, and a new, often animate Subject is introduced instead (see example (257); Subjects in bold, circumstance Theme underlined).

(256)

EO: ***Carrying power tools with your finger on the switch or plugging in power tools that have the switch on*** invites accidents.

GT: *Durch das Tragen des Gerätes mit dem Finger am Schalter oder durch das Verbinden eingeschalteter Geräte werden Unfälle provoziert.*
 '*through the carrying of the tool with the finger on the switch or through the plugging of running tools are accidents invited.*'

[E2G_INSTR_008]

(257)

EO: ***Approaching the mountains from the west***, takes you through 50 miles of escalating hills and forested slopes [...].

GT: *Wenn man vom Westen kommt, fährt man* durch 80 km of langsam höher werdenden Hügeln und bewaldeten Abhängen [...].
 '*when you from the west come drive you* through 80 km of slowly escalating hills and forested slopes [...].'

[E2G_TOU_011]

These results on inanimate Subjects and non-sentient constructions only partially corroborate the results in Freiwald (2016). The same translation procedures that are common in this study could also be found in the register of popular scientific texts: 1. Re-mapping the inanimate entity on a circumstantial Adjunct and introducing a new Human Subject Theme or 2. Changing the semantic requirements of the verb. The first translation procedure also corresponds to Kast's (2012: 156-157) observation that Subjects which

have less prototypical semantic roles in the English originals are sometimes mapped onto different clause elements in the German translations. What is, however, not consistent is the rate with which such constructions are changed. Freiwald (2016: 92) reports that 50.0% of all non-agentive constructions were changed in popular scientific writings, while here the change rate is only 22.2%. Register thus seems to play a crucial role regarding the translations of non-sentient constructions.

Two hypotheses were formulated regarding Subject Theme sentience: H_{2.1.4} hypothesized that inanimate Subject Themes in combination with sentient verbs are significant predictors of Theme Change, while H_{2.1.5} predicted that middle animate Subject Themes, including Machine Subjects, in combination with sentient verbs are not significant predictors of Theme Change. On the basis of these results, H_{2.1.4} can clearly be confirmed and H_{2.1.5} clearly rejected. However, evaluating the underlying assumption that led to these hypotheses is more difficult. Middle animate Subjects were assumed to not predict Change because their animacy status was not considered to be problematic for German translators. In the case of Organization Subjects, this assumption is correct even though the predicted non-significant result is missing. As for the animacy status of Machine Subjects, the results are inconclusive. Machine Subjects produce statistically significant results in fewer regression models than the rest of inanimate Subject Themes. Above that, their estimates on Change are positive but generally lower compared to other inanimate Subjects. At the same time, their frequency of change, especially in terms of Subject sentience, is also not comparable with Organization Subjects. Machine Subjects are therefore somewhere in between middle animate and inanimate Subjects, at least in terms of the way they are translated into German. In terms of change rate and translation procedure, I would consider Machines more similar to concrete and nonconcrete Inanimates than Organization or Animal and therefore deem them the least marked inanimate Subject type in German.

Freiwald (2016: 88) reported that Machine Subjects remained largely unchanged in German translations of popular scientific texts. This was the case even if they were paired with a non-agentive verb, which led him to include Machine Subjects in the middle animate category. Given that the Machine results in this paper come primarily from the instruction manuals register, these two contrasting results clearly point towards register differences. Also, the kind of machine is likely to play an important role regarding verb combinability. In Freiwald (2016), many of the Machine Subjects were self-operating machines, which is clearly relevant when assessing their potential of being agents.

H_{2.1.4} predicted inanimate Subject Themes in non-sentient constructions to be strong predictor variables, which is the case. Yet, again, the underlying assumption that it is the sentence requirements of the verb, rather than the inanimacy of the Subject, which is the cause of these changes cannot be as easily discerned. The results for concrete and non-concrete Inanimates are consistently significant regardless of verb requirements. The results for Place and Process Subjects are occasionally significant and the sentence requirements of the verb only seem to play a minor role. Surprisingly, inanimate Subjects thus appear to be less marked in English generally, even if the verb does not presuppose certain semantic qualities from its first participant. That being said, the estimates of inanimate Subjects in non-sentient constructions are higher, with almost no exception. Consequently, while inanimate Subjects generally appear to be challenging in translations into German, it is still accurate that inanimate Subjects are even more prone to change if they are part of a non-sentient construction.

Lastly, the regression analysis demonstrates significant differences in change between the four analyzed registers. FICTION was arbitrarily chosen as the reference category and INSTR and TOU show a generally positive effect while SPEECH shows a generally negative effect on Change in relation to FICTION. These effects are only occasionally significant for each of the registers; nevertheless, it is evident that registers generally do play a significant role regarding the amount and types of Theme changes in English-German translations.

No hypothesis was formulated regarding the effects of individual registers on Theme Change. Still, the discovery that SPEECH and FICTION are less likely and INSTR more likely to experience change is unanticipated. INSTR, in particular, appears to be a very formulaic register in both languages with little variation between individual texts. However, that might be the reason why clauses in INSTR undergo so much change. If one kind of formulation or lexico-grammatical pattern is problematic in the target language German, it is bound to have a significant impact on the results, since it is repeated so many times in INSTR.

In her analysis of Themes in the registers of fiction and letters to shareholders, Neumann (2014: 303) also found that the German translations of fictional texts were generally in between source and target language originals, whereas translation in letters to shareholders rather followed target language norms. These results match as FICTION demonstrates the same relationship between source and target language in this study as well.

Considering all the Theme features that have already shown to trigger change makes the register results more comprehensible. The two Theme categories that are statistically most likely to cause change are Theme identifiability and Theme sentience. Withal, Carrier Themes and relational processes in general often undergo shifts. INSTR and TOU have by far the highest relative frequency of inanimate Subject Themes at 59.1% and 56.3% respectively, compared to the 18.8% in SPEECH and 9.7% in FICTION. INSTR and TOU are also the two registers with the highest number of non-identifiable Subject Themes. Additionally, TOU is the register that has by far the highest percentage of relational processes. In the light of these facts, the high number of changes in the two registers becomes less surprising.

What is more surprising, however, is the relationship between FICTION and SPEECH. FICTION has the lowest number of non-identifiable and by far the highest number of Human Subject Themes out of all registers. It also has the highest number of Time Themes, which were shown to be fairly resistant to change, and it includes a below average number of relational processes. Nevertheless, the number of changes in SPEECH are significantly lower. As it turns out, this difference is not due to any Theme category in particular; rather, a combination of multiple categories that explains this discrepancy. In comparison to FICTION, SPEECH shows higher frequencies of change in only two of the nine general Theme categories, namely circumstance Theme and Subject Theme identifiability. In the remaining seven categories, FICTION has consistently higher change rates, be it Theme number, Theme length or change of a textual Theme. These multiple disparities add up to the overall significant difference.

It is not immediately clear why SPEECH is below the global average in almost all Theme categories. For instance, SPEECH has the highest number of interpersonal Themes and, as was shown above, the presence of an interpersonal Theme has a significant effect on Change. Nevertheless, SPEECH scores lower than the overall average regarding changes to interpersonal Themes. Apparently, translators try to stay as close as possible to the source text in SPEECH, even more so than they already do in the other registers. Even if a lexico-grammatical pattern is marked in the target language, translators resist a change more often in favor of staying faithful to the original. This may be due to the general significance of a political speech and the importance of capturing its original meaning as exactly as possible. If a translator changes an original active construction to passive in an instruction manual to avoid a non-sentient construction, arguably little meaning is lost or changed. But if active is changed to passive in a political speech, the focus can shift from

the Actor to the Goal, which can significantly change the interpretation of the speech. If the subject matter is of international importance, like war or global warming, especially, a thematic change appears to be undesirable even if the result is stylistically marked in German. For instance, the translation in example (258) tolerates the non-sentient construction rather than change the thematic structure or the verb requirements of the original, presumably because the translator did not want to change the thematic focus or the process type.

(258)

EO: *The horror of September 11th, and the existence of al Qaeda cells in this and over 60 nations around the world, dispel any notion that America's commitment to the defeat of our enemies is mere rhetoric.*

GT: *Der Horror des 11. September und die Existenz von Al-Qaida-Zellen in diesem Land und in über 60 Ländern auf der ganzen Welt, strafen die Annahme Lügen, der politische Wille der Vereinigten Staaten, ihre Feinde zu besiegen, sei reine Rhetorik.*
'the horror of the 11th of September and the existence of al-Qaeda-cells in this country and in over 60 countries across the entire world, belie the assumption the political will of the United States their enemies to defeat is mere rhetoric.'

[E2G_SPEECH_008]

In summary, some of the hypotheses are confirmed by the data, while others are not. Subject Theme identifiability was predicted accurately. The hypotheses on Subject Theme sentience are not entirely accurate but the estimates of non-sentient constructions as well as the more qualitative analyses of translation procedures do support some of the underlying assumptions. Theme number did not have the expected effects in the regression models, but it was argued that this is primarily caused by multicollinearity rather than inaccurate assumptions. The hypotheses on Theme length were technically confirmed by the data; however, the explanations behind the effects were not accurately predicted. It was assumed that explication and simplification were the primary cause of changes to Theme length, but the analysis does not confirm that. In general, the translation features of explication and simplification are a lot less observable than originally hypothesized. The results on marked Themes do suggest some relationship between thematic potential and translation shifts. Nevertheless, the influence of thematic potential is a lot less strong in translations into German than assumed.

11.3 Results: German to English translations

In this section, the six regression models on German to English translations will be presented. There are three combinations of Theme hypotheses that were considered, and the response variable Change was used both as a binary and as a numerical category for each of these combinations. Converging the regression models of German to English translations was more difficult than in the other translation direction, especially for the models that have Change as a binary variable. One of the reasons for this is likely that fewer Theme annotations are available in the German originals. In the Forefield hypothesis, many German Themes do not contain an experiential Theme, so any variable that is tied to experiential Themes, like Theme markedness or Participant Theme, is not available. In the first experiential element hypothesis, experiential Theme measures are, by definition, considered. But since so many Themes in German do not begin with a Subject, the two Subject Theme measures Subject sentience and Subject identifiability are also frequently undefined. That being said, the model with the most converging issues is based on the Subject hypothesis and binary Change, so there must be additional reasons. To successfully converge the three models that use binary Change, some of the predictor variables had to be excluded: Subject sentience in the Forefield and first experiential element hypothesis and Marked Themes and Register in the Subject hypothesis.

Table 27 summarizes the results of the six regression models. The entire results of the regressions are added to the Appendix. As mentioned in the previous section, not all significant variables are consistently significant across all regression models. For this reason, those variables that were statistically significant repeatedly will again be in focus and more importance will be attached to the results of Theme hypotheses that consider more data points.

		Forefield --> FirstExp				FirstExp --> FirstExp				Subject --> Subject			
		Binary		Numerical		Binary		Numerical		Binary		Numerical	
Theme number		0.7819	0.00017	0.059	0.017884	2.5242	< 2e-16	0.4639	< 2e-16	3.443	< 2e-16	0.2995	< 2e-16
Length of the exp. Theme Reference: Middle	Short	-	2.46E-10	-	0.003138	-0.563	5.80E-10	-0.128	0.000851	-0.501	2.18E-05	-0.091	0.01154
	Long	-	3.40E-14	-	7.76E-06	-0.668	8.81E-14	-0.143	0.000103	-0.072	0.50535	-0.014	0.66164
	Very long	-	1.87E-12	-	2.01E-06	-1.065	4.05E-12	-0.231	0.000748	-0.761	5.70E-06	-0.09	0.05054
Marked Themes Reference: Subject	Behalf	0.4887	0.17959	0.616	0.017908	0.4844	0.1823	0.1988	0.367162	-	-	0.1522	0.19229
	Comitative	0.709	0.032	0.506	0.053045	0.7032	0.0331	0.0954	0.673528	-	-	-0.035	0.77417
	Concession	-	0.08121	-	0.092931	-1.113	0.0864	-1.339	0.011328	-	-	0.2595	0.09563
	Condition	-	0.12963	-	0.579236	-0.325	0.0845	-0.636	0.001853	-	-	0.1039	0.18465
	Duration	0.9881	0.01226	0.629	0.017911	1.0158	0.0102	0.2118	0.370139	-	-	0.0192	0.8893
	Guise	1.185	0.03183	0.807	0.007319	1.2745	0.0197	0.3077	0.260808	-	-	0.2008	0.23399
	Matter	1.2026	0.02366	0.658	0.021597	0.9641	0.0394	0.1393	0.590401	-	-	0.1236	0.39844
	Means	1.5498	9.67E-08	0.655	0.00512	1.5482	9.12E-08	0.2237	0.252773	-	-	0.2756	0.00204
	Place	0.7486	1.54E-07	0.505	0.020273	0.7339	2.30E-07	0.101	0.566941	-	-	0.1057	0.10958
	Quality	2.923	1.61E-06	1.041	1.69E-05	2.9235	1.59E-06	0.6799	0.000961	-	-	0.2591	0.02993
	Reason	0.5135	0.13165	0.535	0.042086	0.4395	0.1921	0.039	0.865695	-	-	0.121	0.3056
	Time	-	0.25012	0.143	0.522378	-0.292	0.0549	-0.391	0.034759	-	-	0.0094	0.89663
	Complement	3.2194	< 2e-16	1.248	1.27E-10	3.0837	< 2e-16	0.8103	8.85E-08	-	-	0.2269	0.00061
process	18.557	0.77726	1.023	0.000657	16.044	0.7908	0.4908	0.040426	-	-	-	-	
Participant Role Reference: Carrier	Empty	-	0.8992	0.317	0.009025	-0.006	0.987	0.1994	0.045502	0.3614	0.36463	0.3719	2.17E-07
	Goal	0.1424	0.411	0.16	0.033768	0.1517	0.3772	0.1562	0.01598	-0.109	0.5355	0.0872	0.07734
	Value	0.5729	0.0316	0.259	0.026621	0.5618	0.0346	0.2066	0.066966	0.2274	0.38728	0.0785	0.47122
	Initiator	0.6207	0.03207	0.357	0.153432	0.6097	0.0349	-0.119	0.577266	0.0319	0.92164	0.0373	0.74406

	Phenomenon	0.548	0.0707	0.366	0.000241	0.4861	0.1031	0.3245	0.000504	0.5756	0.07359	0.196	0.00582
	Senser	-0.3384	0.02288	0.243	0.232759	-0.322	0.0289	-0.058	0.720074	-0.102	0.51444	0.0702	0.16037
Non-Exp. Themes Reference: No Textual/No Interpersonal Theme	Textual Yes	-2.497	0.00107	0.087	0.403981	-3.665	5.22E-13	-0.082	0.345671	-2.402	< 2e-16	0.0814	0.11144
	Interpersonal Yes	-1.8825	0.01994	0.184	0.089072	-3.304	4.59E-09	-0.026	0.780816	-1.898	2.41E-06	0.0404	0.46528
Subject Theme Identifiability Reference: Identifiable	Other	1.6161	< 2e-16	0.725	< 2e-16	1.5703	< 2e-16	0.5872	< 2e-16	1.5145	2.66E-16	0.4097	< 2e-16
Subject Theme Sentience Reference: Human	Nonconcrete Inanimate No Sentience	-	-	0.243	0.013962	-	-	0.1931	0.020568	0.4023	0.06281	0.2268	0.00019
	Nonconcrete Inanimate Sentience	-	-	0.465	5.86E-05	-	-	0.2775	0.013554	0.7294	0.00974	0.1699	0.06673
	Place No Sentience	-	-	0.292	0.090914	-	-	0.1508	0.363425	0.3644	0.33782	0.2479	0.00327
	Place Sentience	-	-	0.481	0.017042	-	-	0.3695	0.03988	0.1003	0.82821	0.243	0.03761
Register Reference: FICTION	INSTR	0.2952	0.07679	0.077	0.252757	0.3197	0.0502	0.1869	0.01655	-	-	0.2162	0.00923
	SPEECH	0.8694	4.32E-10	0.319	2.54E-09	0.8653	2.09E-10	0.3946	4.20E-10	-	-	0.3266	2.75E-06
	TOU	0.7708	8.70E-08	0.23	3.82E-05	0.7842	2.73E-08	0.3111	1.38E-06	-	-	0.3848	2.57E-08

Table 27 Summary of regression analyses 7-12

As in the previous summary table, Table 27 only includes those variables that are significant in at least two regression models to increase readability and plausibility of the analyses. However, in categories that are missing from at least one of the models, which are Subject sentence, marked Themes and Register, variables are also included in the table if they are only significant once. Based on these restrictions, Table 27 excludes the marked Theme types *Additive*, *Comparison*, *Frequency*, and *Purpose*, the Participant Roles *Actor*, *Attribute*, *Behaver*, *Beneficiary*, *Existent*, *Existential process*, *Token*, *Sayer*, and *Verbiage*, *Non-identifiable Subject Themes*, and all Subject Sentence types apart from *nonconcrete Inanimates Sentient & Non-Sentient* and *Place Sentient & Non-Sentient*. These variables either produced significant results in only one of the regressions or no significant results at all.

Theme number is a consistently significant variable that has a positive effect on Change. Theme number does not have a reference variable because it was used as a scaled measure. This means that the higher the Theme number in the German originals gets, the more likely it is that the Theme undergoes some kind(s) of change in all Theme hypotheses.

The results on Theme length are also largely consistent. The reference variable is medium-sized Theme, which appears to be the length category that is changed the most given that short, long, and very long Themes all have a negative effect on Theme Change across all models. These effects are universally significant for short Themes. Long and very long Themes are significant in four and five regressions respectively.

The results on marked Themes are less uniform. Two types of marked Themes are clearly reliable predictors of Change, namely Quality and Complement Themes, which demonstrate significant positive effects in all five models that take Theme markedness into consideration. Other marked Themes that have produced significant positive results in three to four models include Duration, Guise, Matter, Means, and Place. The marked Theme types Behalf, Comitative, Reason, and process are inconsistently significant but also have a positive effect on Change. Three marked Themes have a negative effect on Change, namely Condition, Concession, and Time Themes, which means that they were changed less often than the reference category of Subject Themes. However, their results are only significant in one regression each.

The results on participant Themes are also inconsistent. The results of more than half of participant Themes do not significantly deviate from the reference variable of Carrier

Themes in more than one model. Participant Themes that are occasionally significant and have a positive effect on Change include Goal, Identifier, Initiator, and Phenomenon Themes. Senser is the only participant type that deviates significantly from Carriers with a negative effect on Change. The estimates of empty Subjects are not consistently positive or negative. However, their three significant results are all tied to positive effects.

The presence of textual and interpersonal elements in the German Themes does not have a universally positive or negative effect on Theme Change, either. In those cases, in which their results are significant, the effects are always negative. Incidentally, all significant, negative results occur in the models that use binary Change as the response variable.

Subject Themes whose identifiability was not analyzed, mainly semantically empty Subjects and rank-shifted clauses, deviate positively from identifiable Subject Themes. This result is consistently significant.

Moreover, the results of the different types of Subject Theme sentience are largely non-significant. Only four variables produced any significant results, which are not uniform across the regression models. These variables include nonconcrete Inanimates and Place Subjects in combinations with both sentient and non-sentient verbs. They all have a positive effect on Change, which means they are associated with more change compared to the reference category of Human Subject Themes.

Like in translations from English to German, the choice of register also affects the likelihood of changes in this translation direction. The effects of all three registers, INSTR, SPEECH, and TOU, are positive in comparison to the reference register FICTION. This means that Themes in FICTION are overall changed the least. For SPEECH and TOU, these results are consistently significant. For INSTR, this is only the case in two out of five regression models.

11.4 Discussion: German to English translations

In this section, the previously presented significant and non-significant predictor variables in German to English translations will be discussed. Additionally, the accuracy of the hypotheses on Change predictors will be evaluated. For reference, these are the relevant hypotheses that were postulated in Section 5.11:

- H2.2.1 High Theme numbers are a significant negative predictor of Change in the Forefield hypothesis.
 - H2.2.2 High Theme numbers are a significant positive predictor of Change in the first experiential element hypothesis and the Subject hypothesis.
 - H2.2.3 Experiential Themes that are not of medium length are significant positive predictors of Change.
 - H2.2.4 Place, Reason, Means, Frequency, Purpose, Comitative, Quality and Complement Themes are significant predictors of Change in translations from German to English.
 - H2.2.5 Neither middle animate nor inanimate Subject Themes in combination with sentient verbs are significant predictors of Change.
 - H2.2.6 Non-identifiable Subject Themes are significant predictors of Change.
- H2.3 Translation shifts are more common in translations into English than translations into German.

Theme number has been shown to be one of the best and most consistent predictors of Theme Change in German to English translations. In part, this result is unsurprising. If in the first experiential element hypothesis or the Subject hypothesis the German Theme contains more than two elements, it is almost inevitable that some kind of change occurs in the English translation. This is again due to the Finite-second constraint in German. Multiple Theme elements are usually not possible in German unless one of these elements can occupy the left outfield, like a conjunction or Vocative (see Section 3.3). In all other cases, only one early element can be positioned in the Forefield and the other(s) need(s) to be moved to the Mid-

field. If that element is the first experiential element or the Subject, the Theme number automatically increases by one, because the Finite, which is in second position, is an additional Theme element. In translations into English, the Finite Theme cannot be preserved since this would go against English grammar rules. Instead, most translations simply move the Finite to the position after the Subject, resulting in a reduction of Theme elements (see example (259)). Hence, 95.9% of Themes with three or more Theme elements in GO are decreased in Theme number in ET. If the translator wants to preserve the original number of Themes, they need to add a new textual or interpersonal Theme, resulting in a change of non-experiential Themes. This is why, apart from very few exceptions, every German Theme with three or more Theme elements necessarily has to undergo at least one of these changes. These results therefore confirm H_{2.2.2}.

(259)

GO: *Zu lange | wurden | solche Verträge zu Lasten Dritter abgeschlossen [...].*

'For too.long were such agreements at.the expense of.others made [...].'

ET: *For too long, | such agreements have been concluded at the expense of others [...].*

[G2E_SPEECH_007]

This explains the strong positive effect in the first experiential element hypothesis and the Subject hypothesis. However, it does not explain why Theme number is also a significant variable in the Forefield hypothesis. By definition, the Finite cannot be one of the German Theme elements in the Forefield hypothesis, which is why it is not possible for the Finite Theme to cause a decrease of Theme number in ET. Nevertheless, the higher the number of Theme elements in the Forefield hypothesis, the more likely a Theme change occurs. There are only few cases in which more than one element can occupy the Forefield position in German, the most frequent being the aforementioned conjunction. Surprisingly, it is exactly these cases which are changed in the English translations. 37.8% of Themes with two elements in the Forefield undergo a change in Theme number, typically because one is dropped as a Theme element. For instance, in example (260), the textual Theme is abandoned in the translation in favor of a simple Subject Theme. It is also not uncommon that neither of the two original Theme elements are preserved as Theme in English, especially if one of the German Themes is a marked experiential Theme (see example (261)).

(260)

GO: *Aber | das ist doch kein Grund, ein Kind zu schlagen!*
'but | that is still no reason a child to strike!'

ET: *That's no reason to strike a child!*

[G2E_FICTION_006]

(261)

GO: *Meine Damen und Herren, | heute stehe ich sicher das letzte Mal vor ihnen [...].*
'my ladies and gentlemen today stand I probably for.the last time before you [...].'

ET: *This is probably my last appearance in the Bundestag [...].*

[G2E_SPEECH_008]

H2.2.1 hypothesized that higher Theme number would be a reliable negative predictor of Theme Change in the Forefield hypothesis. This hypothesis was formulated assuming that a single textual or interpersonal Theme in German needs to be changed in some way in the English translations. While such cases clearly exist, they do not outweigh the frequent changes to Themes with more than one element. Therefore, not only does the data not confirm this hypothesis, the exact opposite seems to be the case. High Theme numbers undergo change significantly more often than low Theme numbers even in the Forefield Hypothesis. To some extent, this unexpected result can be attributed to an error in reasoning during the design of the hypotheses. A German Theme that is lacking experiential material needs to be addressed by the English translators, typically by adding an additional Theme element. However, Themes that lack an experiential element are not restricted to simple Themes in the German Forefield. A multiple Theme can easily consist of a textual and interpersonal or even two textual elements. Thus, the Finite-second constraint should affect simple and multiple Themes equally in German to English translations. Therefore, a more accurate hypothesis should have stated that Theme number is not a good predictor of Change in the Forefield hypothesis.

And yet, as it turns out, high Theme number is a significant positive predictor variable. Since most multiple Themes in the German Forefield hypothesis consist of a textual and an experiential Theme, the only possible explanation for this significant effect is that the textual elements are moved out of the Theme in the English translations. At closer inspection, this is

precisely the reason for this significant effect as 36.5% of textual Themes are not translated as such in the English translations. This is unexpected since no contrastive difference regarding the use of non-experiential Themes between EO and GO could be detected (see Section 9.1.1). Also, this omission of textual elements suggests that the textual ties within the text are expressed more implicitly in the English translations, which goes against the common translation feature of explicitation.

The use of textual and interpersonal Themes has led to mixed effects in the regression models with some estimates being positive and others negative. In the three cases where p-values are significant, the estimates are all negative, which suggests that the inclusion of non-experiential Themes in GO make changes in the translations less probable. This is a very surprising result and, in my opinion, does not accurately reflect the relationship between the presence of non-experiential Themes and the likelihood of change in German to English translations. If the German original includes a textual Theme, for example, it is 24.5% more likely to be changed in the first experiential element hypothesis than Themes without one. And yet, for some reason, its reported estimate is not only negative but even significant in the binary model. I believe this is due to the relationship between the variables Theme number and Textual/Interpersonal Themes. In the Forefield hypothesis and the First Experiential Element Hypotheses, all multiple Themes must include either a textual or interpersonal Theme or both, so the three variables largely predict the same effect. In Section 11.2, this relationship between the three variables was already argued to explain the statistically insignificant results of Theme number in English to German translations. Also, the function *vif.mer* (Frank 2014) again reports elevated levels for these three variables, which points towards multi-collinearity between them. If the predictor variable Theme number is excluded from the regression models, the results for textual and interpersonal Themes not only become positive in all models but also significant in all but one. I decided to report the results as they are to allow the reader their own interpretation. However, a further discussion of the supposedly negative effect of textual and interpersonal Themes is in my view unnecessary.

Another consistent Change predictor is Theme length. It was originally assumed that medium-sized experiential Themes are among the least likely to be changed in translations. Short Themes, which are often pronouns, were argued to increase in length because the translator wants to make the reference more explicit. Very long Themes were hypothesized

to decrease in length for reasons of simplification or processing. Apparently, none of these assumptions are accurate. Medium Themes are consistently the most likely Themes to be changed. All other length categories have a negative effect on Change and are almost always significant.

In the majority of cases Themes of medium length become long Themes in the English translations. This is a common effect in all registers except for FICTION. However, at closer inspection most of these longer Themes in the English translations are just literal translations of the original Theme without any noticeable deviation in form or function (see example (262); Subjects in bold). They just happen to be long enough to fall in the long category in English while still being part of the medium category in German.

(262)

GO: ***Der Rahmen um das PIP-Bild** ist bei Standard-PIP zunächst grün.*

'the frame round the PIP-picture is in standard-PIP initially green.'

ET: ***The frame round the PIP picture** is initially green in standard PIP.*

[G2E_INSTR_009]

Potentially, GO uses terminology that is longer and less common in English, which is why their direct translations are above average in size. This interpretation is supported by the fact that many of the Themes that changed from medium to long are only slightly above the threshold of long Subjects. The global average of a long Theme in ET is 31.4 characters, but the average of long Themes that were changed from a medium Theme is only 25.7 characters, which is closer to the 19-character minimum.

At this point, it is important to re-iterate that the length categories that were used to analyze ET are based on the length of experiential Themes in EO. Hence, any general, contrastive length difference between English and German should already be accounted for in these calculations. This result is also the exact opposite of the findings in English to German translations, where long Themes often turn into medium Themes even though they represent literal translations. This suggests that ET require more characters to express the same meanings, while GT require fewer characters and that this effect is not tied to any contrastive spelling differences. A more thorough investigation of this effect is needed.

There is one other category in which medium Themes are changed at an above-average rate and that is Theme type. Many medium experiential Themes are non-Subject Themes, which are reliable predictors of change, as will be discussed below. Short Themes, on the other hand, are almost exclusively Subject Themes (with the exception of some deictic circumstances like *hier* and *jetzt*). 43.2% of Subjects are short in length, as opposed to only 12.2% of circumstances. Accordingly, at least to some extent, it is not the medium length itself that is responsible for the high frequency of change but rather the fact that circumstance Themes, which are changed often in ET, often happen to be of medium length. In example (263), the circumstance Theme, which is of medium length, (in bold) is omitted in the translation in favor of a Subject Theme. Nevertheless, H2.2.3 is falsified.

(263)

GO: ***In Deutschland*** *haben wir der BaFin praktisch die Aufsicht über unseren gesamten Finanzsektor anvertraut.*

'in Germany *have we the BaFin basically the supervision over our total financial sector entrusted.'*

ET: *The task of supervising virtually the entire financial sector has been assigned in to the BaFin, the Federal Financial Supervisory Authority.*

[G2E_SPEECH_002]

With regard to marked Themes, a variety of circumstance and Complement Themes were reported as good predictors of positive Change in German to English translations. However, as was hypothesized originally, it is not marked Themes generally that undergo frequent change, but it depends on the kind of meaning they express, especially in the case of circumstance Themes. One of the most common Theme types to be changed in translation into English are Complement Themes. This is hardly surprising given the fact that Complement Themes in GO outweigh Complement Themes in EO more than eight to one. This finding is also consistent with previous studies that also worked with the CroCo Corpus (for example Neumann 2014; Freiwald 2016; Niemietz, Neumann, and Freiwald 2017). In fact, 95.9% of Complement Themes are changed in ET and the majority of these changes involve a shift in Theme type, which is almost always a shift to a Subject Theme. Surprisingly few cases involve a change in voice, which would allow the participant role to stay in Theme position as the

Subject of a passive clause (see example (264); all Complements in bold, all Subjects underlined). In the vast majority of cases the Complement is moved to a position behind the verb and the original Subject is made the experiential Theme in the English translation (see example (265)). The few cases in which an original Complement Theme stays a Complement Theme in the translations typically involve Attribute Complements in relational processes (see example (266)). Complement Themes in material or mental processes that remain thematic in the translation are rare, though they do exist (see example (267)). These translation procedures largely mirror the results in Freiwald (2016), where Complement Themes were also significant predictors of Change and most often changed by moving them out of the Theme and into the Rheme.

(264)

GO: **Ausgiebige Spaziergänge** unternehmen Ausflügler im Wald rund um das Benediktinerkloster Stift Neuburg im Stadtteil Ziegelhausen.
'*extensive strolls undertake trippers in.the forest around the Benedictine.cloister Stift Neuburg in.the district Ziegelhause.*'

ET: Extensive strolls can be made in the forest around the Neuburg Benedictine Cloister in the Ziegelhausen section of the city.

[G2E_TOU_005]

(265)

GO: **Kleine Mengen** können Sie kurz in der Pfanne anbraten und im Mikrowellengerät fertig-garen.
'*small amounts can you briefly in the pan sear and in the microwave finish.*'

ET: You can sear **small quantities** briefly in the frying pan and finish cooking them in the microwave appliance.

[G2E_INSTR_006]

(266)

GO: **Nicht so klein, aber noch älter** ist der Gunkel [...].
'*not so small but even older is the Gunkel [...].*'

ET: **Not all that small but even older** is Gunkel [...].

[G2E_TOU_022]

(267)

GO: *Das Buch, das ich so krampfhaft am Leibe führe, habe ich erst heute morgen bei Arno's Antik, einem Trödellden am Salzhafen, erworben.*

'the book that I so tightly to.the body press have I only this morning at Arno's Antiques a junk.shop at.the Salzhafen purchased.'

ET: *The book I'm hugging so tightly I acquired only this morning at Arno's Antiques, a junk shop beside the Salzhafen.*

[G2E_FICTION_004]

The second most reliable predictor of Change is Quality Themes. Quality Themes are not particularly common in GO. However, they were still hypothesized to undergo change frequently given the remarkable discrepancy in thematic potential of Quality Themes between EO and GO (see Section 9.1.3). So, while Quality Themes in German are atypical Themes in comparison to all other circumstance types, they are even more marked in English, outmatched only by Complement and Predicator Themes. It thus comes as no surprise that 93.5% of all German Quality Themes are changed in translations. In his regression analysis of marked Themes in German-English translations, Freiwald (2016: 62) also found that Quality Themes are significant Change predictors in the popular-scientific register. For the most part, such a Quality Theme is simply moved to a Rheme position in English and the original Subject takes its place instead (see example (268); circumstance of Quality in bold), which was also the most common translation procedure reported in Freiwald (2016: 64). In total, only five Quality Themes stay Quality Themes in all four sub-corpora. This is clearly an effect of normalization: Quality Themes are so marked in the English system that the translators felt obliged to change the original in favor of a more authentic target text.

(268)

GO: *Geschwind hob sie ihn auf [...].*

'quickly picked she it up [...].'

ET: *She quickly picked it up [...].*

[G2E_FICTION_008]

A variety of other marked Theme types have positive estimates, but their results are only inconsistently significant. These types include Duration, Guise, Matter, Means, Place, Behalf, Comitative, and Predicator Themes. Let me address Predicator Themes first by saying that

their inconsistent results are misleading. Predicator Themes could only be analyzed in four of the six regression models. Their results in the binary models are not statistically significant, which is due to their large standard errors (see Appendix). This stems from the fact that every single one of the 17 Predicator Themes in German was changed in ET. Predicator Themes are a strong predictor of Change in translations into English, likely the strongest of them all. None of the Predicator Themes remain Predicator Themes in the target text but are usually shifted to Subject Themes. In most cases, the Predicator Themes also remain the process in the translation, just not in Theme position. Yet, in some cases some aspects of the process can also become the new Subject if it is lexico-grammatically possible like in (269), where the verb *übernachten* (*staying the night*) is turned into the noun *accommodations*, which serves as the Subject of the translations.

(269)

GO: *Übernachten kann man in Bremen-Vegesack auf dem Schulschiff "Deutschland", [...].*

'stay.the.night can you in Bremen-Vegesack on the research.vessel "Deutschland", [...].'

ET: *Accommodation is available in Bremen-Vegesack on the research vessel "Deutschland", [...].*

[G2E_TOU_009]

After assessing the contrastive differences regarding thematic potential in Section 9.2, eight types of marked Themes were identified as potential challenges in the revised version of H2.2.4: Place, Reason, Means, Frequency, Purpose, and Comitative Themes as well as Quality and Complement Themes, which have already been discussed above. Six of these eight types have at least partially produced significant results. However, despite their differences in thematic potential, Frequency and Purpose Themes do not undergo translation shifts at a significantly different frequency than Subject Themes. Moreover, three further circumstances that were not included in H2.2.4 are significant predictors of Change: Behalf, Matter, and Guise.

Regarding Purpose Themes, it is noticeable that many are realized as hypotactic clauses introduced by the conjunction *um* (*to*). In fact, 43.9% of Purpose Themes are realized as hypotactic clauses in GO. In EO, the number of Purpose Themes realized as clauses is even higher at 95.2%. It appears that circumstances of Purpose are more likely to be thematic in English if they come in the form of a clause. What is also noticeable is that the few cases of

Purpose Themes that are changed in the English translations are primarily realized as prepositional phrases in the original, with only 21.4% of changed Purpose Themes being hypotactic clauses. Unfortunately, Theme form was not analyzed in this study, but these results clearly suggest that form has an effect on the likelihood of a translation shift. The majority of circumstance Theme in the form of clauses are left intact, while circumstance Themes in the form of phrases are changed repeatedly. This may be one of the reasons why English translators do not feel pressured to remove Purpose Themes in the translations, even if they are slightly over-represented. However, a complete analysis of Theme types and forms would be necessary to be certain of this explanation.

In the case of Frequency, no such pattern can be observed, as all Frequency Themes in German are adverbs or prepositional phrases. 61.1% of Frequency Themes stay Frequency Themes in the translation. In all cases where the Frequency Theme is changed, it is moved to a rhematic position in English. Neither the cases that are changed nor those that remain unchanged appear in any way noteworthy. Consequently, despite the difference in thematic potential, English translators do not seem to be challenged by Frequency Themes in German. In the majority of cases, they decide to retain the thematic structure of the original, which is a clear sign of shining-through.

Means, Place, Comitative, and Reason Themes were hypothesized to be significant predictors of Change due to their contrastive difference regarding thematic potential and this prediction turned out to be accurate for the most part. Means Themes have a generally high rate of change at 80.7%, which is very consistent across all four registers. Means Themes in German are mostly realized as groups and phrases. However, the few cases where circumstances of Means are realized as clauses always remain unchanged in the English translations. There are three common translation procedures that the translators use to resolve a marked Means Theme. By far the most common translation involves a movement of the Means Adjunct to the back of the clause (see example (270); all circumstances of Means in bold). Another less common shift also involves moving the Means out of the Theme but realizing it as a different clause element, for instance a Complement (see example (271)). Lastly, translators also sometimes keep the meaning of Means in Theme position but turn it into the Subject of the

target clause (see example (272)). This last procedure is particularly common in INSTR, where the original Subject is often semantically empty.⁹³

(270)

GO: ***Durch längeres Drücken der roten Funktionstaste im Normalbetrieb** können Sie das CI-Modul manuell initialisieren.*

'through longer pressing the red function.key in.the normal.mode can you the CI-module manually initialize.'

ET: *You can initialise the CI module manually in normal operation **by holding down the red function key for some time.***

[G2E_INSTR_003]

(271)

GO: ***Mit konkreten Schritten** kommen wir diesem Ziel näher [...].*

'with concrete steps move we to.this goal closer [...].'

ET: *We have taken concrete steps to move us closer to achieving this goal [...].*

[G2E_SPEECH_001]

(272)

GO: ***Mit dieser Funktion** wird es ermöglicht, Programminformationen des eingeschalteten Programms anzusehen.*

'with this function is it made.possible programme.information of.the current programme view.'

ET: *This function allows you to view programme information related to the current programme.*

[G2E_INSTR_003]

The Comitative circumstance is a very heterogeneous category in the analyses. As was already mentioned in Section 6.3, only very few Comitative Themes in the data involve a person with whom the process is carried out together. Instead, most Comitatives involve a material component or a concept, which explain the process or with which the process is made possible. Most prepositional phrases were analyzed as Comitatives if they began with *mit/with* unless they could clearly be categorized as Means or Reason. A discussion of Comitatives is therefore very challenging because the different kinds of meanings are so diverse.

⁹³ There is also a fourth option, which includes simply omitting the marked Theme, which was already discussed in Chapter 10.

That being said, the most common translation procedure of dealing with Comitatives is still to keep it formally and functionally intact, but to move it out of the Theme in the translation (see example (273); all circumstances of Comitative in bold). A re-mapping onto a different clause element in the target clause occasionally occurs as well but is generally very rare (see example (274)).

(273)

GO: **Mit Terroristen wie Osama bin Laden** werden wir nicht verhandeln können.
'**with terrorists like Osama bin Laden** will we not negotiate be.able.'

ET: We will not be able to negotiate **with terrorists like Osama bin Laden**.

[G2E_SPEECH_003]

(274)

GO: **Mit 18 Liften und 14 Hütten** können sportliche Genießer hier ganz nach Wunsch zwischen Abfahrt und Einkehr wechseln.
'**with 18 lifts and 14 huts** can athletic connoisseurs here totally as desired between going.downhill and coming.back switch.'

ET: The 18 lifts and 14 huts allow athletic connoisseurs to change as they like between skiing and relaxing.

[G2E_TOU_005]

Some of the distinctive features of circumstances of Place in German to English translations were already discussed in the analysis in Sections 10.1.3 and 10.2. It thus comes as no surprise that the use of a Place Theme turns out to be a significant predictor of Change in the inferential analyses. What is so surprising about Place is that their thematic potential is almost identical in GO and ET. The reason why so many Place Themes are nevertheless not translated as such is not due to thematic potential but to the fact that GO includes a much higher number of Theme circumstances generally, regardless of position. In Chapter 10, it was shown that not all of these missing circumstances of Place just disappear in the translations (although that does happen as well), but that most are realized as other clause elements. As was the case with previous marked Themes, another productive translation procedure of dealing with Place Themes is to move it out of the Theme (see example (275); circumstances of Place in bold). However, this shift is noticeably less common in the case of Place Themes.

(275)

GO: *Vor dem Fenster fällt Schnee.*

'in.front.of the window falls snow.'

ET: *Snow was falling in front of the window.*

[G2E_FICTION_007]

It appears that English translators are very creative when it comes to the translation of Place and they try to keep a balance between adhering to the source language as well as the target language norms. On the one hand, they adopt the thematic potential of circumstances of Place of the source language, which results in a higher number of Place Themes in ET compared to EO. On the other hand, they follow the general frequencies of Place of the target language and translate many of the original Place Themes as different clause elements. This results in a lower number of circumstances of Place in ET compared to GO. The translation of Place Themes in English may be one of the best examples of normalization and shining-through effects in combination.

Lastly, Reason Themes also produced significant results but only in a single regression model; this result should therefore not be over-interpreted. The translation of Reason Themes is very similar to that of Purpose Themes, which makes sense as they are very close in meaning. Reason Themes in German are often realized as hypotactic clauses introduced by the subordinators *weil* (*because*) and *da* (*because, since*). However, 86.6% of Reason Themes that are changed are not clauses but groups or phrases. In other words, if the German Reason Theme has the form of a hypotactic clause, it is much less likely to be changed. In comparison to Purpose, the number of Reason Themes as clause is considerably lower, which explains why the effect of Reason on Change was greater than the effect of Purpose.

Three types of circumstance Themes have positive significant effects even though they were not included in the hypothesized group of marked circumstances: Behalf, Matter, and Guise. Behalf was considered to be included in H_{2.2.4} because there is a noticeable difference in its thematic potential between EO and GO. However, this consideration was ultimately dismissed because, despite this difference, the thematic potential of Behalf in English is still fairly high. For this reason, it was assumed that translators can easily keep most Behalf Themes in the same position in English even if that means that they are marginally

overrepresented. This assumption did not turn out to be the case. 66.7% of Behalf Themes are changed in some respect; most involve a shift in Theme type from circumstance to Subject Theme. In these cases, the Behalf Theme is almost always moved to a later position in the clause (see example (276); all circumstances of Behalf in bold). Using (parts of) the Behalf Themes as the new Subject is rare but does occur in clauses like (277). This suggests that differences in thematic potential do matter in translations even if the potential is generally high in both languages.

(276)

GO: **Für die Schulkinder** war es noch zu früh, kaum einer verliess so früh das Lager.

'**for the schoolchildren** was it still too early, hardly anyone left this early the camp.'

ET: It was too early still **for the schoolchildren**, hardly anyone left the camp this early.

[G2E_FICTION_002]

(277)

GO: Doch **auch für Kinder** ist in Deutschland immer etwas los!

'but **also for children** is in Germany always something happening!'

ET: Something for children is always happening in Germany!

[G2E_TOU_005]

The significant results of Matter and Guise are completely unexpected. There are differences in thematic potential regarding these two circumstance types, but it is EO that has the higher thematic potential for both of these circumstance types. In the case of circumstances of Guise, this difference is substantial, which was used as the primary reason to explain their significant results in English to German translations in Section 11.2. Based on these results, it was unlikely for Matter and Guise to be a problem in translation into English and yet they are changed significantly more often than Subject Themes.

Freiwald (2016: 62) also reports Matter Themes as a significant predictor of Change in translations from German to English. In the popular-scientific register, the most common translation procedure involves turning the Matter circumstance into the Subject of the clause, especially if the German original only includes a semantically light or empty Subject. This appears to be a register-specific procedure since none of the examples in this data set exemplify a similar translation procedure.

Two interesting differences can be noticed between English and German regarding circumstances of Matter. Firstly, while their thematic potential is almost identical in the two languages, the overall frequency of Matter Themes is much higher in German. Hence, the frequent changes to Matter Themes may be one translation procedure to adjust their frequency to that of the target language. Secondly, the way in which Matter Themes are used in the two languages is very different. If a clause opens with a Matter circumstance in English originals, the Matter is typically more detached from the rest of the clause. Matter Themes in English arguably function similarly to Comment Adjuncts by reminding the hearer in what context the following clause needs to be understood (see example (278); all circumstances of Matter in bold). In German, Matter Themes can also be used in this way. However, it is much more likely to find Matter Themes that are more integrated in the independent clause (see example (279)). Such Matter Themes in German could be argued to be post-modifiers of a Complement or a Subject, which were detached and moved to the front of the clause. Unsurprisingly, most of these kinds of Matter Themes are resolved in the English translations by moving them behind the element that they are connected to. Accordingly, while these technically belong to the same circumstance category, they are clearly different in both form and function. This explains the high number of changes in the English translations.

(278) **When it comes to trade**, enlargement has had a more mixed impact on the United States' relationship with accession countries, in large measure due to the extremely lengthy accession period.

[E2G_SPEECH_014]

(279)

GO: **Über das anschliessende Verfahren zur Ratifizierung in den Mitgliedstaaten** hat europaweit eine Diskussion eingesetzt.

'**about the subsequent procedure for the ratification in the member states** has Europe-wide a discussion begun.'

ET: A Europe-wide discussion on the subsequent ratification procedure in the member states has begun.

[G2E_SPEECH_014]

A similar effect can be observed in the case of Guise Themes. Circumstances of Guise represent yet another category that included a variety of heterogeneous types of meanings. In

English, Guise mainly serve as a category to label detached predicatives (Biber et al. 1999: 136), which are fairly common in English, but not so much in German. 67.9% of Guise Themes in EO can be analyzed as this kind of construction, often in the form of a non-finite clause. In GO, there are only three cases of Guise Themes in total that are clauses, which incidentally all remain Guise Themes in the translations. The vast majority of Guise Themes in German are prepositional phrases introduced by *als* (*as*), which are sometimes kept in Theme position and sometimes moved to the back of the clause in the translations (see example (280); circumstances of Guise in bold). Again, the circumstance category may be the same in both of these cases but form and function are still very different. This shows how difficult it is to simply transfer the categories of one language to another.

(280)

GO: ***Als internationale Kongressdestination*** rangiert es unter den ersten vier Städten weltweit.

'as international conference.destination ranks it among the first four cities worldwide.'

ET: *Vienna also ranks among the first four cities in the world **as an international conference destination**.*

[G2E_TOU_018]

Finally, three non-Subject Themes are generally less likely to be changed compared to Subject Themes, which are Condition, Concession, and Time Themes. Even though no extra hypothesis was formulated, this result was to be expected. These three circumstances are the only circumstance types that have a thematic potential of over 50% in EO, which means that they are on average more likely than the Subject to be used as the Theme of the clauses they occur in. Thus, as far as frequency is concerned, the Rheme is a more marked position for these three circumstances than the Theme. For this reason, it can be expected that translators do not feel the need to frequently change their thematic status if there is no contrastive reason for them to do so. Since there are a number of marked Themes that are changed regularly, translators may be even more inclined to not change Condition, Concession, and Times Themes in order not to deviate from the source text even more than they already have. It needs to be pointed out that the negative effects of these three Themes are only significant in one of the five regression models each, which is why an indisputable deviation from Subject Themes cannot be attested here. Nevertheless, the results do clearly show that these

three circumstance Themes are not more marked than Subject Themes in German to English translations.

Regarding H_{2.2.4}, the regression models only partially confirm the hypothesis. Two of the eight marked Theme types that were hypothesized in H_{2.2.4} to be reliable predictors do not deviate significantly from Subject Themes. Of the remaining six types, many are not consistently significant. Additionally, three types of marked Themes have a significant positive effect on Change even though they were not included in H_{2.2.4}. That being said, I would still argue that the hypothesis is mostly accurate and that thematic potential in general is a very reliable predictor of translation shifts. The high frequency of shifts concerning Matter and Guise were explained by formal and functional differences, which the model of thematic potential could not factor in based on the analyses that were done. Behalf Themes were considered as a potential Change predictor because of the TPot difference between the two languages but were ultimately dismissed. And while inconsistent significant effects are difficult to interpret, the estimates for all hypothesized Theme types are positive and often just shy of the necessary p-values. Lastly, all circumstance types that have a TPot of over 50% in English have negative effects on Change, which are occasionally significant. Obviously, thematic potential is not the sole predictor of Theme Change, but it has clearly proven to be a useful tool in predicting translation shifts in German to English translations.

These results in marked Themes are largely consistent with the findings of Freiwald (2016). He found that in translations into English four non-Subject Themes had significant, positive effects on Change: Matter, Quality, Viewpoint, and Complement Themes (Freiwald 2016: 62). Three of these marked Themes also lead to significant results in this study. Viewpoint circumstances were very common in the popular scientific register. However, in the four registers included in this study they were only used nine times in total, and only four times in Theme position, which is why they did not even enter in the regression model here. Six further circumstance Themes were significant in this study but not in Freiwald (2016), which is largely due to limited data since four of these six circumstances were not frequent enough to be analyzed in Freiwald (2016). This comparison shows that English Theme markedness is fairly consistent across registers.

In the previous regressions on English to German translations, Carrier, the reference category for Participant Themes, was among the most likely Participants to be changed in translations. In the opposite translation direction, they rank more in the middle of the markedness scale and do not deviate significantly from most other participant roles. Five Participant Themes do, however, differ from Carrier Themes and show occasionally significant results: Empty Subject, Goal, Value, Initiator, and Phenomenon Themes. The significant results for Goal and Phenomenon Themes are easily explained. In theory, no participant role is tied to any particular clause element. However, in practice, first participants typically map onto Subjects and second participants onto Complements. There are many Goal and Phenomenon Themes in the German originals which are Subjects in *MOOD*, which means that the verb is in passive voice. However, a substantial amount of Goal and Phenomenon Themes are also Complement Themes. It was already shown above that Complement Themes, especially ones that are not Attributes, have a high likelihood of being changed in English translations.

Empty Subject Themes refer to grammatical Subjects that are not tied to a participant role in *TRANSITIVITY*. Empty Subject Themes have a positive effect on Change in four models and a negative effect in two models in German to English translations. All significant results are tied to positive estimates, so empty Subject Themes do seem to be a good predictor of change. Semantically empty Subjects exist in both English and German. However, there are many German constructions that include such an empty Subject, which do not exist in English. These include for instance expletive *es*-Subjects that only act as a placeholder in German if a Subject is missing (Engel 2004: 164). Naturally, clauses like (281) require some kind of change in the English translations (Subjects in bold). But even in those cases where an English equivalent does exist, translators often opt for a semantically heavier Subject in the target clause (see example (282); Subjects in bold).

(281)

GO: **Es** locken kostbare, glitzernde Kleinigkeiten, wienerische Kunstwerke aus Glas und Porzellan, Kleidung in wienerisch-elegantem Design oder im internationalen Trend und die berühmten, auf der Zunge schmelzenden Verführungen aus Schokolade.

it lures priceless glittering objects, Viennese masterpieces out.of glass and porcelain, clothes in Viennese-elegant design or in.the international trend and the famous, on the tongue melting temptation out.of chocolate.'

ET: **One** can choose between priceless glittering objects, Viennese masterpieces made from glass and porcelain, clothes of timeless Viennese elegance or of international trend and the famous seductive chocolates whose tastes linger for a long time.

[G2E_TOU_022]

(282)

GO: **Es** geht um Ihre Ideen, Ihren Einsatz.

it is about your ideas, your commitment.'

ET: **We** need your ideas, your commitment.

[G2E_SPEECH_005]

A change in participant role is the most common change involving empty Subject Themes, followed by a change in Subject identifiability and sentience. Since empty *es* Subjects cannot be analyzed in terms of their identifiability or sentience, it is unsurprising that a change in participant role is often accompanied by changes in these two categories as well. This also explains why the category has higher estimates in the numerical models, as semantically empty Subjects often undergo a multitude of changes.

Initiator Themes are reported to have a positive impact on Change, but it is not the participant role itself that is changed frequently. In fact, Initiator Themes have a below average change rate in the category of participant roles. What is changed fairly frequently with Initiators is their identifiability. Many of the Initiators in GO refer to forces of nature, objects, and even abstract concepts, which happen to frequently include an indefinite or zero article. In the English translations, these Subjects remain in Theme position, but their article is repeatedly changed to a definite article, thus turning them from non-identifiable to identifiable Subject Themes (see example (283); Subjects in bold).

(283)

GO: ***Haus und Garten** lassen die Schaffenskraft des Expressionisten Emil Nolde lebendig werden.*

*'**house and garden** let the creative.power.of.the expressionist Emil Nolde to.live come.'*

ET: ***The house and garden, of the artist Emil Nolde**, bring his expressionism to life.*

[G2E_TOU_010]

In Section 10.1.5, it was shown that ET are generally less identifiable than the German originals, so this effect goes against this principle. In truth, I believe this effect to be mere coincidence despite the significant results in two regression models. This belief is supported by the fact that Initiators do not deviate from Carriers significantly in the Subject hypothesis, which takes into consideration considerably more instances of Initiator Themes.

Value Themes, on the other hand, are changed frequently in terms of their participant role. By far the most common type of change involves a re-ordering of Token and Value (see example (284); Token in bold, Value underlined). In identifying relational processes, Token and Value can come in any order but it is generally more common in both English and German to find the Token, the more concrete participant, preceding the more abstract Value. English translations often re-establish this unmarked order.

(284)

GO: *Das erste sind **seine Augen** [...].*

*'the first.thing are **his eyes** [...].'*

ET: ***His eyes** are the first thing [...].*

[G2E_FICTION_005]

Lastly, there is one Participant Theme which deviates from Carrier Themes negatively, namely Senser Themes. Senser Themes are significantly less likely to undergo a change in two models. This result is surprising due to the fact that German can realize Sensers as Complements in active clauses, which is an unmarked construction in German (Steiner and Teich 2004: 155) but does not exist in English. However, despite the fact that German Complement Sensers are considered unmarked by Steiner and Teich (2004), they are fairly rare in the data (4.3% of mental processes) and are counterbalanced by the much higher number of Subject Sensers. Subject Sensers score below average in global change rate in German to

English translations. Mental processes are mostly found in FICTION and SPEECH and often reveal important information about the internal world of the character or speaker. It seems that translators try to avoid a change of this process type to not lose this valuable insight.

Subject Theme identifiability is generally not a good predictor of Theme Change in German to English translations. It was hypothesized that non-identifiable Subjects are changed significantly more often than identifiable Subjects because of explicitation. While its effect on Change is consistently positive and also significant in one regression, this is not enough evidence to support H_{2.2.6}. The translation of Subject Themes whose identifiability cannot be analyzed deviates significantly from the translation of identifiable Subject Themes. However, this group of Subjects is very dissimilar and was already partially discussed in different contexts (for example semantically empty *es* Subjects).

Similarly, Subject Theme sentience is also not a very reliable category to project Theme changes in German to English translations. Since the semantic mappings onto clause elements is generally freer in English than in German (Hawkins 1986: 67), different kinds of inanimate Subjects were not expected to pose much of a challenge for English translators. This expectation turned out to be accurate for the most part. However, two Subject types do deviate significantly from Human Subject: nonconcrete Inanimates and Place Subjects in both sentient and non-sentient constructions. The fact that both of these Subject types produce significant results regardless of the semantic restrictions of the verb already suggests that the reason for their higher rate of change is not one of semantic mismatches between Subject and verb.

The most likely change of nonconcrete Inanimates is a change in Subject sentience, where the nonconcrete Inanimate Subject is turned into a Subject whose sentience was not analyzed. As a reminder, to save time during the annotations, some constructions, where animacy of the Subject or the semantic requirements of the verb were deemed irrelevant, were not analyzed in terms of Subject sentience. This includes, for example, relational processes and passive clauses. Incidentally, these are also the most likely kinds of changes that involve nonconcrete Inanimates. Many of these inanimate Subject Themes are used as Actors in the German originals, which are changed to Subject Carriers in the translations, regardless of whether the process requires sentience or agency (see example (285)). Another surprisingly

common type of change turns the nonconcrete Inanimate Actor into a Goal Subject in a passive clause, which is why the Subject Sentience category changes as well (see example (286)). Hence, most nonconcrete Inanimate Subjects keep their semantic meaning in the English translations despite the high change rate of the Subject Theme sentence.

(285)

GO: *Ein Hauch Ritterromantik umgibt noch immer die vielen Burgruinen [...].*
'a touch knight.romanticism surrounds still the many fortress.ruins [...].'

ET: *The romantic notion of knights in shining armour is alive today in the many fortress ruins [...].*

[G2E_TOU_012]

(286)

GO: *Falls das Fernsehgerät nur über eine Antennenbuchse verfügt, so erfolgt die Wiedergabe über das angeschlossene Antennenkabel.*
'if the television only an antenna.socket has, then takes.place the playback via the connected antenna.cable.'

ET: *For TVs with an antenna socket, playback is performed via the connected antenna cable.*

[G2E_INSTR_008]

The exact same translation procedures can also be observed for Place Subjects in Theme position. Sentience and participant role changes are by far the most common and the majority of Place Subjects that are changed become Carriers or Goals in the English translations. Additionally, Place Subjects show a very high rate of changes to marked Themes, which is due to the fact that surprisingly many German clauses that contain a Place Subject are opened by a Complement. And as was shown before, Complement Themes are one of the strongest predictors of change in ET. It is not immediately apparent why German Place Subjects are so likely to be accompanied by Complement Themes. A common construction in German tourism leaflets, where almost all Place Subjects occur, is to name a touristic attraction in the form of a Complement followed by the verb *bieten* (*offer*) followed by the Place Subject. Such clause constructions may be designed to grab the reader's attention and point them towards the place where they can experience what they are looking for. Given the high markedness of Complement Themes in English, this constituent order can usually not remain unchanged in the English translations and it is typically resolved by either changing the order of Subject

and Complement (see example (287); all Complements in bold, all Subjects underlined) or by turning the original Complement Theme into a Subject Theme (see example (288)).

(287)

GO: ***Fachkompetente Rehabilitation*** bieten *die Spezialkliniken im Ort.*
'***competent rehabilitation*** offers *the special.clinics in.the city.*'

ET: *The specialized clinics right in the city* offer ***competent rehabilitation.***

[G2E_TOU_004]

(288)

GO: ***Spezielle Gourmet-Arrangements*** bieten *die Hotels "Sonne" und "Neptun" [...].*
'***special gourmet-arrangements*** offer *the hotels "Sonne" and "Neptun" [...].*'

ET: *Special gourmet arrangements* are offered by the hotels "Sonne" and "Neptun", [...].

[G2E_TOU_009]

Apart from this last translation procedure, most of the Place Subjects also remain Place Subjects, just like with nonconcrete Inanimates, and it is rather a change in process or participant role distribution that indirectly affects Subject sentience. This tendency to change the process is again largely independent of semantic restrictions of the verb. Based on these results, H_{2.2.5} is technically falsified, though the underlying premise is mostly accurate.

Lastly, the category Register is again telling in terms of Theme changes. All registers deviate positively from the reference variable FICTION, which means that FICTION is the register that generally undergoes the lowest number of Theme changes. SPEECH and TOU have similarly large estimates and are consistently significant, which puts them at the far end of the change spectrum. Given its smaller estimates and inconsistently significant results, INSTR is apparently positioned somewhere in between. In the other translation direction, SPEECH was changed the least out of the four registers and it was INSTR and TOU which were most prone to change. Thus, the relationship between register and rate of Theme change appears to be language dependent.

Unlike in the previous analysis of register effects, there is no one or two contrastive features which clearly explain why one register undergoes more change than the others. The most reliable predictors of Change in German-English translations are Theme heaviness,

Theme number and certain marked Themes. However, none of the registers is uniquely characterized by any of these features. The reality is that FICTION and INSTR score slightly below average and TOU and SPEECH generally score slightly above average in most change categories. In comparison to the opposite translation direction, the frequency with which Theme changes occur is much more similar between the four registers. It also needs to be pointed out that each register has a lot of data points so even the slightest variation between them can already result in statistical significance. These results on FICTION again mirror Neumann's (2014) results, who also noticed that English translations of fictional texts are more similar to the source language norms compared to letters to shareholders.

To summarize, some of the hypotheses are verified by the data, while others are not. For the most part, the translation of non-Subject Themes was predicted accurately. The effects of higher Theme numbers in the Subject and the first experiential element hypothesis are also as expected. On the contrary, Theme numbers in the Forefield hypothesis and Theme length generally revealed surprising translation procedures.

Translation procedures in response to challenging Theme structures are diverse. However, there is a general tendency of English translators to not dissociate the semantic meaning from its clause element. Despite the generally more flexible mapping of semantic meaning onto grammatical function in English, most of the form-meaning pairings in the originals are also preserved in the translations. In the case of marked Themes, this is accomplished by moving Theme elements into the Rheme. In the case of marked Subject Theme sentience, a change in process type is mostly preferred over a re-mapping of the Subject. This tendency of preserving original form-meaning pairings is accompanied by other Theme-related changes, for instance a change in Theme number or experiential Theme length.

As the final, overarching hypothesis, H_{2.3} claims that more translation shifts occur in translations into English compared to translations into German. This hypothesis was formulated, assuming that English translations align more with target language rules, while German translations show more traces of the source language. And in fact, the number of shifts is higher in ET than in GT and the difference is statistically significant in terms of both the general presence of a change to the original Theme (binary Change; $\chi^2 = 14.145$, $df = 1$, $p\text{-value} = 0.0001692$) and the number of thematic changes altogether (numerical Change; $\chi^2 = 11.438$, $df = 1$, $p\text{-value} = 0.0007195$). These results thus confirm H_{2.3} and clearly show that

Theme changes are more common in translations into English. Hasselgård (1997) makes a very similar observation in her analysis on Theme translations between English and Norwegian. She found that the translations generally preserved the Theme structure of the originals but that Norwegian translations adhered to the source language even more so than English translations. Evert and Neumann (2017) analyzed a variety of grammatical measures in English and German original and translated texts and also found that the shining-through effect was generally more observable in German translations. They attribute this finding to differences in prestige between source and target language. It is difficult to indisputably argue that these findings corroborate Evert and Neumann's (2017) interpretation since the contrastive differences of Theme may simply restrict English translations more heavily than in the other translation direction. Nevertheless, the results do, at the very least, support their assumption.

12 Summary of the results

This twelfth chapter serves as a summary of the previous five results chapters. The results that were presented and discussed so far have been numerous, which is why a collection of the most important insights gained seems appropriate at this point. Naturally, the explanations will be rather brief to keep the length of this chapter manageable.

Chapter 7 dealt with the Theme structure analysis in German originals. Three different Theme hypotheses of German were compared, namely the Forefield, the first experiential element, and the Subject hypothesis. This comparison was only made for Theme markedness and Theme number, as these were the only thematic aspects in German that were noticeably affected by the extent of the Theme. Self-evidently, the further the Theme potentially extends into the clause, the higher the average Theme number may get. The frequency of marked Themes also varied noticeably between Theme hypotheses. On the one hand, this was due to clause element positioning: circumstances, for instance, are not commonly placed in the first post-verbal position. Their relative frequency thus decreased between the Forefield hypothesis and the first experiential element hypothesis because the latter did not include many more circumstance Themes. On the other hand, the decrease of marked Themes was also an effect of dilution since the number of additional Subject Themes in the Subject hypothesis, for example, diluted the relative frequency of all other Theme types.

One of the most interesting results that was found during the comparison of different Theme types in German was that circumstance Themes and textual Themes are in an inversely proportional relationship across all four registers that were analyzed. This could simply be a coincidence, but a comparison with the EO subcorpus does not reveal the exact same relationship in English despite the fact that the registers are otherwise quite similar in most thematic aspects. It was argued that the reverse relationship between circumstance Themes and textual Themes rests upon their similar positional restrictions in the German clause. Both circumstance Themes and textual Themes are positioned in the Forefield of the clause if they are to come early. If they are not in the Forefield, they likely occupy a medial or late position in the Midfield. It is very uncommon for circumstances and textual elements to occupy the starting point of the Midfield. In other words, circumstances and textual elements compete for the same position in the clause, in particular the same Theme position, which is

the Forefield. The frequencies of Complement Themes were not in any way systematically related to circumstance or textual Themes since Complements can be positioned in the Forefield as well as at the beginning of the Midfield, especially if they are identifiable. This last observation was interpreted as evidence that the Theme in German is not made up of the Forefield alone. If only the Forefield was thematic in German, the Complement would also have to compete with circumstances and textual elements for this one Theme position. The fact that they do not suggests that the first post-verbal position also has thematic meaning.

At this point I would like to go back to one of the discussion points of the fifth chapter, which is also one of the main questions of the entire thesis: Is the German Theme restricted to the Forefield or can thematic elements also enter the Midfield. Drawing on both the theoretical and the empirical considerations, I am confident in claiming that Theme in German is not congruent with Forefield and that the German Theme can also be located in the Midfield if the first experiential element does not occupy the pre-verbal position. I would therefore argue against the Forefield hypothesis and in favor of the first experiential element hypothesis in German. In total, there are five reasons that lead me to this interpretation:

1. I do not believe a single textual or interpersonal element is sufficient for the hearer to interpret the message, neither in English nor in German. A Theme needs an anchorage to the realm of experience if it truly is to serve as a meaningful point of departure of the message. The assumption that a near-literal translation of an English clause into German is interpreted differently just because the finite verb is one position to the left remains unconvincing to me.
2. On a similar note, to my mind, there is not enough convincing evidence for why Germanic languages are possibly the only language family that do not require an obligatory experiential Theme. The fact that the finite verb has a fairly fixed second position in the clause does not explain satisfactorily why speakers of Germanic languages (apart from English) conceptualize their points of departure fundamentally differently from the rest of the language families.
3. Steiner and Teich's (2004: 172-173) main reason for rejecting the first post-verbal element in German as an element of Theme is that the first post-verbal position is the default place of identifiable referents, which is a question of informational meaning. It is true that

identifiability has an influence on the constituent order in the Midfield, as identifiable referents typically precede non-identifiable ones (Götze and Hess-Lüttich 2002: 485). However, identifiability is not the only factor that governs Midfield sequencing and even with all of these different constraints, different arrangements of one and the same German clause are still possible in the majority of cases. Even if the Midfield contains both an identifiable and a non-identifiable referent, the non-identifiable element can head the Midfield if it is contextually motivated, which was found in the data repeatedly. Besides, it is not unheard of that choices in one system can influence choices in another system, as is the case for MOOD and THEME. It is thus not implausible to assume that INFORMATION simply increases the likelihood of certain Themes over others but does not render thematic choices superfluous. I will say that there are cases where the different constraints on the Midfield are so strong that no positional flexibility remains. This is the case, for example, if the Subject is a first person, nominative pronoun, where no other element can precede the Subject in the Midfield without making the clause ungrammatical. In such circumstances, I will concede that the speaker does not have any thematic choice unless they want to change the clause entirely. However, this is not the only example where only a single clause element is a reasonable choice for Theme. The same holds true for declarative clauses in English and German that contain an intransitive verb and no circumstances. In such a case, only the Subject remains as a potential Theme candidate, which leaves the speaker with no other option either. Nevertheless, the Subject is still considered a meaningful Theme element.

4. Additionally, in practice, the differences in identifiability constraints between Forefield and Midfield have been shown to be rather minor. The majority of German Subject Themes are identifiable, regardless of position. In fact, German Subject Themes have an identical likelihood of being identifiable as English Subject Themes, which shows that there is a similar natural relationship between the point of departure and given information in German as has been discussed for English (see Section 5.1.1). When comparing Subject Themes in the Forefield and Subject Themes that occupy the position immediately after the finite verb, the number of identifiable Themes does increase; but only by 0.9%. In other words, the additional identifiability constraints that affect the early elements in the Midfield and do not affect elements in the Forefield are only observable once per 100

clauses. Given this small difference, I do not find it convincing that these identifiability considerations of the Midfield would disqualify early elements from being interpreted as the point of departure of the message.

5. The final reason is the one outlined above that Complements would have to compete with circumstances and textual elements for thematic status if only the Forefield position had thematic meaning. Apparently, they do not, though, and that is because they can also easily occupy the first post-finite position, which circumstances, and textual elements cannot. If this post-finite position was not thematic, one would expect a similar antiproportional relationship.

The analysis of German circumstances showed that there are some circumstance types that appear to have a natural predisposition to being used as points of departure, namely Condition and Concession. Most other types of circumstances have a relatively similar thematic potential of about 40%, with only few circumstances being much lower than that. Thematic potential is not directly linked to general frequency since circumstances of Concession, for instance, are used relatively rarely, while circumstances of Place, the most frequent circumstance type overall, have only an average TP.

Differences in the frequency of non-sentient constructions were found when comparing middle animate to inanimate Subjects. As expected, Machine Subjects were in between these two groups but behaved more similarly to the middle animate group. Most of the middle animate and inanimate Subject types came from one register in particular. For example, all Animal Subjects are used in FICTION, Organization Subjects are common in SPEECH, most Place Subjects can be found in TOU, and Machine Subjects are largely used in INSTR. Concrete and nonconcrete Inanimates were the only Subject types that were consistently used across the registers, but given their very broad membership criteria, these Subjects varied a lot across the register in the kinds of meanings they expressed.

There was hardly any Theme measure that was not predicted by the variable register. Everything from Theme number and marked Themes to participant Themes, circumstance distributions, Subject Theme identifiability, and non-sentient Theme constructs varied noticeably between the registers. It is thus fair to argue that the thematic space is highly dependent on the register in German.

Of the four registers that were analyzed, FICTION and SPEECH were the most alike in terms of Theme. Both used a similar number of marked Themes and textual and interpersonal Themes. They were also the two most balanced registers regarding process types and circumstance types. FICTION included a considerably higher number of behavioral processes compared to all other registers but apart from that the distributions in FICTION and SPEECH were very similar and close to the overall average. TOU included a high number of circumstance Themes, in particular Place, but only few textual and interpersonal Themes. The distribution between material and relational processes, which were the two most common process types across the board, were most balanced in this register. INSTR was the most noteworthy register in the German original subcorpus. It included by far the highest number of circumstance Themes at around 40% and at the same time the lowest number of non-experiential Themes. INSTR was also the register with the highest number of participant Themes from material processes. Conditions were particularly common in INSTR, where they are used almost four times as often as in any other register.

Chapter 8 dealt with the Theme structure analysis in English originals. Similar to German, three Theme hypotheses of English were compared: The first element, the first experiential element, and the Subject hypothesis. This comparison was again only made for Theme markedness and Theme number, as these were the only thematic aspects that were truly affected by the extent of the Theme. The fact that the statistics for most Theme aspects are influenced so little by the Theme hypothesis means that a Theme analysis like the one in Neumann (2014), based on the first element in the clause, is a viable option in quantitative Theme studies of English and German.

Despite the fact that circumstance Themes are considered generally marked in English (Halliday and Matthiessen 2014: 98), three circumstance types had a thematic potential of 50% or higher, which means they were more likely to be used thematically than rhematically: Condition, Concession, and Time. Unlike in German, the rest of the circumstances did not group around an average TPot but their TPot steadily decreased.

English is said to have only weak semantic constraints on its Subjects to make up for the lack of positional flexibility (Kast 2012: 148). Nevertheless, there is a clear divide between the use of middle animate Subjects with sentient verbs and inanimate Subjects in English as well. Again, Machine Subjects behaved a lot like the other middle animate Subjects. This goes

to show that considerations of animacy, sentience, and agency remain relevant in English even if the restrictions are supposedly looser compared to German.

Many of the same register effects that were found in GO could also be observed in EO. Theme number was largely predicted by the register, and so was the use of marked Themes. The effects of register on process type and participant Theme distributions were also highly significant. The visual representation of circumstance Themes suggested noticeable differences between the four registers, yet the statistical tests did not confirm this interpretation. However, the use of the different Subject Theme types and the affiliated likelihood of non-sentient construction were again tightly linked to register.

Material and relational processes were the two most frequent process types in EO and were distributed similarly in the overall results. Yet, within one register, their differences in use did diverge more noticeably. Material processes were very prominent in INSTR and SPEECH, while relational processes were particularly common in TOU. FICTION was the most evenly distributed register regarding process types. Textual Themes were very common in FICTION and SPEECH, which explains why they also had a higher number of Theme elements in the first experiential element hypothesis. Interestingly enough, the numbers of multiple Themes were comparable between the four registers in the Subject hypothesis, which suggests that the types of elements preceding the Subject in English are different between the registers but that the likelihood of any non-Subject constituent to open up the clause is generally the same. This was interpreted as an effect of processing considerations (Diessel 2005). In terms of marked Themes, the registers were fairly close together in frequency, with FICTION having the lowest number and INSTR the highest number of marked Themes. The distribution of circumstance Theme types was imbalanced between the registers. INSTR was dominated by Condition Themes, while TOU included a high number of Place Themes. The circumstance frequencies in FICTION and SPEECH were most like the overall average with a slight tendency towards time-related circumstance Themes. Just as in German, the Subject Theme types varied considerably between registers and many types came predominantly from one register in particular.

Chapter 9 included the contrastive analysis, where the English and German original corpora were compared in terms of their thematic structures. The different Theme hypotheses

were again distinguished for average Theme number and Theme markedness. As was originally assumed, differences in Theme number between English and German were highly dependent on the choice of Theme hypothesis. In the comparison between the Forefield hypothesis in German and the first experiential element hypothesis in English, Theme numbers were higher in English due to the strong restrictions on the number of elements in the German Forefield. This is in line with Steiner and Teich's (2004: 174) description of multiple Themes in German and also mirror the results in Freiwald (2016). In the three other Theme comparisons, the German Theme number was consistently higher. This was inevitable for the comparison between the Forefield hypothesis in German and the first element hypothesis in English, as the English Theme consisted of just one element by definition. In the first experiential element hypotheses, the higher Theme numbers in German were caused exclusively by the additional, obligatory Finite Themes, which shows that the use of textual and interpersonal Themes (apart from Finite Themes) are very comparable between the two languages. The difference in Theme number grew even larger in the Subject hypothesis because of the higher number of marked Themes in German.

This difference in marked Themes was one of the most pronounced differences in the contrastive Theme analysis. This was particularly evident for Complement Themes, which were almost eight times more frequent in GO than in EO. The differences in circumstance Themes were less pronounced between the two languages but still significantly higher in German.

The use of process types was remarkably similar overall and the few differences that were found were not statistically significant. Even when comparing the same registers, frequency differences were only marginal. The largest differences regarding participant Themes and process types were found in INSTR, which is much more materially-oriented in German and more focused on relational processes in English.

Circumstance Theme types were distributed unevenly. German included a lot more Place Themes, while Time and Condition Themes were more representative of English circumstantial Themes. The thematic potentials of individual circumstance types were different between the languages, with most circumstances having a higher potential in German than in English. Notable exceptions include Time, Comparison, and Guise. That being said, the TPot was fairly similar for most circumstances: Overall, German circumstances have a probability of 38.9% while English circumstances have a probability of 33.9% to become the Theme of their clause.

This rather small difference does not explain the large discrepancy in total numbers of circumstance Themes that was found between the original corpora. The reason why GO includes so many more circumstance Themes than EO is that apparently German speakers use a considerable higher number of circumstances in general, regardless of position. Naturally, this also increases the number of circumstance Themes that were found. In other words, the positional flexibility of circumstances is not that different between English and German, German just has more of them. The same cannot be said for Complements, which are clearly more mobile in German.

Hawkins (1986: 67) claims that English is less restricted than German regarding the semantic mappings to clause elements. While the number of inanimate Subject types and sentient verbs was generally higher in EO, the differences were relatively small and also not statistically significant. There were even more Place Subjects as part of non-sentient construction in German, which is counter-intuitive given that GO already included a much higher number of circumstances of Place. It appears as if German is more accepting of unorthodox Subject-Verb combination than originally assumed, a process that is also predicted in König and Gast (2009) and Königs (2011).

Another surprising revelation was that Subject Theme identifiability was basically identical in EO and GO. One might assume that the number of identifiable Subject Themes in German was higher due to the restrictions on early referents in the German Midfield. Subjects in EO had a higher probability by 0.1% to be identifiable, a statistically non-significant difference.

Most of the thematic differences that were found overall were also evident across all four registers. For example, Place Themes were more common in German and Time Themes more common in English, which also affected the individual registers. Register-specific differences in Theme were scarce because the thematic spaces of individual registers were very similar in English and German. FICTION and INSTR, especially, had very similar thematic distributions in EO and GO, regarding almost all Theme measures. Political speeches in English included more material processes while German speeches were more marked by mental processes. This difference was in large parts attributed to a difference in the expression of modality. Many of the additional mental processes in German were used to express the desires of the speaker in cases where they did not want to explicitly state what had to be done in the

future but rather what they themselves wished would happen. In English, such hedges were linguistically realized mostly through modal verbs, which were very uncommon in SPEECH in GO. Vocative Themes were also more common in GO, which suggests that there is more interaction between the participants in the speech event in German. Apart from these differences, the register was thematically very similar. The greatest register difference between EO and GO was found in TOU and concerned the use of process types, which was already mentioned above. TOU in English was a lot more static and included more relational and existential processes to describe places and events. TOU in German was more dynamic by featuring more material processes, which were used to depict the reader as a participant in a guided tour, who travels to the places that are discussed in the leaflet.

Chapter 10 was concerned with differences in Theme in both translation directions. The results that were shared in this chapter were largely descriptive and similar in structure to the previous three results chapters. Regarding multiple Themes, the same effects that were found in the contrastive analysis were also evident in the analysis of translation, which is that it highly depends on the choice of Theme hypothesis which of the languages included more Theme elements on average. If the Forefield hypothesis in German and the first experiential element hypothesis in English are considered, the Theme numbers in EO and ET were higher than in GO and GT. If the same hypotheses were used for both languages, the average Theme number in the German subcorpora were in turn higher. When comparing Theme based on the first experiential element hypothesis in both languages, no significant difference in Theme number could be found if the Finite Themes in GO and GT were disregarded. It is thus fair to say that Theme number differences are predominantly caused by the additional Finite Themes in German.

Process type distribution and participant Themes were again remarkably similar across all four sub-corpora. The number of relational processes increased, while the number of material processes decreased in both translation directions. These small differences were, however, not statistically significant.

GO included a significantly higher number of marked Themes than EO. The frequency of marked circumstance Themes in ET and GT were in between those of the original subcorpora, which represents a combination of shining through (Teich 2003: 145) and normalization (Baker 1996: 176-177) in translation. The same held true for Complement Themes in the

direction of English to German translations, as the number of marked Complement Themes in GT were in the middle of EO and GO. However, in the opposite translation direction, ET included fewer Complement Themes than GO and even EO. This was a clear example of normalization and over-correction, where the translators used fewer Complement Themes than even English authors of original texts.

The general distribution of circumstance Theme types looked rather similar across the four sub-corpora. The most notable differences regarded Place Themes, which were used more often in GO, and Condition Themes, which were better representatives of EO. If the numbers of GO and EO diverged, ET and GT were again located in between the two original subcorpora in most cases. The comparison of circumstance TPots revealed an astonishing result: Despite the fact that the number of circumstance Theme was significantly lower in ET compared GO, the thematic potential of circumstances was almost identical in the two sub-corpora. However, if the thematic potential of a circumstance is the same, their frequency of occurrence should also be the same, unless the two corpora include different numbers of circumstances overall. And this is exactly the case as the number of circumstances, regardless of position, decreased in ET by 11.6%. The same effect could be found in the opposite translation direction, where the number of circumstances overall increase by 12.6% in GT. The frequencies of three circumstance types in particular varied heavily between the subcorpora, namely Means, Place, and Quality.

At first, it seemed as if this imbalance was caused by translators adding or omitting circumstantial information in the translations. And while such cases were found, especially in the case of deictic expression, a more detailed analysis revealed that most of the missing circumstances were expressed in different ways in EO and ET. Many of the circumstances of Place and Means in GO were turned into the Subject of the clause in ET, which is possible because of the less restricted mappings of semantic meaning and syntactic function. The opposite effect could be observed in translations from English to German, where the additional circumstances of Place and Means were used to avoid inanimate Subjects in GT. Circumstances of Quality were oftentimes analyzed as a different circumstance type in EO and ET, which in part explains their lower numbers. That being said, the number of cases where an existing Quality Theme in GO was simply omitted in ET was highest for this circumstance type.

The use of inanimate Subjects and non-sentient constructions was very comparable in the original and translated texts. The frequencies of middle animate Subjects as part of non-sentient constructions in particular did not change noticeably in either translation direction. Regarding inanimate Subjects, the differences became larger, especially between EO and GT, but not large enough to be statistically significant. This once again shows that such constructions become more and more accepted in German and are changed in translations only occasionally.

Subject Theme identifiability increased in translations from English to German and decreased in translations from German to English. The higher number of identifiable Subjects in GT was predicted but the decrease in the other translation direction came as a surprise. That being said, the difference between GO and ET was not significant. It was assumed that the frequency of identifiable Subject Themes would increase because of the identifiability restrictions in the German Midfield. If the English original clause includes a non-identifiable Subject Theme which would have to be placed in an early Midfield position in the German translation, the translator may opt to change its identifiability to avoid a marked Midfield sequence. However, the results show that identifiability goes up in the Midfield as well as in the Forefield. Subject Themes in German are not restricted in terms of their identifiability if they are positioned in the Forefield, which is why the original explanation cannot account for all of the differences. A closer analysis of the changed sentence pairs revealed that differences in identifiability were very register-dependent. In FICTION, identifiability changes affected identifiable and non-identifiable Subject equally. In INSTR, the higher number of identifiable Subjects were often a byproduct of changes to inanimate Subjects. One common way of changing an inanimate Subject in the German translation was to change the original Subject to a circumstance and add a second person pronoun as the new Subject of the translation, referring to the reader. Pronouns are of course identifiable. And in SPEECH, there seemed to be different conventions regarding the inclusion or omission of articles in combination with abstract nouns. There are multiple examples where the English original Subject is an abstract noun, which is lacking an article, and where a definite article is added in the German translation. This addition of articles is again independent of position, which suggests that this is simply a more common way of realizing abstract nominal groups in German.

Chapter 11 included the inferential analysis of translations and translation changes. The two translation directions were separated in this chapter because the results were based on larger, independent linear regression models and a side-by-side analysis would have lacked clarity. The linear models included predictor variables from all the Theme aspects that were discussed prior, which were tested in terms of how well they predicted a change to the thematic structure in the translation. Change was operationalized in two ways: first binary, where only the two options of changed and unchanged were distinguished, and second numerical, where the number of changes were added up to individual Change values.

In translations from English to German, short Themes had a significant effect on Change. Most of the short Themes in English were single pronouns acting as Subjects that were oftentimes positioned in an early Midfield position in German, which is the default place for identifiable Subject pronouns. As a consequence, the new first experiential element was often longer than the short Subject or the number of Theme elements increased because the new Subject Theme now followed a Finite Theme in German. Additionally, long and very long Themes were also regularly changed regarding their length in the Subject hypothesis. Nonetheless, most of these long and very long Subject Themes were not actually changed in any meaningful way in the German translation. The literal translation of the Subject Themes just happened to be short enough for them to be part of the next smaller length category.

Non-Subject Themes were consistently strong predictors of Change in the Subject hypothesis, but this was just because the translators had to add an additional Finite Theme between the circumstance and the Subject Theme, which increased the number of Theme elements and thus constituted a Theme change. Therefore, every case of circumstance Theme in English-German translations needed to at least include a change in Theme number, which explains why every circumstance type was significant. In the other two Theme hypotheses, the addition of Finite Themes was not obligatory, which is why their analysis were more meaningful for this Theme aspect. Circumstance Themes were on the whole not very good predictors of translation changes in translations into German. There were even two circumstance types, Condition and Time, that negatively predicted Change, which means that they were significantly less likely to be changed than Subject Themes. Behalf and Guise Themes were the only circumstance types that had significant, positive effects on Change; however, these

effects were not consistently observable in the different linear regressions. The fact that German has a weaker notion of markedness has been discussed repeatedly throughout this thesis. However, following the results of Freiwald (2016), it was assumed that there are some circumstance types which are more marked in the German Theme than in the English one, which the contrastive analysis confirmed, and that they are changed systematically in German translations, which the inferential analysis of translations did not confirm. Even though there are differences in thematic potential of individual circumstance types, the positional flexibility in German seems to allow even more marked Theme structures in German translations.

Surprisingly, Carrier Themes were prone to translation shifts even though the number of relational processes increases in GT. However, this appears to be largely a side-effect of a different Theme category. Apparently, an above average number of Carrier Themes are either long or very long Themes, which have already been shown to be reliable Change predictors. Furthermore, existential process Themes and Phenomenon Themes were also shown to be systematically changed in GT. Notwithstanding, both of these participant Themes were primarily changed in one particular register, which suggests that it is rather a register-specific effect. Most of the existential process Themes came from TOU where they were often changed to relational processes. This was particularly common if the original existential clause also included a circumstance of Place, which was then turned into the Attribute. The majority of Phenomenon Themes was part of SPEECH and primarily involved in changes to the voice or the order of Senser and Phenomenon.

As expected, non-identifiable Subject Themes were a significant Change predictor. If the original English Theme was a multiple Theme that included a non-identifiable Subject alongside other Theme elements, the German translator had to either change the number or the order of the Themes or accept to have a non-identifiable nominal group early in the Midfield, which is an atypical position in German. Moreover, as was commented on before, there were some more abstract nominal group that were more likely to be accompanied by definite articles in German, especially in SPEECH.

The results of Subject Theme sentience were among the most surprising in this translation direction. Nearly all Subject types deviated significantly from the reference variable Human Subject regarding their likelihood of change. This was especially surprising for Subjects of

middle animacy like Organization and arguably Machine. However, a closer look at individual translations of Organization Subject Themes revealed that most of their changes are not changes to the Subject animacy but to Theme number because a high number of Organization Subjects were accompanied by Time Themes in EO. As was explained above, any circumstance Theme in English involved a change in Theme number in the Subject hypothesis. Machine Subjects, on the other hand, were changed much more readily in terms of their Subject type, which calls into question whether they are indeed to be treated like other middle animate Subjects in German.

Inanimate Subjects also underwent systematic changes. However, such changes were common regardless of the sentience requirements of the verb. Apparently, the animacy of the Subject alone often warranted a change in the German translations. This result was not expected and suggested that the relationship between semantics and clause functions is generally more restricted in German regardless of the verb requirements. This result was even more surprising given the non-significant differences of middle animate and inanimate Subjects in Chapter 10. That being said, the estimates for inanimate Subjects were consistently higher if they were paired with a sentient verb so it is still accurate that non-sentient constructions are more marked in German.

INSTR and TOU on the whole were more likely to include Theme changes than FICTION and SPEECH. This result can be explained by the different Theme categories that were already discussed as being strong Change predictors, especially Subject Theme identifiability and Subject Theme sentience. INSTR and TOU included a higher number of non-identifiable and inanimate Subject Themes than the other two registers, which is why they also include more Theme changes. SPEECH had a significant, negative effect on Change compared to FICTION, which is unexpected given that they are so similar in terms of the Theme structure in EO. SPEECH was less likely to undergo change in almost all Theme categories, which suggests that translators are generally less inclined to make changes to the thematic space in political speeches.

Theme number was hypothesized to be one of the strongest predictor variables in English to German translations given the Finite-second constraint and the significant results regarding Theme number differences between English and German in the previous chapters. Astonishingly, Theme number was consistently not significant in the logistic regressions. It was

argued that this is largely due to the relationship between Theme number and textual and interpersonal Themes. The inclusion of textual and interpersonal Themes significantly increased the likelihood of a Theme change. In the first experiential element hypothesis, any Theme number that is higher than one has to include a textual or interpersonal Theme. In the Subject hypothesis, a multiple Theme can also consist of a marked Theme followed by a Subject Theme, but the majority of multiple Themes still included a textual or interpersonal Theme. Thus, it was argued that these three variables take away much of their explanatory power, which resulted in the non-significant results of Theme number. If the regression models excluded textual and interpersonal Themes as predictor variables, Theme number had a consistently positive, significant effect on Change.

In the translation direction from German to English, a similar relationship between Theme number and the presence of textual and interpersonal Themes was suspected. This time, Theme number was, as expected, a strong predictor of Change in all logistic regressions but textual and interpersonal Themes allegedly had a significant, negative effect on Change, even though the inclusion of a non-experiential Theme is the only way to construct multiple Themes in the Forefield and the first experiential element hypothesis. This was again interpreted as a case of multicollinearity, which was supported by the fact that with the exclusion of the variable Theme number, the estimates of textual and interpersonal Theme all turned positive and were significant in all but one model.

Multiple Themes in GO were often reduced in number in ET in the first experiential element and the Subject hypothesis. The main reason for this reduction of Theme elements is the obligatory Finite Theme in German that is excluded in the English translations. These results were unsurprising given the significant difference in Theme elements between GO and EO caused by the extra Finite Themes. In the Forefield hypothesis, higher Theme numbers also had significant, positive effects on Change, which was the exact opposite effect than what was originally hypothesized. It was assumed that simple Themes that consist of only a single textual or interpersonal Theme would be systematically changed in ET by adding the obligatory experiential Theme to the target clause. Such cases were numerous and yet they were counteracted by another common change made to multiple Themes in GO. In the Forefield hypothesis, there are only few elements that do not take up the entire Forefield but allow

a second clause element, namely conjunctions and some interpersonal elements like Vocatives. Such multiple Themes were often reduced in number ET, typically by moving or omitting the non-experiential Theme. This translation procedure was so prominent that it outweighed the increase of Theme elements in the case of simple Themes.

All hypotheses regarding Theme length in German to English translations turned out to be inaccurate. It was assumed that short Themes would be increased, and long Themes decreased in length as effects of explicitation and simplification. In fact, it was the medium-sized Themes that were most subject to change. The most common type of change was an increase in length from medium to long Themes in the English translation. However, the in-depth analysis revealed that these longer Themes did not carry any additional information or had a more complex grammatical structure. Most of them appeared to be literal translations of the German Theme, which just happened to be long enough in characters to count as the next longer category. Given that the exact opposite effect was found in English to German translations, a more thorough analysis of Theme length is warranted.

Differences in Theme markedness between English and German were discussed at length throughout this thesis. Non-Subject Themes were therefore expected to have a significant, positive effect on Change. However, marked Themes were not assumed to be universally challenging. Instead, it was hypothesized that those marked Themes that had a noticeably higher thematic potential in GO than in EO were likely candidates for change in ET, which applied to eight marked Theme types. This hypothesis was largely accurate. Of these eight types, six were associated with higher rates of change: Place, Reason, Means, Comitative, Quality, and Complement Themes. Duration and Purpose Themes did not have significant, positive estimates despite their lower thematic potential in EO. Additionally, three other types of marked Themes, which were not included in the hypothesis, did deviate significantly from the reference category of Subject Themes, namely Behalf, Matter, and Guise. A closer analysis of the erroneously evaluated circumstance Themes revealed that many of the unexpected changes were tied to formal aspects. Generally, circumstance Themes were much more likely to be kept in Theme position in ET if they were realized as clauses rather than phrases. This impression was not evaluated statistically, which leaves room for a more detailed analysis of marked Themes and form for the future. Nevertheless, the majority of marked Themes was predicted accurately on the basis of thematic potential. Additionally,

three circumstance Themes had significant, negative estimates on Change: Condition, Concession, and Time. Incidentally, these are also the only types that had a thematic potential of 50% or higher, which means that they are just as likely if not more likely to be chosen as the Theme of a clause over the Subject. This clearly shows that thematic potential is a factor that English translators, knowingly or unknowingly, take into consideration when translating German texts.

The translations of different participant Themes were not particularly noteworthy in German to English translations. Participant roles that are more likely to be mapped onto Complements, like Goal and Phenomenon, were also more likely to be changed, which is unsurprising given how marked Complement Themes are in English. Initiator Themes were also changed regularly but mostly in terms of their identifiability rather than in terms of their participant role or process type. If a Value participant in an identifying relational process is used as the Theme of the source clause, it is often moved to the Rheme in ET. Most Senser Themes remained unchanged in the English target text despite the fact that German allows the mapping of the Senser role onto a Complement. Nevertheless, the majority of Senses in GO are Subject Themes, which stayed mostly intact in translations into English. Given that most mental processes can be found in FICTION, this effect may be highly register-dependent.

Subject Theme identifiability was generally inconsequential in terms of Theme changes. Unlike German, English is not formally restricted in its positioning of identifiable and non-identifiably references. There is a strong preference for using given information as Theme in English (Halliday and Matthiessen 2014: 120), but that does not seem to keep English translators from preserving non-identifiable Subject Themes in the target text.

Similarly, Subject sentience was also not a particularly reliable predictor of Change. Two types of inanimate Subject Themes had significant, positive estimates regardless of the verb requirements: nonconcrete Inanimates and Place Subjects. However, the more detailed analysis revealed that these Subject types were not changed in terms of their animacy or sentience but rather in terms of their process types or participant roles that indirectly affected Subject sentience. Many such Subject types were turned into Carrier Themes or Goal Themes as part of a passive construction but kept their inanimate status. Moreover, Place Subjects were often accompanied by Complement Themes in the source clause, which have already

been shown to be strong Change predictors. It is thus safe to say that Subject animacy and verb requirements are a non-issue in translations from German to English.

The different registers also deviated significantly from one another, with SPEECH and TOU being overall more prone to thematic change in ET than FICTION and INSTR. This can again be tied to some of the previously discussed reliable Change predictors. SPEECH has the highest Theme number in GO, which was often reduced in ET. The majority of Place Subject Themes and marked Complement Themes was found in TOU, which also explains its higher change numbers. That being said, the rates of change were overall very comparable between the four registers and it was seldom a single change category but rather multiple elevated change categories that caused the significant results.

Lastly, it was shown that changes to Theme were more numerous in translations into English than in translations into German. This means that English translators are more willing to change the thematic structure of the source clauses to be more like the thematic structure of the target language. German translations, on the other hand, showed more signs of shining through. Evert and Neumann (2017) made a similar observation comparing multiple linguistic phenomena in English and German translations. They interpret this effect as a difference in language prestige. Although the analyses in this thesis only considered a single linguistic feature, the results on Theme can be seen as further support of this interpretation.

13 Conclusion

In the introduction of this thesis, two main aims were formulated: One, I wanted to ascertain all facets and subtleties of the Theme in systemic functional linguistics, both in terms of its general function and its language-specific formal realizations. Two, I wanted to use these theoretical insights and analyze how contrastive differences of Theme affected translations between English and German empirically. In this final chapter, I will briefly summarize each chapter of this thesis and consider whether these aims were achieved. Above that, the chapter includes a critical assessment of the methodology and an outline of possible future steps and concludes with some final remarks.

In Chapter 2, basic terminologies and concepts in translation studies were outlined. It was proposed that translating is the impossible task of transferring a text in one language into an equivalent version of that text in a different language. It was described as impossible because this transfer into a different language system inevitably requires changes by the translator both on the formal and the functional level. Such changes can be grouped into different translation procedures and the translator must decide which of these is most appropriate in each context. As a consequence of these changes, translated texts are supposedly characterized by translation features that set them apart from original texts.

Chapter 3 was an excursion into the grammatical structure of English and German clauses. The focus here was mostly on German given its unusual position of the verbal units and its three topological fields. For the purposes of this study, the topological fields Forefield and Midfield are highly relevant as they divide the Theme and the Rheme in German, as argued by Steiner and Teich (2004). It was shown that these fields do not only differ in terms of their position in the clause but also regarding the number, the types, and the order of elements that they can include. It was important to discuss differences in clause structures first because they are vital in understanding some of the formal thematic differences between English and German. Chapter 3 also included a discussion of the relationship between grammar and semantics by highlighting some of the contrastive difference in Subject-Verb combinability between English and German.

Chapter 4 presented a quick overview of the framework of SFL. This included central concepts relevant to the goals of the thesis as well as a detailed discussion of the three main

metafunctions in SFL and their relationships between each other. As the main system of the textual metafunction, THEME was briefly introduced here as well. THEME and MOOD are uniquely intertwined as the thematic structure mirrors the primary purposes of the different speech functions expressed through mood. The relationship between TRANSITIVITY and THEME is more subtle. Nevertheless, the choice of process types, participant roles, and circumstance types has a verifiable effect on Theme probabilities as well.

The central theoretical chapter was Chapter 5, which included a thorough discussion of Theme in SFL. First, Theme was contrasted with other linguistic concepts, which are either related to Theme in terms of meaning or even also called Theme in other linguistic frameworks. Following that, different approaches to the functional meaning of Theme were presented alongside some of the most common lines of criticism. The formal realization in English was presented, including the central thematic aspects of Theme markedness and multiple Themes, followed by a short summary of Theme form in other languages to identify the main similarities and differences in Theme realization across a multitude of languages. While the formal realization of Theme differs from language to language, early positioning, and the necessity of a relationship to the experiential realm were two aspects of Theme that most languages shared. Subsequently, the formal realization of Theme in German, which was primarily based on Steiner and Teich (2004), was discussed and compared to Theme in English. One of the most striking differences between English and German Theme is the restriction of one, potentially non-experiential Theme in German, which is caused by the relationship between informational meaning and Midfield sequencing. However, it was also argued that a case can be made for multiple Theme hypotheses in both languages. The chapter was concluded with insights from the state of the art on Theme in translations and the presentation of the hypotheses.

Chapter 6 outlined the methodology, which included a presentation of the CroCo corpus, the four registers and the annotation tools used in this project. This chapter also contained a detailed explanation of annotation decisions and concluded with an outline of the statistical tests.

Chapters 7 to 11 contained the empirical results and discussion of Theme in the different subcorpora. Chapters 7 and 8 presented the thematic structures of English and German in isolation with a focus on register differences in each of these languages. In Chapter 9, these

results were used to compare Theme distributions contrastively to highlight some of the main differences between English and German Theme. Chapters 10 and 11 were concerned with the relationship between the original subcorpora and their translations, first in general terms and then with the help of regression analyses. Chapter 12 summarized the most important findings of the previous five chapters.

When I decided to make Theme the focus of my Ph.D. project, I was well aware that a thorough and comprehensive discussion in two different languages was a challenging and potentially impossible task. I do believe that the fifth chapter represents a discussion of Theme that is unique in the systemic functional community regarding detail and relation to other frameworks, viewpoints, and languages. I also think I make a strong argument for the possibility of different formal definitions of Theme, especially in German. That being said, the descriptions of Theme in this thesis are also not free of the two main criticisms that have been brought forward repeatedly in regards to Theme: One, that the functional descriptions of Theme are too metaphorical and vague and two, that the different kinds of elements that are said to be thematic have little in common regarding their meanings. From a formal perspective, I do believe that the fronting of an element is meaningful and a result of choice both in English and German and I also believe that the functions of these fronted elements are comparable between the two languages. While the meanings of textual, interpersonal, and experiential elements are fundamentally different, they do, in my opinion, contribute to the interpretation of clauses and texts if used in Theme position, albeit in very different ways. Whether this commonality is too abstract to be meaningful, as Fawcett (2007: 137) suggests, remains debatable. While I do not claim to have solved the mystery that is Theme, I do hope that I have shed some light on the discussion and expanded the understanding of Theme in the linguistic community.

What has been most surprising to me was how much the empirical results were able to inform the theoretical discussions. The analysis of a variety of different Theme hypotheses revealed some interactions between early elements that allowed meaningful conclusion regarding Theme. On the basis of this, I am confident in saying that Theme needs a relation to the experiential realm both in English and German. Whether or not the Theme necessarily includes the Subject is a meaningful debate and should be related to the research question. Especially in less quantitative analyses of Theme and thematic progression, in which the

meaning of individual texts is worked out, the inclusion of the obligatory Subject Theme is likely crucial. However, in a quantitative, clause-based analysis of Theme, as was performed in this thesis, the Subject hypothesis, in my opinion, not only masks the significance of marked Themes, it also does not reveal many additional insights compared to the first experiential element hypothesis and even the Forefield/first element hypothesis. In summary, while the theoretical discussion of the extent of Theme is in my eyes meaningful, the choice of Theme hypothesis in a quantitative analysis is largely inconsequential.

The effects of Theme in English and German on translation have been outlined and discussed meticulously in the final results chapters. I am confident in saying that the second main goal of this thesis has been attained. Without repeating some of the results again, the translation results have clearly shown that Theme is very relevant in translations between English and German and that translators consciously or subconsciously take note of thematic differences between the two languages and try to balance them in their translations. The analyses have also shown that the similarities and differences of Theme between source and target text are an amalgamation of different influencing factors, like register aspects, contrastive differences, translation inherent tendencies, language prestige, and others.

The methodology was very appropriate for the aims of this project and there are only few methodological decisions that I would have changed in hindsight. One of the greatest shortcomings of the annotations are, of course, that they were only carried out by a single annotator. This not only makes the annotations less objective, but it also leaves room for simple oversights. It was also not ideal that there was a gap of two and a half years between the first and the last register that was annotated since the quality of the annotations clearly improved over time.

The UAM CorpusTool is an incredibly useful tool especially because it allows you to design your own annotation schemes and autocode elements based on existing annotations. Nonetheless, the transfer of the corpus tool data to Excel was very time consuming and required elaborate Excel functions. It may have been easier to do all the annotations in Excel directly, which would have likely increased annotation time but would have simplified the analysis tremendously.

I consider the majority of my annotation decisions to be appropriate. However, I should have not grouped elements in the same category if their meanings were too dissimilar. This

was particularly the case for circumstances. Given the high number of 22 circumstance types in SFL, I wanted to avoid introducing additional groups and generally tried to find a category that best described the elements in question. Unfortunately, this practice resulted in circumstance categories like Guise and Comitative that were only loosely related in meaning and that lead to some poorly designed hypotheses. Given the amount of data points, I should have been more willing to introduce additional annotation categories.

One of my central arguments regarding German Theme was that, at the very least, additional Theme hypotheses are also plausible, and that the consideration of Theme forms is meaningful. The extent of Theme in English has been a point of discussion in the state of the art for decades, which is why I decided to include the results of different Theme hypotheses to highlight quantitative similarities and differences. Additionally, I wanted to avoid alienating potential readers, given the plurality of opinions regarding Theme realization. Contrasting the different hypotheses in the early chapters was very meaningful and illuminated different Theme aspects. However, dealing with three Theme combinations in the translation results was time-consuming and potentially sacrificed clarity to some extent, especially during the inferential analyses. Perhaps, it would have been beneficial to decide on only one Theme hypothesis each; however, I did not want to forgo this level of detail especially because the data was available from my annotations. Given the scope of this project, I stand behind this decision, but for future projects, I will refrain from working with multiple Theme hypotheses.

The statistical tests were chosen appropriately. However, this thesis includes a high number of tests, which make type I and II errors almost inevitable. Ideally, I would have liked to run fewer tests but to forgo testing for some assumptions and not for others also seemed imperfect. The distinction between binary and numerical change was appropriate in theory but it potentially further contributed to a lack of clarity in Chapter 11. For future research, I would only use numerical Change as the predicted variable of a mixed model revolving around Theme. The preference of numerical Change over binary Change is mostly a practical one since those models that were based on binary Change often did not converge and arguably misrepresented Theme aspects that were consistently changed in translations as was the case with Predicator Themes in translations from German to English.

While the Theme analysis of this Ph.D. thesis is quite detailed, there is still some room for future work. Like many other Theme-related research projects, this thesis only focused on declarative clauses. To my knowledge, there is no detailed, empirical analysis of Theme structure in imperative and interrogative clauses, neither for English nor for German. This area of Theme is a clear shortcoming in the state of the art, one that I would have liked to address in this thesis but had to give up due to its already large scope.

Including additional registers would of course also increase the reliability and generalizability of the results. In the contrastive analysis, some Theme aspects were remarkably similar overall, yet very different between the registers, so a different set of registers would have potentially also changed the results and the interpretation. An obvious register to include would be a spoken register given that all the registers analyzed in this project are written ones. Spoken language likely has a unique thematic structure especially in terms of multiple Themes and Theme markedness.

One thematic aspect that was unfortunately disregarded in this project was formal structure. The results suggested that circumstances in particular are used very differently if they are realized as a phrase or a clause. Additionally, different types of phrases and different kinds of subordinating conjunctions are also likely to make a difference in the positioning of circumstances. Theme length was included in the inferential analysis of translations, but the results were inconclusive as the Theme length category would often change without any noticeable change to the form or function of the translated Theme. This result was particularly surprising since the division of length categories already considered contrastive differences. Theme form and Theme length are thus aspects that should be highlighted more in research on Theme.

The inclusion of additional languages would also be very fruitful to disentangle contrastive from translation-inherent influencing factors. Besides, the disregard of thematic progression could be considered one of the major shortcomings of my Theme analyses. Both of these aspects could lead to very interesting research endeavors.

The corpuslinguistic analysis of Theme was tremendously insightful and I want to continue corpus-based translation studies on Theme as well as other linguistic phenomena. However, a change of methodology may compensate for some of the shortcomings of this thesis. Theme is criticized as being too vaguely defined and different thematic elements have

arguably very different effects on the interpretation of the message. Berry (1996) went away from just theoretically discussing the nature of Theme and instead tried to identify its function by asking authors to describe the prioritized meanings of their texts. While I do not necessarily agree with her method, I applaud the attempt of linking the proposed functional meaning of Theme to the perceived meaning of clauses and texts judged by the language users. Similarly, I could imagine a future research project where Theme-induced differences in clause and text interpretation are analyzed with the help of psycholinguistic measuring tools and reader interviewing. This would bring the empirically-based understanding of the nature of Theme one step further.

As my final remark, I would like to again thank all the people that directly or indirectly contributed to the creation of this thesis. When I started working with Theme more than six years ago, I would have never assumed that this linguistic concept would take up such a significant part of my life. My work on Theme has revealed many interesting insights into both the English and the German language to me and I have gained a much deeper understanding of translation in general and Theme translation in particular. However, the more I work with Theme, the more I also realize how much I have yet to grasp and how much interesting research in this field is potentially still ahead of me.

References

- Ackerman, Farrell, and John C. Moore. 2001. *Proto-Properties and Grammatical Encoding: A Correspondence Theory of Argument Selection*. Stanford: CSLI Publications.
- Adam, Séverine, and Cécile Delettres. 2016. "Textstrategische Funktion indefiniter Komplement-NPs am Satzanfang. Eine kontrastive Untersuchung Dt./Frz." In *Variation im europäischen Kontrast: Untersuchungen zum Satzanfang im Deutschen, Französischen, Norwegischen, Polnischen und Ungarischen*, edited by Martine Dalmas, Cathrine Fabricius-Hansen, and Horst Schwinn, 119–55. Berlin, Boston: Walter de Gruyter.
- Admoni, Wladimir. 1962. "Die umstrittenen Gebilde der deutschen Sprache von heute. II. Der Satzrahmen." *Muttersprache* 6: 161–71.
- Admoni, Wladimir. 1970. *Der deutsche Sprachbau*. 3rd ed., revised and extended. München: C.H. Beck'sche Verlagsbuchhandlung.
- Ahearn, Laura M. 2001. "Language and Agency." *Annual Review of Anthropology* 30: 109–37.
- Altenberg, Bengt. 1998. "Connectors and Sentence Openings in English and Swedish." In *Corpora and Cross-Linguistic Research: Theory, Method, and Case Studies*, edited by Stig Johansson and Signe Oksefjell, 115–43. Language and Computers: Studies in Practical Linguistics 24. Amsterdam, Atlanta: Rodopi.
- Altmann, Hans. 1981. *Formen der «Herausstellung» im Deutschen: Rechtsversetzung, Linksversetzung, Freies Thema und verwandte Konstruktionen*. Linguistische Arbeiten 106. Tübingen: Max Niemeyer Verlag.
- Andersen, Thomas Hestbæk, Uwe Helm Petersen, and Flemming Smedegaard. 2001. *Sproget som ressource: Dansk systemisk funktionel lingvistik i teori og praksis*. Odense: Odense Universitetsforlag.
- Bach, Emmon. 1962. "The Order of Elements in a Transformational Grammar of German." *Language* 38 (3): 263–69. <https://doi.org/10.2307/410785>.
- Baker, Mona. 1992. *In Other Words: A Coursebook on Translation*. London, New York: Routledge.
- Baker, Mona. 1993. "Corpus Linguistics and Translation Studies." In *Text and Technology*, edited by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 233–50. Amsterdam: John Benjamins Publishing Company.
- Baker, Mona. 1995. "Corpora in Translation Studies: An Overview and Some Suggestions for Future Research." *Target* 7 (2): 223–43.
- Baker, Mona. 1996. "Corpus-Based Translation Studies: The Challenges That Lie Ahead." In *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*, edited by Harold Somers, 175–86. Amsterdam: John Benjamins.
- Bartlett, Tom. 2014. *Analysing Power in Language: A Practical Guide*. London, New York: Routledge.
- Bartoń, Kamil. 2018. *MuMin: Multi-Model Inference*. R package version 1.10.0. <https://CRAN.R-project.org/package=MuMin>.

- Bassola, Péter, and Horst Schwinn. 2016. "Markierte Vorfeldbesetzungen im Deutschen." In *Variation im europäischen Kontrast: Untersuchungen zum Satzanfang im Deutschen, Französischen, Norwegischen, Polnischen und Ungarischen*, edited by Martine Dalmas, Cathrine Fabricius-Hansen, and Horst Schwinn, 229–59. Berlin, Boston: Walter de Gruyter.
- Bates, Douglas, Martin Maechler, Ben Bolker, and Steve Walker. 2015. "Fitting Linear Mixed-Effects Models Using lme4." *Journal of Statistical Software* 67 (1): 1–48.
- Becher, Viktor. 2010. "Abandoning the Notion of "Translation-Inherent" Explication: Against a Dogma of Translation Studies." *Across Languages and Cultures* 11 (1): 1–28.
- Behrens, Bergljot, and Kåre Solfeld. 2014. "Discourse Role Guiding Clause Types and Position in Translation." In *Adverbials in Use: From Predicative to Discourse Functions*, edited by Laure Sarda, Shirley Carter-Thomas, Benjamin Fagard, and Michael Charolles, 267–96. Louvain-La-Neuve: Presses universitaires de Louvain.
- Bell, Roger T. 1991. *Translation and Translating: Theory and Practice*. London, New York: Longman.
- Beneš, Eduard. 1959. "Začátek německé věty z hlediska aktuálního členění větného ("The Beginning of the German Sentence from the Point of View of Functional Sentence Perspective")." *ČMF* 41: 205–17.
- Beneš, Eduard. 1971. "Die Besetzung der ersten Position im deutschen Aussagesatz." In *Fragen der strukturellen Syntax und der kontrastiven Grammatik*, edited by Hugo Moser, 160–82. Düsseldorf: Pädagogischer Verlag Schwann.
- Beneš, Eduard. 1973. "Thema-Rhema-Gliederung und Textlinguistik." In *Studien zur Texttheorie und zur Deutschen Grammatik: Festgabe für Hans Glinz zum 60. Geburtstag*, edited by Horst Sitta and Klaus Brinker, 42–62. Düsseldorf: Pädagogischer Verlag Schwann.
- Benson, James D., William S. Greaves, and Glen Stillar. 1992. "Motivating Text and Construing Context in Two Film Reviews." Paper presented to the 19th International Systemic Functional Congress, July 13-18, 1992.
- Bernardini, Silvia, and Federico Zanettin. 2004. "When Is a Universal Not a Universal? Some Limits of Current Corpus-Based Methodologies for the Investigation of Translation Universals." In *Translation Universals*, edited by Anna Mauranen and Pekka Kujamäki, 51–62. Amsterdam: John Benjamins Publishing Company.
- Berry, Margaret. 1975. *An Introduction to Systemic Linguistics: I Structures and Systems*. New York: St. Martin's Press.
- Berry, Margaret. 1987. "The Functions of Place-Names." *Leeds Studies of English* 18: 71–88.
- Berry, Margaret. 1989. "Thematic Options and Success in Writing." In *Language and Literature: Theory and Practice. A Tribute to Water Grauberg*, edited by Joanna M. Channell, Christopher Butler, and Richard A. Cardwell, 62–78. Nottingham: University of Nottingham.
- Berry, Margaret. 1995. "Thematic Options and Success in Writing." In *Thematic Development in English Texts*, edited by Mohsen Ghadessy, 55–84. London, New York: Pinter Publishers.
- Berry, Margaret. 1996. "What Is Theme? - A(nother) Personal View." In *Meaning and Form: Systemic Functional Interpretations: Meaning and Choice in Language: Studies for Michael*

- Halliday*, edited by Margaret Berry, Christopher Butler, Robin Fawcett, and Huang Guowen, 3–64. Norwood: Ablex Publishing Corporation.
- Bestgen, Yves, and Wietske Vonk. 2000. "Temporal Adverbials as Segmentation Markers in Discourse Comprehension." *Journal of Memory and Language* 42 (1): 74–87. <https://doi.org/10.1006/jmla.1999.2670>.
- Biber, Douglas. 1988. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press.
- Biber, Douglas. 1995. *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge: Cambridge University Press.
- Biber, Douglas, and Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad, and Randi Reppen. 1998. *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Bierwisch, Manfred. 1973. "Regeln für die Intonation deutscher Sätze." In *Untersuchungen über Akzent und Intonation im Deutschen*. 3rd ed., 99–201. Berlin: Akademie-Verlag.
- Bisiada, Mario. 2018. "The Editor's Invisibility." *Target* 30 (2): 288–309.
- Bloor, Thomas, and Meriel Bloor. 1995. *The Functional Analysis of English: A Hallidayan Approach*. London, New York: Arnold.
- Blum-Kulka, Shoshana. 1986. "Shifts of Cohesion and Coherence in Translation." In *Interlingual and Intercultural Communication*, edited by Juliane House and Shoshana Blum-Kulka, 17–35. Tübingen: Narr.
- Boost, Karl. 1964. *Neue Untersuchungen zum Wesen und zur Struktur des deutschen Satzes: Der Satz als Spannungsfeld*. 5th ed. Berlin: Akademie-Verlag.
- Brinkmann, Hennig. 1971. *Die Deutsche Sprache: Gestalt und Leistung*. 2nd ed., revised and extended. Düsseldorf: Pädagogischer Verlag Schwann.
- Brown, Gillian, and George Yule. 1983. *Discourse Analysis*. Cambridge: Cambridge University Press.
- Burchardt, Aljoscha, Katrin Erk, Anette Frank, Andrea Kowalski, Sebastian Padó, and Manfred Pinkal. 2009. "Using FrameNet for the Semantic Analysis of German: Annotation, Representation, and Automation." In *Multilingual FrameNets in Computational Lexicography: Methods and Applications*, edited by Hans C. Boas, 209–44. Trends in linguistics. Studies and monographs 200. Berlin: Mouton de Gruyter.
- Caffarel, Alice. 2004. "Metafunctional Profile of the Grammar of French." In *Language Typology: A Functional Perspective*, edited by Alice Caffarel, James R. Martin, and Christian M.I.M. Matthiessen, 77–137. Current Issues in Linguistic Theory 253. Amsterdam: John Benjamins.
- Catford, John C. 1965. *A Linguistic Theory of Translation: An Essay in Applied Linguistics*. Oxford: Oxford University Press.

- Chafe, Wallace L. 1976. "Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View." In *Subject and Topic: Symposium on Subject and Topic (University of California, Santa Barbara, 1975)*, edited by Charles N. Li, 26-55. New York: Academic Press.
- Corbett, John. 2009. "Theme, Field and Genre: Thematic Realisations in Academic Articles and Their Popularisations." In *Text Type and Texture: In Honour of Flo Davies*, edited by Gail Forey and Geoff Thompson, 70–93. London, Oakville: Equinox.
- Croft, William. 1993. "Case Marking and the Semantics of Mental Verbs." In *Semantics and the Lexicon*, edited by J. Pustejovsky, 55–72. Studies in Linguistics and Philosophy 49. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-011-1972-6_5.
- Croft, William. 1996. "'Markedness' and 'Universals': From the Prague School to Typology." In *Multiple Perspectives on the Historical Dimensions of Language*, edited by Kurt R. Jankowsky, 15–21. Münster: Nodus Publikationen.
- Crompton, Peter. 2006. "The Effect of Position on the Discourse Scope of Adverbials." *Text & Talk* 26 (3): 245–79.
- Cruse, David A. 1973. "Some Thoughts on Agentivity." *Journal of Linguistics* 9 (1): 11–23.
- Dalmas, Martine. 2008. "Aspekte der Topikalisierung." In *Tradition und Geschichte im literarischen und sprachwissenschaftlichen Kontext: Unter Mitwirkung von Sonja Pachali, Aigi Heero, Merle Jung, Helju Ridali und Maris Saagpakk*, edited by Mari Tarvas, 83–94. Frankfurt am Main: Peter Lang.
- Daneš, František. 1964a. "A three-level approach to syntax." In *Travaux linguistiques de Prague 1*, edited by Josef Vachek, 225–40. Prague: Académie Tchèque des Sciences.
- Daneš, František. 1964b. "Téma // (Základ) // Východisko Výpovědi: [Theme // (Foundation) // Starting Point of the Utterance]." *Slovo a slovesnost* 25 (2): 148–49.
- Daneš, František. 1970. "One Instance of Prague School Methodology: Functional Analyses of Utterance and Text." In *Method and Theory in Linguistics*, edited by Paul L. Garvin, 132–46. The Hague, Paris: Mouton.
- Daneš, František. 1974. "Functional Sentence Perspective and the Organization of the Text." In *Papers on Functional Sentence Perspective*, edited by František Daneš, 106–28. Berlin: De Gruyter.
- Davidse, Kristin. 1987. "M.A.K. Halliday's Functional Grammar and the Prague School." In *Functionalism in Linguistics*, edited by René Dirven and Vilém Fried, 39–79. Amsterdam: John Benjamins.
- Davidson, Donald. 2001. *Essays on Actions and Events*. 2nd ed. Oxford: Clarendon Press.
- Davies, F. 1988. "Reading Between the Lines: Thematic Choice as a Device for Presenting Writer Viewpoint in Academic Discourse." *The ESPecialist* 9 (2): 173–200.
- Davies, F. 1997. "Marked Theme as a Heuristic for Analysing Text-Type, Text and Genre." In *Applied Languages: Theory and Practice in ESP*, edited by Jordi Piqué and David J. Viera, 45–79. Valencia: Universitat de València.
- Delaere, Isabelle. 2015. "Do Translations Walk the Line? Visually Exploring Translated and Non-Translated Texts in Search of Norm Conformity." PhD Thesis, Universiteit Gent.

- DeLancey, Scott. 1984. "Notes on Agentivity and Causation." *Studies in Language* 8 (2): 181–213.
- Diessel, Holger. 2005. "Competing Motivations for the Ordering of Main and Adverbial Clauses." *Linguistics* 43 (3): 449–70.
- Dik, Simon C. 1980. *Studies in Functional Grammar*. London, New York: Academic Press.
- Doherty, Monika. 1996. "Passive Perspectives; Different Preferences in English and German: A Result of Parameterized Processing." *Linguistics* 34: 591–643.
- Doms, Stefan, Bernard de Clerck, and Sonia Vandepitte. 2016. "Non-Human Agents as Subjects in English and Dutch: A Corpus-Based Translation Study." In *Atypical Predicate-Argument Relations*, edited by Thierry Ruchot and Pascale van Praet, 87–112. Amsterdam: John Benjamins Publishing Company.
- Downing, Angela. 1991. "An Alternative Approach to Theme: A Systemic Functional Perspective." *WORD* 42 (2): 119–43.
- Dowty, David R. 1991. "Thematic Proto-Roles and Argument Selection." *Language* 67: 547–619.
- Drach, Erich. 1963. *Grundgedanken der Deutschen Satzlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Duff, Alan. 1981. *The Third Language: Recurrent Problems of Translation into English*. Oxford, New York, Toronto, Sydney, Paris, Frankfurt: Pergamon Press.
- Dürscheid, Christa. 1989. *Zur Vorfeldbesetzung in deutschen Verbzweit-Strukturen*. Trier: Wissenschaftlicher Verlag Trier.
- Eggins, Suzanne. 1994. *An Introduction to Systemic Functional Linguistics*. London: Pinter Publishers.
- Eisenberg, Peter. 1994. *Grundriß der deutschen Grammatik*. 3rd ed., revised. Stuttgart, Weimar: Metzler.
- Engel, Ulrich. 1970. *Regeln zur Wortstellung*. Forschungsberichte des Instituts für deutsche Sprache 5. Mannheim.
- Engel, Ulrich. 1974. "Syntaktische Besonderheiten der deutschen Alltagssprache." In *Gesprochene Sprache: Jahrbuch 1972*, edited by Hugo Moser, 199–228. Sprache der Gegenwart 26. Düsseldorf: Schwann.
- Engel, Ulrich. 1988. *Deutsche Grammatik*. Heidelberg: Julius Groos Verlag.
- Engel, Ulrich. 2004. *Deutsche Grammatik: Neubearbeitung*. München: Iudicium Verlag.
- Enkvist, Nils Erik. 1973. "Theme Dynamics and Style: An Experiment." *Studia Anglica Posnaniensia* 5: 127–35.
- Erdmann, Peter. 1990a. *Discourse and Grammar: Focussing and Defocussing in English*. Tübingen: Niemeyer.
- Erdmann, Peter. 1990b. "Fokuskonstruktionen im Deutschen und Englischen." In *Kontrastive Linguistik*, edited by Claus Gnutzmann, 69–83. forum Angewandte Linguistik 19. Frankfurt am Main, Bern, New York, Paris: Peter Lang.

- Esau, Helmut. 1973. "Order of the Elements in the German Verb Constellation." *Linguistics* 11 (98): 20–40.
- Eskola, Sari. 2004. "Untypical Frequencies in Translated Language: A Corpus-Based Study on a Literary Corpus of Translated and Non-Translated Finnish." In *Translation Universals*, edited by Anna Mauranen and Pekka Kujamäki, 83–99. Amsterdam: John Benjamins Publishing Company.
- Etzensperger, Jürg. 1979. *Die Wortstellung der deutschen Gegenwartssprache als Forschungsobjekt: Mit einer kritisch referierenden Bibliographie*. Studia Linguistica Germanica 15. Berlin, New York: Walter de Gruyter.
- Evert, Stefan, and Stella Neumann. 2017. "The Impact of Translation Direction on Characteristics of Translated Texts: A Multivariate Analysis for English and German." In *Empirical Translation Studies*, edited by Gert de Sutter, Marie-Aude Lefer, and Isabelle Delaere, 47–80. Berlin/Boston: De Gruyter.
- Fabricius-Hansen, Cathrine. 2016. "Unbestimmtheit im Vorfeld: Deutsche und norwegische Wikipedia-Artikel im Vergleich." In *Variation im europäischen Kontrast: Untersuchungen zum Satzanfang im Deutschen, Französischen, Norwegischen, Polnischen und Ungarischen*, edited by Martine Dalmas, Cathrine Fabricius-Hansen, and Horst Schwinn, 75–118. Berlin, Boston: Walter de Gruyter.
- Faucher, E. 1976. "Exceptions à la règle V 2?" *Cahiers d'Allemand* 11: 27–47.
- Fawcett, Robin. 1980. *Cognitive Linguistics and Social Interaction: Towards an Integrated Model of a Systemic Functional Grammar and the Other Components of a Communicating Mind*. Exeter linguistic studies 3. Heidelberg: Groos.
- Fawcett, Robin. 1981. "Generating a Sentence in Systemic Functional Grammar." In *Readings in Systemic Linguistics*, edited by M.A.K. Halliday and James R. Martin, 146–83. London: Batsford Academic and Educational Ltd.
- Fawcett, Robin. 2007. "The Many Types of 'Theme' in English: Their Syntax, Semantics and Discourse Functions." Accessed September 25, 2020. <http://www.isfla.org/Systemics/Print/Papers/Fawcett-ThemePaperv3.pdf>.
- Fawcett, Robin. 2008. *Invitation to Systemic Functional Linguistics Through the Cardiff Grammar: An Extension and Simplification of Halliday's Systemic Functional Grammar*. 3rd ed. London, Oakville: Equinox.
- Fillmore, Charles J. 1968. "The Case for Case." In *Universals in Linguistic Theory*, edited by Emmon Bach and R. T. Harms, 1–90. New York: Holt, Rinehart & Winston.
- Fillmore, Charles J., Christopher R. Johnson, and Miriam R.L. Petruck. 2003. "Background to FrameNet." *International Journal of Lexicography* 16 (3): 235–50.
- Firbas, Jan. 1964a. "From Comparative Word-Order Studies: (Thoughts on V. Mathesius' Conception of the Word-Order System in English Compared with That in Czech)." *Brno Studies in English* 4: 111–28.
- Firbas, Jan. 1964b. "On defining the theme in functional sentence analysis." In *Travaux linguistiques de Prague 1*, edited by Josef Vachek, 267–80. Prague: Académie Tchécoslovaque des Sciences.

- Firbas, Jan. 1987. "On the Delimitation of the Theme in Functional Sentence Perspective." In *Functionalism in Linguistics*, edited by René Dirven and Vilém Fried, 137–56. Amsterdam: John Benjamins.
- Firbas, Jan. 1989. "Degree of Communicative Dynamism and Degrees of Prosodic Prominence." *Brno Studies in English* 18: 21–66.
- Firbas, Jan. 1992. "On Some Basic Problems of Functional Sentence Perspective." In *Advances in Systemic Linguistics: Recent Theory and Practice*, edited by Martin Davies and Louise Ravelli, 167–88. London, New York: Pinter Publishers.
- Ford, Cecilia E. 1993. *Grammar in Interaction: Adverbial Clauses in American English Conversations*. Cambridge: Cambridge University Press.
- Francis, Gill. 1989. "Thematic Selection and Distribution in Written Discourse." *WORD* 40 (1-2): 201–21.
- Francis, Gill. 1990. "Theme in the Daily Press." *Occasional papers in systemic linguistics* 4: 51–87.
- Frank, Austin. 2014. *Mer-Utils.R*. Accessed September 30, 2020. <https://raw.githubusercontent.com/aufrank/R-hacks/master/mer-utils.R>.
- Frawley, William. 1984. "Prolegomenon to a Theory of Translation." In *Translation: Literary, Linguistic and Philosophical Perspectives*, edited by William Frawley, 159–75. Newark: University of Delaware Press.
- Freiwald, Jonas. 2016. "You Say Theme, I Say Thema: A Corpus-Based Approach to Theme in English and German from an SFL Perspective." Staatsarbeit, RWTH Aachen University.
- Fries, Peter H. 1981. "On the Status of Theme: Arguments from Discourse." *Forum Linguisticum* 6 (1): 1–38.
- Fries, Peter H. 1995a. "A Personal View of Theme." In *Thematic Development in English Texts*, edited by Mohsen Ghadessy, 1–19. London, New York: Pinter Publishers.
- Fries, Peter H. 1995b. "Patterns of Information in Initial Position in English." In *Discourse in Society: Systemic Functional Perspectives. Meaning and Choice in Language: Studies for Michael Halliday*, edited by Peter H. Fries and Michael Gregory, 47–66. Norwood: Ablex Publishing Corporation.
- Fries, Peter H. 1995c. "Themes, Methods of Development, and Texts." In *On Subject and Theme: A Discourse Functional Perspective*, edited by Ruqaiya Hasan and Peter H. Fries, 317–59. Amsterdam: John Benjamins.
- Fries, Peter H. 1997. "Theme and New in Written English." In *Functional Approaches to Written Text: Classroom Applications*, edited by Tom Miller, 230–43. Washington.
- Fries, Peter H. 2009. "The Textual Metafunction as a Site for a Discussion of the Goals of Linguistics and Techniques of Linguistic Analysis." In *Text Type and Texture: In Honour of Flo Davies*, edited by Gail Forey and Geoff Thompson, 8–44. London, Oakville: Equinox.
- Fries, Peter H., and Gill Francis. 1992. "Exploring Theme: Problems for Research." *Occasional papers in systemic linguistics* 5: 45–60.

- García García, Marco, Beatrice Primus, and Nikolaus P. Himmelmann. 2018. "Shifting from Animacy to Agentivity." *Theoretical Linguistics* 44 (1-2): 25–39.
- Garretson, Gregory. 2004. "Coding Practices Used in the Project Optimal Typology of Determiner Phrases." <http://npcorpus.bu.edu/html/documentation/index.html>.
- Ghadessy, Mohsen. 1995. "Thematic Development and Its Relationship to Registers and Genres." In *Thematic Development in English Texts*, edited by Mohsen Ghadessy, 129–46. London, New York: Pinter Publishers.
- Ghadessy, Mohsen, and Yanjie Gao. 2000. "Thematic Organization in Parallel Texts: Same and Different Methods of Development." *Text* 20 (4): 461–88.
- Gómez-González, María Ángeles. 2001. *The Theme-Topic Interface: Evidence from English*. Amsterdam: John Benjamins Publishing Company.
- Götze, Lutz, and Ernest W. B. Hess-Lüttich. 2002. *Grammatik der deutschen Sprache: Sprachsystem und Sprachgebrauch*. Gütersloh, München: Wissen Media Verlag.
- Granger, Sylviane. 1996. "From CA to CIA and Back: An Integrated Approach to Computerized Bilingual and Learner Corpora." In *Languages in Contrast: Papers from a Symposium on Text-Based Cross-Linguistic Studies, Lund 4-5 March 1994*, edited by Karin Aijmer, Bengt Altenberg, and Mats Johansson, 37–52. Lund Studies in English 88: Lund University Press.
- Greenberg, Joseph H. 1990. "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In *On Language: Selected Writings of Joseph H. Greenberg*, edited by Keith Denning and Suzanne Kemmer, 40–70. Stanford: Stanford University Press.
- Grimes, Joseph Evans. 1975. *The Thread of Discourse*. The Hague: Mouton.
- Gruber, Jeffrey S. 1967. "Look and See." *Language* 43 (4): 937–47.
- Gundel, Jeanette K. 1988. *The Role of Topic and Comment in Linguistic Theory*. New York, London: Garland Publishing.
- Gundel, Jeanette K., Nancy Hedberg, and Ron Zacharski. 1993. "Cognitive Status and the Form of Referring Expressions in Discourse." *Language* 69 (2): 274–307.
- Haftka, Brigitta. 1978. "Bekanntheit und Neuheit als Kriterien für die Anordnung von Satzgliedern." *Deutsch als Fremdsprache* 15: 157–64.
- Haider, Hubert. 2000. "OV Is More Basic Than VO." In *The Derivation of VO and OV*, edited by Peter Svenonius, 45–68. Amsterdam: John Benjamins Publishing Company.
- Hakulinen, Auli. 1989. "Some Notes on Thematics, Topic, and Typology." In *Text and Discourse Connectedness: Proceedings of the Conference on Connexity and Coherence, Urbino, July 16-21, 1984*. Vol. 16, edited by Emel Sözer, János S. Petőfi, and Maria-Elisabeth Conte, 53–64. Studies in language companion series (SLCS) v. 16. Amsterdam: John Benjamins Publishing Company.
- Halliday, M.A.K. 1967a. "Notes on Transitivity and Theme in English: Part 1." *Journal of Linguistics* 3 (1): 37–81.
- Halliday, M.A.K. 1967b. "Notes on Transitivity and Theme in English: Part 2." *Journal of Linguistics* 3 (2): 199–244.

- Halliday, M.A.K. 1968. "Notes on Transitivity and Theme in English: Part 3." *Journal of Linguistics* 4 (2): 179–215.
- Halliday, M.A.K. 1970. "Language Structure and Language Function." In *New Horizons in Linguistics*, edited by John Lyons, 140–65. Middlesex: Penguin Books.
- Halliday, M.A.K. 1973. *Explorations in the Functions of Language*. London: Edward Arnold.
- Halliday, M.A.K. 1974. "Interview with M. A. K. Halliday." In *Discussing Language*, edited by Herman Parret, 81–120. The Hague, Paris: Mouton.
- Halliday, M.A.K. 1976. "Theme and Information in the English Clause." In *Halliday: System and Function in Language: Selected Papers*, edited by Gunther Kress, 174–88. Oxford: Oxford University Press.
- Halliday, M.A.K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Arnold.
- Halliday, M.A.K. 1979. "Modes of Meaning and Modes of Expression: Types of Grammatical Structure, and Their Determination by Different Semantic Functions." In *Function and Context in Linguistic Analysis: A Festschrift for William Haas*, edited by David J. Allerton, Edward Carney, and David Holcroft, 57–79. Cambridge: Cambridge University Press.
- Halliday, M.A.K. 1984. "On the Ineffability of Grammatical Categories." In *The Tenth LACUS Forum*, edited by Alan Manning, Pierre Martin, and Kim McCalla, 3–18. Columbia: Hornbeam Press.
- Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*. 1st ed. London: Edward Arnold.
- Halliday, M.A.K. 1991. "Corpus Studies and Probabilistic Grammar." In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, edited by Karin Aijmer and Bengt Altenberg, 30–43. London, New York: Longman.
- Halliday, M.A.K. 1993. "The Construction of Knowledge and Value in the Grammar of Scientific Discourse: Charles Darwin's *the Origin of Species*." In *Writing Science: Literacy and Discursive Power*, edited by M.A.K. Halliday and James R. Martin, 86–105. Pittsburgh series in composition, literacy, and culture. Pittsburgh: University of Pittsburgh Press.
- Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. 2nd ed. London: Arnold.
- Halliday, M.A.K. 1998. "On the Grammar of Pain." *Functions of Language* 5 (1): 1–32.
- Halliday, M.A.K., and William S. Greaves. 2008. *Intonation in the Grammar of English*. London, Oakville: Equinox.
- Halliday, M.A.K., and Ruqaiya Hasan. 1985. *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*. Geelong: Deakin University Press.
- Halliday, M.A.K., and Christian M.I.M. Matthiessen. 1999. *Construing Experience Through Meaning: A Language-Based Approach to Cognition*. London, New York: Continuum.
- Halliday, M.A.K., and Christian M.I.M. Matthiessen. 2014. *Halliday's Introduction to Functional Grammar*. 4th ed. London, New York: Routledge.

- Halliday, M.A.K., and Edward McDonald. 2004. "Metafunctional Profile of the Grammar of Chinese." In *Language Typology: A Functional Perspective*, edited by Alice Caffarel, James R. Martin, and Christian M.I.M. Matthiessen, 305–96. *Current Issues in Linguistic Theory* 253. Amsterdam: John Benjamins.
- Hansen, Silvia. 2003. *The Nature of Translated Text: An Interdisciplinary Methodology for the Investigation of the Specific Properties of Translations*. Saarbrücken: DFKI/Universität des Saarlandes.
- Hansen, Silvia, and Elke Teich. 1999. "Kontrastive Analyse von Übersetzungskorpora: ein funktionales Modell." In *Multilinguage Corpora. Codierung, Strukturierung, Analyse: 11. Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung*, edited by Jost Gippert, 311–22. Prague: enigma corporation.
- Hansen-Schirra, Silvia. 2011. "Between Normalization and Shining-Through." In *Multilingual Discourse Production: Diachronic and Synchronic Perspectives*. Vol. 12, edited by Svenja Kranich, 133–62. *Hamburg Studies on Multilingualism* 12. Amsterdam: John Benjamins Publishing Company.
- Hansen-Schirra, Silvia, and Stella Neumann. 2012. "3 Corpus Enrichment, Representation, Exploitation, and Quality Control." In *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*, edited by Silvia Hansen-Schirra, Stella Neumann, and Erich Steiner, 35–52. Berlin: de Gruyter Mouton.
- Hansen-Schirra, Silvia, Stella Neumann, and Erich Steiner. 2007. "Cohesive Explicitness and Explication in an English-German Translation Corpus." *Languages in Contrast* 7 (2): 241–65.
- Hansen-Schirra, Silvia, Stella Neumann, and Erich Steiner, eds. 2012. *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*. Berlin: de Gruyter Mouton.
- Hasan, Ruqaiya, and Peter H. Fries, eds. 1995a. *On Subject and Theme: A Discourse Functional Perspective*. Amsterdam: John Benjamins.
- Hasan, Ruqaiya, and Peter H. Fries. 1995b. "Reflections on Subject and Theme: An Introduction." In *On Subject and Theme: A Discourse Functional Perspective*, edited by Ruqaiya Hasan and Peter H. Fries, xiii–xlv. Amsterdam: John Benjamins.
- Haspelmath, Martin. 2006. "Against Markedness (And What to Replace It with)." *Journal of Linguistics* 42 (1): 25–70. <https://doi.org/10.1017/S0022226705003683>.
- Hasselgård, Hilde. 1997. "Sentence Openings in English and Norwegian." In *Corpus-Based Studies in English: Papers from the 17th International Conference on English Language Research on Computerized Corpora*, edited by Magnus Ljung, 3–20. Amsterdam: Rodopi.
- Hasselgård, Hilde. 1998. "Thematic Structure in Translation Between English and Norwegian." In *Corpora and Cross-Linguistic Research: Theory, Method, and Case Studies*, edited by Stig Johansson and Signe Oksefjell, 145–67. *Language and Computers: Studies in Practical Linguistics* 24. Amsterdam, Atlanta: Rodopi.
- Hasselgård, Hilde. 2000. "English Multiple Themes in Translation." In *Contrastive Studies in Syntax*, edited by Alex Klinge, 11–38. *Copenhagen Studies in Language* 25. Frederiksberg: Samfundslitteratur.

- Hasselgård, Hilde. 2004. "Thematic Choice in English and Norwegian." *Functions of Language* 11 (2): 187–212. <https://doi.org/10.1075/fol.11.2.03has>.
- Hatim, Basil, and Ian Mason. 1997. *The Translator as Communicator*. London, New York: Routledge.
- Hawkins, John A. 1986. *A Comparative Typology of English and German: Unifying the Contrasts*. London, Sydney: Croom Helm.
- Hawkins, John A. 1992. "Syntactic Weight versus Information Structure in Word Order Variation." In *Informationsstruktur und Grammatik*, edited by Joachim Jacobs, 196–219. Opladen: Westdeutscher Verlag.
- Hawkins, John A. 2000. "The Relative Order of Prepositional Phrases in English: Going Beyond Manner–Place–Time." *Language Variation and Change* 11 (3): 231–66.
- Heidolph, Karl Erich, Walter Flämig, and Wolfgang Motsch. 1981. *Grundzüge einer deutschen Grammatik*. Berlin: Akademie-Verlag.
- Heilmann, Arndt, Tatiana Serbina, Jonas Freiwald, and Stella Neumann. 2020. "Animacy and Agentivity of Subject Themes in English-German Translation." *Lingua* 102813. <https://doi.org/10.1016/j.lingua.2020.102813>.
- Hoberg, Ursula. 1981. *Die Wortstellung in der geschriebenen deutschen Gegenwartssprache*. München: Max Hueber Verlag.
- Hockett, Charles F. 1958. *A Course in Modern Linguistics*. New York: Macmillan.
- Höhle, Tilman N. 2019a. "Explikationen für 'normale Betonung' und 'normale Wortstellung'." In *Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle*, edited by Stefan Müller, Marga Reis, and Frank Richter. 2nd ed., revised, 107–91. Berlin: Language Science Press.
- Höhle, Tilman N. 2019b. "Topologische Felder." In *Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle*, edited by Stefan Müller, Marga Reis, and Frank Richter. 2nd ed., revised, 7–89. Berlin: Language Science Press.
- Hopper, Paul J., and Sandra A. Thompson. 1980. "Transitivity in Grammar and Discourse." *Language* 56 (2): 251–99. <https://doi.org/10.2307/413757>.
- House, Juliane. 2008. "Beyond Intervention: Universals in Translation?" *Trans-kom. Zeitschrift für Translationswissenschaft und Fachkommunikation* 1 (1): 6–19.
- Huber, Walter, and Werner Kummer. 1974. *Transformationelle Syntax des Deutschen I*. München: Wilhelm Fink Verlag.
- Huddleston, Rodney. 1988. "Constituency, Multi-Functionality and Grammaticalization in Halliday's Functional Grammar." *Journal of Linguistics* 24 (1): 137–74. <https://doi.org/10.1017/S0022226700011592>.
- Huddleston, Rodney. 1991. "Further Remarks on Halliday's Functional Grammar: A Reply to Matthiessen and Martin." *Occasional papers in systemic linguistics* 5: 75–129.
- Hudson, Richard. 1986. "Systemic Grammar: Review of an Introduction to Functional Grammar by M. A. K. Halliday, and Systemic Linguistics by C. S. Butler." *Linguistics* 24: 791–815.

- Johansson, Mats. 2002. "Clefts in English and Swedish: A Contrastive Study of IT-Clefts and WH-Clefts in Original Texts and Translations." PhD Thesis, University of Lund.
- Johansson, Stig. 1997. "Using the English-Norwegian Parallel Corpus: A Corpus for Contrastive Analysis and Translation Studies." In *International Conference on Practical Applications in Language Corpora: Łódź, Poland, 10-14 April 1997: Proceedings*, edited by Barbara Lewandowska-Tomaszczyk and Patrick J. Melia, 282–96. Łódź: Łódź University Press.
- Johansson, Stig. 2004. "Viewing Languages Through Multilingual Corpora, with Special Reference to the Generic Person in English, German, and Norwegian." *LiC* 4 (2): 261–80. <https://doi.org/10.1075/lic.4.2.05joh>.
- Kade, Otto. 1968. *Zufall und Gesetzmäßigkeit in der Übersetzung*. Leipzig: Verlag Enzyklopädie.
- Kast, Marlene. 2012. "8 Variation Within the Grammatical Function 'Subject' in English-German and German-English Translations." In *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*, edited by Silvia Hansen-Schirra, Stella Neumann, and Erich Steiner, 147–60. Berlin: de Gruyter Mouton.
- Katz, Jerrold J. 1980. "Chomsky on Meaning." *Language* 56 (1): 1–41. <https://doi.org/10.2307/412641>.
- Kenny, Dorothy. 2001. *Lexis and Creativity in Translation: A Corpus-Based Study*. Manchester: St. Jerome Publishing.
- Kies, Daniel. 1988. "Marked Themes with and Without Pronominal Reinforcement: Their Meaning and Distribution in Discourse." In *Pragmatics, Discourse and Text: Some Systemically-Inspired Approaches*, edited by Erich Steiner and Robert Veltman, 47–75. London: Pinter Publishers.
- Kim, Mira. 2007. "A Discourse Based Study on Theme in Korean and Textual Meaning in Translation." PhD Thesis, Macquarie University.
- Kim, Mira, and Zhi Huang. 2012. "Theme Choices in Translation and Target Readers' Reactions to Different Theme Choices." *T&I Review* 2: 79–111.
- Kim, Mira, and Christian M.I.M. Matthiessen. 2015. "Ways to Move Forward in Translation Studies: A Textual Perspective." *Target* 27 (3): 335–50. <https://doi.org/10.1075/target.27.3.01kim>.
- Kirkwood, Henry W. 1970. "Some Systemic Means of Functional Sentence Perspective in English and German." *IRAL - International Review of Applied Linguistics in Language Teaching* 8 (2): 103–114.
- Klein, Yvonne. 2007. "Übersetzungsspezifische Eigenschaften: Eine korpusbasierte Studie am Beispiel der Kohäsion." Dissertation, Universität des Saarlandes.
- Koller, Werner. 2011. *Einführung in die Übersetzungswissenschaft*. 8th ed., revised. Tübingen: Francke Verlag.
- König, Ekkehard, and Volker Gast. 2009. *Understanding English-German Contrasts*. 2nd ed., revised. Berlin: Erich Schmidt Verlag.

- Königs, Karin. 2011. *Übersetzen Englisch - Deutsch: Lernen mit System*. 3rd ed., revised. München: Oldenbourg Verlag.
- Korner-Nievergelt, Franzi, Tobias Roth, Stefanie von Felten, Jérôme Guelat, Bettina Almasi, and Pius Korner-Nievergelt. 2015. *Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS and Stan*. Amsterdam: Elsevier.
- Kruger, Haidee. 2017. "The Effects of Editorial Intervention: Implications for Studies of the Features of Translated Language." In *Empirical Translation Studies*, edited by Gert de Sutter, Marie-Aude Lefer, and Isabelle Delaere, 113–55. Berlin/Boston: De Gruyter.
- Kruger, Haidee, and Bertus van Rooy. 2012. "Register and the Features of Translated Language." *Across Languages and Cultures* 13 (1): 33–65.
- Kuno, Susumu. 1973. *The Structure of the Japanese Language*. Chicago, London: The MIT Press.
- Kunz, Kerstin A. 2010. *Variation in English and German Nominal Coreference: A Study of Political Essays*. Frankfurt am Main: Peter Lang.
- Lakoff, George. 1977. "Linguistic Gestalts." *Papers from the 13th Regional Meeting of the Chicago Linguistic Society*, 236–87.
- Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: Cambridge University Press.
- Lasch, Alexander. 2016. *Nonagentive Konstruktionen des Deutschen*. Sprache und Wissen (SuW) 25. Berlin/Boston: De Gruyter.
- Laviosa-Braithwaite, Sara. 1996. "The English Comparable Corpus (ECC): A Resource and a Methodology for the Empirical Study of Translation." PhD Thesis, Centre for Translation and Intercultural Studies, University of Manchester.
- Leech, Geoffrey. 1991. "The State of the Art in Corpus Linguistics." In *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, edited by Karin Aijmer and Bengt Altenberg, 8–29. London, New York: Longman.
- Lenerz, Jürgen. 1977. *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Gunter Narr.
- Li, Suqin. 2015. "A Description of the Theme Structure of Bai Clause." *OJML* 5 (6): 528–38. <https://doi.org/10.4236/ojml.2015.56046>.
- Lindeberg, Ann-Charlotte. 1985. "Abstraction Levels in Student Essays." *Text* 5: 327–46.
- Liu, Xiangjun, and Xiaohu Yang. 2013. "Thematic Progression in English–Chinese Translation of Argumentative Classics: A Quantitative Study of Francis Bacon's 'Of Studies' and Its 11 Chinese Translations." *Perspectives* 21 (2): 272–88. <https://doi.org/10.1080/0907676X.2011.615940>.
- Lødrup, Helve. 1993. "Subjects and Thematic Roles in English and Norwegian." *Norsk lingvistisk tidsskrift: NLT* 11: 105–24.
- Lötscher, Andreas. 1983. *Satzakzent und funktionale Satzperspektive im Deutschen*. Tübingen: Max Niemeyer Verlag.

- Lyons, John. 1968. *Introduction to Theoretical Linguistics*. Cambridge: Cambridge University Press.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Martin, James R. 1983. "Participant Identification in English, Tagalog and Kate." *Australian Journal of Linguistics* 3 (1): 45–74.
- Martin, James R. 1986. "Grammaticalising Ecology: The Politics of Baby Seals and Kangaroos." In *Semiotics, Ideology, Language*, edited by Terry Threadgold, E. A. Grosz, Gunther Kress, and M.A.K. Halliday, 225–68. Sydney Studies in Society and Culture 3. Sydney: Sydney Association for Studies in Society and Culture.
- Martin, James R. 1989. *Factual Writing: Exploring and Challenging Social Reality*. 2nd ed. London: Oxford University Press.
- Martin, James R. 1992a. *English Text: System and Structure*. Amsterdam: John Benjamins Publishing Company.
- Martin, James R. 1992b. "Theme, Method of Development and Existentiality: The Price of Reply." *Occasional papers in systemic linguistics* 6: 147–83.
- Martin, James R. 1995. "More Than What the Message Is About: English Theme." In *Thematic Development in English Texts*, edited by Mohsen Ghadessy, 223–58. London, New York: Pinter Publishers.
- Martin, James R. 2004. "Metafunctional Profile of the Grammar of Tagalog." In *Language Typology: A Functional Perspective*, edited by Alice Caffarel, James R. Martin, and Christian M.I.M. Matthiessen, 255–304. Current Issues in Linguistic Theory 253. Amsterdam: John Benjamins.
- Martin, James R., Christian M.I.M. Matthiessen, and Clare Painter. 1997. *Working with Functional Grammar*. London: Arnold.
- Martin, James R., Clare Painter, and Christian M.I.M. Matthiessen. 2006. *Working with Functional Grammar*. London: Arnold.
- Mathesius, Vilém. (1939) 1975. "O Tak Zvaném Aktuálním Členění Věty: [On the so-called Functional Sentence Perspective]." In *Harvard Studies in Syntax and Semantics*, edited by Susumu Kuno, 467–80. Cambridge: Harvard University Press.
- Matthiessen, Christian M.I.M. 1988. "Representational Issues in Systemic Functional Grammar." In *Systemic Functional Approaches to Discourse*, edited by James D. Benson and William S. Greaves, 136–75. Norwood: Ablex Publishing Corporation.
- Matthiessen, Christian M.I.M. 1992. "Interpreting the Textual Metafunction." In *Advances in Systemic Linguistics: Recent Theory and Practice*, edited by Martin Davies and Louise Ravelli, 37–81. London, New York: Pinter Publishers.
- Matthiessen, Christian M.I.M. 1993. "Register in the Round: Diversity in a Unified Theory of Register Analysis." In *Register Analysis: Theory and Practice*, edited by Mohsen Ghadessy, 221–92. London: Pinter.
- Matthiessen, Christian M.I.M. 1995a. *Lexicogrammatical Cartography: English Systems*. Tokyo: International Language Sciences Publishers.

- Matthiessen, Christian M.I.M. 1995b. "THEME as an Enabling Resource in Ideational Knowledge Construction." In *Thematic Development in English Texts*, edited by Mohsen Ghadessy, 20–55. London, New York: Pinter Publishers.
- Matthiessen, Christian M.I.M., and John A. Bateman. 1991. *Systemic Linguistics and Text Generation: Experience from Japanese and English*. London: Pinter.
- Matthiessen, Christian M.I.M., and James R. Martin. 1991. "A Response to Huddleston's Review of Halliday's Introduction to Functional Grammar." *Occasional papers in systemic linguistics* 5: 5–74.
- Matthiessen, Christian M.I.M., Kazuhiro Teruya, and Marvin Lam. 2010. *Key Terms in Systemic Functional Linguistics*. London: Continuum.
- Mauranen, Anna. 1993. "Theme and Prospection in Written Discourse." In *Text and Technology*, edited by Mona Baker, Gill Francis, and Elena Tognini-Bonelli, 95–114. Amsterdam: John Benjamins Publishing Company.
- Mauranen, Anna. 2004. "Corpora, Universals and Interference." In *Translation Universals*. Vol. 48, edited by Anna Mauranen and Pekka Kujamäki, 65–82. Amsterdam: John Benjamins Publishing Company.
- Mauranen, Anna. 2007. "Universal Tendencies in Translation." In *Incorporating Corpora: The Linguist and the Translator*, edited by Gunilla von Anderman and Margaret Rogers, 32–48. Clevedon: Multilingual Matters.
- McEnery, Tony, and Andrew Wilson. 2001. *Corpus Linguistics: An Introduction*. 2nd ed. Edinburgh: Edinburgh University Press.
- Mode, Donatien. 1987. *Syntax des Vorfelds: Zur Systematik und Didaktik der deutschen Wortstellung*. Reihe Germanistische Linguistik 74. Tübingen: Max Niemeyer Verlag.
- Montemayor-Borsinger, Ann. 2009. "Text-Type and Texture: The Potential of Theme for the Study of Research Writing Development." In *Text Type and Texture: In Honour of Flo Davies*, edited by Gail Forey and Geoff Thompson, 108–24. London, Oakville: Equinox.
- Müller, Gereon. 1999. "Optimality, Markedness, and Word Order in German." Accessed September 26, 2020. <https://home.uni-leipzig.de/muellerg/mu32.pdf>.
- Müller, Stefan. 2005. "Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung." *Linguistische Berichte* 203: 297–330.
- Müller, Stefan, Marga Reis, and Frank Richter, eds. 2019. *Beiträge zur deutschen Grammatik: Gesammelte Schriften von Tilman N. Höhle*. 2nd ed., revised. Berlin: Language Science Press.
- Munday, Jeremy. 1998. "Problems of Applying Thematic Analysis to Translation Between Spanish and English." *Cadernos de Tradução* 1 (3): 183–213.
- Munday, Jeremy. 2001. *Introducing Translation Studies: Theories and Applications*. London: Routledge.
- Neumann, Stella. 2003. *Textsorten und Übersetzen: Eine Korpusanalyse englischer und deutscher Reiseführer*. Frankfurt am Main: Peter Lang.

- Neumann, Stella. 2012. "11 Register-Induced Properties of Translations." In *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*, edited by Silvia Hansen-Schirra, Stella Neumann, and Erich Steiner, 191–209. Berlin: de Gruyter Mouton.
- Neumann, Stella. 2014. *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German*. Trends in Linguistics, Studies and Monographs 251. Berlin, Boston: De Gruyter.
- Neumann, Stella, Jonas Freiwald, and Arndt Heilmann. forthcoming. "On the Use of Multiple Methods in Empirical Translation Studies: A Combined Corpus and Experimental Analysis of Subject Identifiability in English and German." In *Extending the Scope of Corpus-Based Translation Studies*, edited by Sylviane Granger and Marie-Aude Lefer. London, New York: Bloomsbury.
- Neumann, Stella, and Silvia Hansen-Schirra. 2012. "2 Corpus Methodology and Design." In *Cross-Linguistic Corpora for the Study of Translations: Insights from the Language Pair English-German*, edited by Silvia Hansen-Schirra, Stella Neumann, and Erich Steiner, 21–33. Berlin: de Gruyter Mouton.
- Newmark, Peter. 1988. *A Textbook of Translation*. New York: Prentice Hall International.
- Nida, Eugene Albert. 1964. *Toward a Science of Translating: With Special Reference to Principles and Procedures Involved in Bible Translating*. Leiden: Brill.
- Nida, Eugene Albert, and Charles R. Taber. 1969. *Theorie und Praxis des Übersetzens: Unter besonderer Berücksichtigung der Bibelübersetzung*. London: Weltbund der Bibelgesellschaften.
- Niemietz, Paula, Stella Neumann, and Jonas Freiwald. 2017. "Shifts in Theme and Subject Realization in English-German Translation." In *Challenging Boundaries in Linguistics: Systemic Functional Perspectives*, edited by Stella Neumann, Rebekah Wegener, Jennifer Fest, Paula Niemietz, and Nicole Hützen, 331–57. Frankfurt am Main: Peter Lang.
- O'Donnell, Mick. 2008a. "Demonstration of the UAM CorpusTool for Text and Image Annotation." In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies Demo Session - HLT '08*, edited by Jimmy Lin, 13–16. Morristown, NJ, USA: Association for Computational Linguistics.
- O'Donnell, Mick. 2008b. "The UAM CorpusTool: Software for Corpus Annotation." In *Applied Linguistics Now: Understanding Language and Mind: La Lingüística Aplicada Actual: Comprendiendo El Lenguaje Y La Mente*, edited by Carmen M. Bretones Callejas, José F. F. Sánchez, José R. Ibáñez Ibáñez, María E. García Sánchez, Cortés de los Ríos, María Enriqueta, María Sagrario Salaberri Ramiro, María Soledad Cruz Martínez, Nobel A. Perdu Honeyman, and Blasina Cantizano Márquez, 1433–48. Almería: Universidad de Almería.
- Oettinger, Anthony G. 1960. *Automatic Language Translation: Lexical and Technical Aspects, with Particular Reference to Russian*. Harvard monographs in applied science 8. Cambridge: Harvard University Press.
- Olohan, Maeve. 2002. "Corpus Linguistics and Translation Studies: Interaction and Reaction." *Linguistica Antverpiensia* 1: 419–29.

- Olsen, Leslie, and Rod Johnson. 1989. "A Discourse-Based Approach to the Assessment of Readability." *Linguistics and Education* 1: 207–31.
- Paul, Hermann. 1968. *Deutsche Grammatik*. 5th ed. 3rd vol. Halle: Niemeyer.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. www.R-project.org.
- Ramsay, Violetta. 1987. "The Functional Distribution of Preposed and Postposed *If* and *When* Clauses in Written Discourse." In *Coherence and Grounding in Discourse: Outcome of a Symposium, Eugene, Oregon, June 1984*, edited by Russell S. Tomlin, 383–408. Amsterdam: John Benjamins Publishing Company.
- Rashidi, Linda S. 1992. "Towards an Understanding of the Notion of Theme: An Example from Dari." In *Advances in Systemic Linguistics: Recent Theory and Practice*, edited by Martin Davies and Louise Ravelli, 189–204. London, New York: Pinter Publishers.
- Ravelli, Louise. 1995. "A Dynamic Perspective: Implications for Metafunctional Interaction and an Understanding of Theme." In *On Subject and Theme: A Discourse Functional Perspective*, edited by Ruqaiya Hasan and Peter H. Fries, 187–234. Amsterdam: John Benjamins.
- Rørvik, Sylvi. 2004. "Thematic Progression in Translation of Fiction from English to Norwegian." In *Translation and Corpora: Selected Papers from the Göteborg Symposium 18-19 October 2003*, edited by Karin Aijmer and Hilde Hasselgård, 149–61. Göteborg.
- Rose, David. 2001. "Some Variations in Theme Across Languages." Accessed January 13, 2021. <https://www.researchgate.net/publication/252348010>.
- Rose, David. 2004. "Metafunctional Profile of the Grammar of Pitjantjatjara." In *Language Typology: A Functional Perspective*, edited by Alice Caffarel, James R. Martin, and Christian M.I.M. Matthiessen, 479–536. *Current Issues in Linguistic Theory* 253. Amsterdam: John Benjamins.
- Rosengren, Inger. 1993. "Wahlfreiheit mit Konsequenzen - Scrambling, Topikalisierung und FHG im Dienste der Informationsstrukturierung." In *Wortstellung und Informationsstruktur*, edited by Marga Reis, 251–312. Tübingen: Max Niemeyer Verlag.
- Ross, John R. 1970. "Gapping and the Order of Constituents." In *Progress in Linguistics: A Collection of Papers*, edited by Karl E. Heidolph and Manfred Bierwisch, 249–59. The Hague: Mouton.
- Ryshina-Pankova, Marianna V. 2006. "Creating Textual Worlds in Advanced Learner Writing: The Role of Complex Theme." In *Advanced Language Learning: The Contribution of Halliday and Vygotsky*, edited by Heidi Byrnes, 164–83. London, New York: Continuum.
- Sampson, Geoffrey. 1980. *Schools of Linguistics*. Stanford: Stanford University Press.
- Schachter, Paul, and Fe T. Otanes. 1972. *Tagalog Reference Grammar*. Berkeley: University of California Press.
- Schiffrin, Deborah. 1992. "Conditionals as Topics in Discourse." *Linguistics* 30 (1): 165–97. <https://doi.org/10.1515/ling.1992.30.1.165>.

- Serbina, Tatiana. 2015. "A Construction Grammar Approach to the Analysis of Translation Shifts: A Corpus-Based Study." PhD Thesis, RWTH Aachen University.
- Seuren, Pieter A. M. 1998. *Western Linguistics. An Historical Introduction*. Oxford: Blackwell.
- Shlesinger, Miriam. 1991. "Interpreter Latitude Vs. Due Process. Simultaneous and Consecutive Interpretation in Multilingual Trials." In *Empirical Research in Translation and Intercultural Studies: Selected Papers of the Transif Seminar, Savonlinna, 1988*, edited by Sonja Tirkkonen-Condit, 147–55. Tübingen: Gunter Narr.
- Silverstein, Michael. 1976. "Hierarchy of Features and Ergativity." In *Grammatical Categories in Australian Languages*, edited by R. M. W. Dixon, 112–71. Canberra: Australian Institute of Aboriginal Studies.
- Singmann, Henrik, Ben Bolker, Jake Westfall, and Frederik Aust. 2017. *Afex: Analysis of Factorial Experiments*. <https://CRAN.R-project.org/package=afex>.
- Sohn, Ho-min. 1980. "Theme-Prominence in Korean." *Korean Linguistics* 2:1–19. <https://doi.org/10.1075/kl.2.01hms>.
- Stainton, Caroline. 1993. "Metadiscourse and the Analytical Text: A Genre-Based Approach to Children's Written Discourse." PhD Thesis, University of Manchester.
- Steiner, Erich. 1997. "An Extended Register Analysis as a Form of Text Analysis for Translation." In *Modelle der Translation: Models of Translation: Festschrift für Albrecht Neubert*, edited by Heide Schmidt and Gerd Wotjak, 235–56. Leipziger Schriften 2. Frankfurt am Main: Vervuert Verlagsgesellschaft.
- Steiner, Erich. 2001a. "Intralingual and Interlingual Versions of a Text - How Specific Is the Notion of 'Translation'?" In *Exploring Translation and Multilingual Text Production: Beyond Content*, edited by Erich Steiner and Colin Yallop, 161–90. Berlin, New York: Mouton de Gruyter.
- Steiner, Erich. 2001b. "Translations English-German: Investigating the Relative Importance of Systemic Contrasts and the Text-Type "Translation". " SPRIKreports. Reports from the project Languages in Contrast (Språk i kontrast).
- Steiner, Erich. 2004. *Translated Texts: Properties, Variants, Evaluations*. Saarbrücker Beiträge zur Sprach- und Translationswissenschaft 4. Frankfurt am Main: Lang.
- Steiner, Erich, and Wiebke Ramm. 1995. "On Theme as a Grammatical Notion for German." *Functions of Language* 2 (1): 57–93. <https://doi.org/10.1075/fol.2.1.04ste>.
- Steiner, Erich, and Elke Teich. 2004. "Metafunctional Profile of the Grammar of German." In *Language Typology: A Functional Perspective*, edited by Alice Caffarel, James R. Martin, and Christian M.I.M. Matthiessen, 139–84. Current Issues in Linguistic Theory 253. Amsterdam: John Benjamins.
- Swart, Peter de. 2014. "Prepositional Inanimates in Dutch: A Paradigmatic Case of Differential Object Marking." *Linguistics* 52 (2). <https://doi.org/10.1515/ling-2013-0069>.
- Taylor, Christopher. 1993. "Systemic Linguistics and Translation." *Occasional papers in systemic linguistics* 7: 87–103.

- Teich, Elke. 2003. *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts*. Text, translation, computational processing 5. Berlin, New York: Mouton de Gruyter.
- Teruya, Kazuhiro. 2004. "Metafunctional Profile of the Grammar of Japanese." In *Language Typology: A Functional Perspective*, edited by Alice Caffarel, James R. Martin, and Christian M.I.M. Matthiessen, 185–254. *Current Issues in Linguistic Theory* 253. Amsterdam: John Benjamins.
- Thai, Minh Duc. 2004. "Metafunctional Profile of the Grammar of Vietnamese." In *Language Typology: A Functional Perspective*, edited by Alice Caffarel, James R. Martin, and Christian M.I.M. Matthiessen, 397–431. *Current Issues in Linguistic Theory* 253. Amsterdam: John Benjamins.
- Thompson, Geoff. 2006. "Theme in Text." In *Encyclopedia of Language and Linguistics*. Vol. 12, edited by Keith Brown. 2nd ed. 14 vols, 658–68. Oxford: Elsevier.
- Thompson, Geoff. 2007. "Unfolding Theme: The Development of Clausal and Textual Perspectives on Theme." In *Continuing Discourse on Language: A Functional Perspective*, edited by Ruqaiya Hasan, Christian Matthiessen, and Jonathan Webster, 671–96. London: Equinox.
- Thompson, Geoff. 2014. *Introducing Functional Grammar*. 3rd ed. London: Routledge.
- Thompson, Geoff, and Susan Thompson. 2009. "Theme, Subject and the Unfolding of Text." In *Text Type and Texture: In Honour of Flo Davies*, edited by Gail Forey and Geoff Thompson, 45–69. London, Oakville: Equinox.
- Thompson, Sandra A. 1978. "Modern English from a Typological Point of View: Some Implications of the Function of Word Order." *Linguistische Berichte* 54: 19–35.
- Thomson, Elizabeth A. 2005. "Theme Unit Analysis: A Systemic-Functional Treatment of Textual Meanings in Japanese." *Functions of Language* 12 (2): 151–79.
- Toury, Gideon. 1980. *In Search of a Theory of Translation*. Tel Aviv: Tel Aviv University.
- Toury, Gideon. 2012. *Descriptive Translation Studies – and Beyond*. 2nd ed., revised. Amsterdam: John Benjamins Publishing Company.
- Trávníček, František. 1962. "O tak zvaním aktualním clenění větám." [On so-called functional sentence perspective]. *Slovo a slovesnost* 22: 163–71. As cited in Vasconcellos, Muriel Havel de. 1985. "Theme and Focus: Cross-Language Comparison via Translations from Extended Discourse." PhD Thesis, Georgetown University.
- Trosborg, Anna. 1997. "Translating Hybrid Political Texts." In *Analysing Professional Genres*, edited by Anna Trosborg, 145–58. Amsterdam: John Benjamins Publishing Company.
- Trubetzkoy, Nikolaj. 1931. "Die phonologischen Systeme." *Travaux du Cercle Linguistique de Prague* 4: 96–116.
- Uszkoreit, Hans. 1987. *Word Order and Constituent Structure in German*. Stanford: Center for the Study of Language and Information.
- Uzonyi, Pál, and Viktória Dabóczi. 2016. "Quantitative Korpusuntersuchungen zur Phrasenstruktur von deutschen und ungarischen linken Feldern." In *Variation im europäischen*

- Kontrast: Untersuchungen zum Satzanfang im Deutschen, Französischen, Norwegischen, Polnischen und Ungarischen*, edited by Martine Dalmas, Cathrine Fabricius-Hansen, and Horst Schwinn, 53–74. Berlin, Boston: Walter de Gruyter.
- van de Velde, Marc. 1978. "Zur mehrfachen Vorfeldbesetzung im Deutschen." In *Wortstellung und Bedeutung: Akten des 12. Linguistischen Kolloquiums*, edited by Maria-Elisabeth Conte, Anna G. Ramat, and Paolo Ramat, 131–41. Tübingen: Max Niemeyer Verlag.
- Vanderauwera, Ria. 1985. *Dutch Novels Translated into English: The Transformation of a 'Minority' Literature*. Amsterdam: Rodopi.
- Vasconcellos, Muriel Havel de. 1985. "Theme and Focus: Cross-Language Comparison via Translations from Extended Discourse." PhD Thesis, Georgetown University.
- Vasconcellos, Muriel Havel de. 2008. "Text and Translation: The Role of Theme and Information." *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies* 27: 45–66.
- Verstraete, Jean-Christophe. 2004. "Initial and Final Position for Adverbial Clauses in English: The Constructional Basis of the Discursive and Syntactic Differences." *Linguistics* 42 (4): 819–53.
- Vinay, Jean-Paul, and Jean Darbelnet. 1995. *Comparative Stylistics of French and English: A Methodology for Translation*. Amsterdam: John Benjamins Publishing Company.
- Virtanen, Tuija. 2014. "Sentence-Initial Adverbials in Written Texts: On Discourse Functions and Cognitive Motivations." In *Adverbials in Use: From Predicative to Discourse Functions*, edited by Laure Sarda, Shirley Carter-Thomas, Benjamin Fagard, and Michael Charolles, 103–32. Louvain-La-Neuve: Presses universitaires de Louvain.
- Volansky, Vered, Noam Ordan, and Shuly Wintner. 2015. "On the Features of Translationese." *Digital Scholarship in the Humanities* 30 (1): 98–118.
- Voyles, Joseph. 1978. "German as an SOV Language." *Linguistische Berichte* 54: 1–18.
- Ward, Gregory L., and Ellen F. Prince. 1991. "On the Topicalization of Indefinite NPs." *Journal of Pragmatics* 16: 167–77.
- Waugh, Linda. 1982. "Marked and Unmarked: A Choice Between Unequals in Semiotic Structure." *Semiotica* 38 (3-4): 299–318.
- Weinrich, Harald. 1993. *Textgrammatik der deutschen Sprache*. Mannheim: Dudenverlag.
- Whittaker, Rachel. 1995. "Theme, Processes and the Realization of Meanings in Academic Articles." In *Thematic Development in English Texts*, edited by Mohsen Ghadessy, 105–28. London, New York: Pinter Publishers.
- Wiechmann, Daniel, and Elma Kerz. 2013. "The Positioning of Concessive Adverbial Clauses in English: Assessing the Importance of Discourse-Pragmatic and Processing-Based Constraints." *English Language and Linguistics* 17 (1): 1–23.
- Winter, Werner. 1961. "Relative Häufigkeit syntaktischer Erscheinungen als Mittel zur Abgrenzung von Stilarten." *Phonetica* 7: 193–216.
- Xiao, Qun. 1991. *Toward Thematic Selection in Different Genre: Fables and Recipes*.

- Yan, Fang, Edward McDonald, and Cheng Musheng. 1995. "On Theme in Chinese: From Clause to Discourse." In *On Subject and Theme: A Discourse Functional Perspective*, edited by Ruqaiya Hasan and Peter H. Fries, 235–73. Amsterdam: John Benjamins.
- Zaenen, Annie, Jean Carletta, Gregory Garretson, Joan Bresnan, Andrew Koontz-Garboden, Tatiana Nikitina, M. Catherine O'Connor, and Tom Wasow. 2004. "Animacy Encoding in English: Why and How." In *DiscAnnotation '04: Proceedings of the 2004 ACL Workshop on Discourse Annotation*, 118–25.
- Zifonun, Gisela, Ludger Hoffmann, and Bruno Strecker. 1997. *Grammatik der deutschen Sprache*. Berlin: De Gruyter.

Appendix

Chapter 7: Theme in German Originals

	Overall	FICTION	INSTR	SPEECH	TOU
Total Clauses	4901	1306	755	1580	1261
Total processes	3158	856	408	1085	809
Material	1122	280	217	314	311
Subject Actor	769	241	105	231	192
Subject Goal	179	9	85	42	43
Subject Initiator	62	15	5	12	30
Complement Goal	80	12	17	18	33
Other	32	3	5	11	13
Mental	386	127	11	198	50
Subject Senser	291	88	4	175	24
Subject Phenomenon	50	14	3	14	19
Complement Phenomenon	36	22	4	5	5
Other	9	3	0	4	2
Relational	1292	320	154	436	382
Attributive	1032	283	120	329	300
Subject Carrier	874	249	115	284	226
Complement Attribute	150	31	5	42	72
Other	8	3	0	3	2
Identifying	260	37	34	107	82
Subject Token	165	24	27	75	39
Subject Value	92	12	7	30	43
Other	3	1	0	2	0
Verbal	234	66	23	112	33
Subject Sayer	186	59	18	86	23
Subject Verbiage	14	1	4	7	2
Complement Verbiage	23	5	1	11	6
Other	11	1	0	8	2
Behavioral	72	53	0	10	9
Subject Behaver	72	53	0	10	9
Existential	52	10	3	15	24
Process	29	8	3	11	7
Complement Existent	23	2	0	4	17

Table A.1 Participants and process types in the Forefield hypothesis

	Overall	FICTION	INSTR	SPEECH	TOU
Total Clauses	4901	1306	755	1580	1261
Total processes	3517	976	470	1214	857
Material	1287	335	259	367	326
Subject Actor	887	286	125	271	205
Subject Goal	214	13	107	50	44
Subject Initiator	65	16	5	15	29
Complement Goal	82	13	17	18	34
Other	39	7	5	13	14
Mental	436	143	16	218	59
Subject Senser	331	99	7	194	31
Subject Phenomenon	53	15	5	13	20
Complement Phenomenon	37	22	4	6	5
Complement Senser	15	7	0	5	3
Relational	1385	349	164	470	402
Attributive	1111	311	129	354	317
Subject Carrier	946	277	122	304	243
Complement Attribute	155	31	5	47	72
Other	10	3	2	3	2
Identifying	274	38	35	116	85
Subject Token	176	24	28	83	41
Subject Value	94	13	7	30	44
Other	4	1	0	3	0
Verbal	260	77	26	123	34
Subject Sayer	209	69	20	96	24
Subject Verbiage	14	1	4	7	2
Complement Verbiage	23	5	1	11	6
Other	14	2	1	9	2
Behavioral	75	56	0	10	9
Subject Behavior	75	56	0	10	9
Existential	74	16	5	26	27
Process	51	14	5	22	10
Complement Existent	23	2	0	4	17
Empty Subjects	70	31	1	30	8

Table A.2 Participants and process types in the first experiential element hypothesis

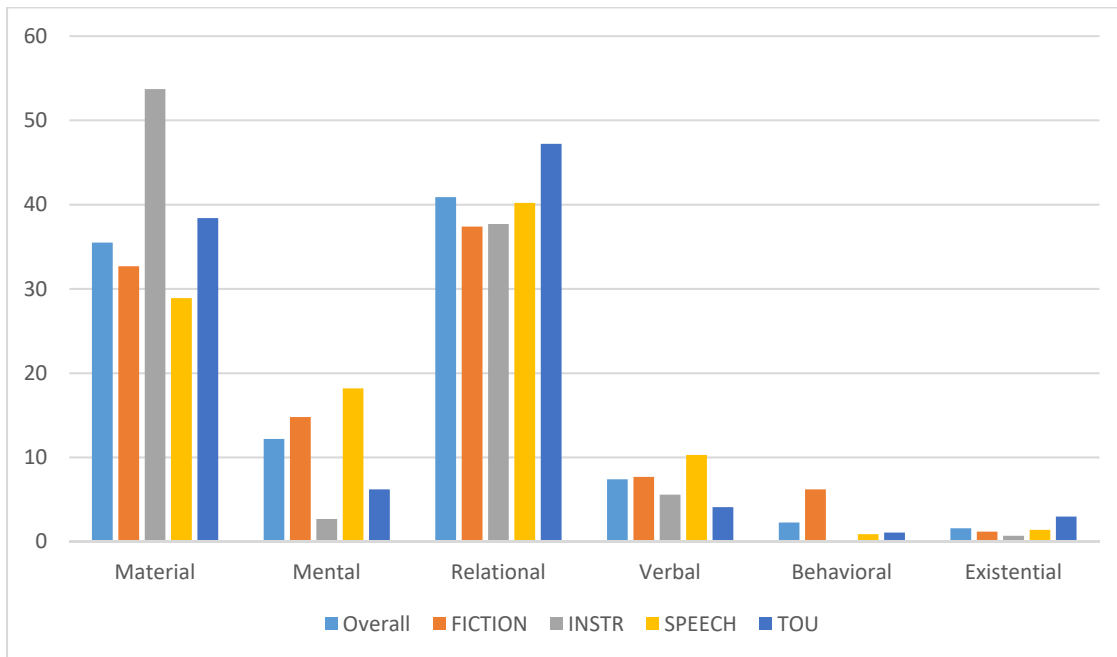


Figure A.1 Process type distribution in the Forefield hypothesis

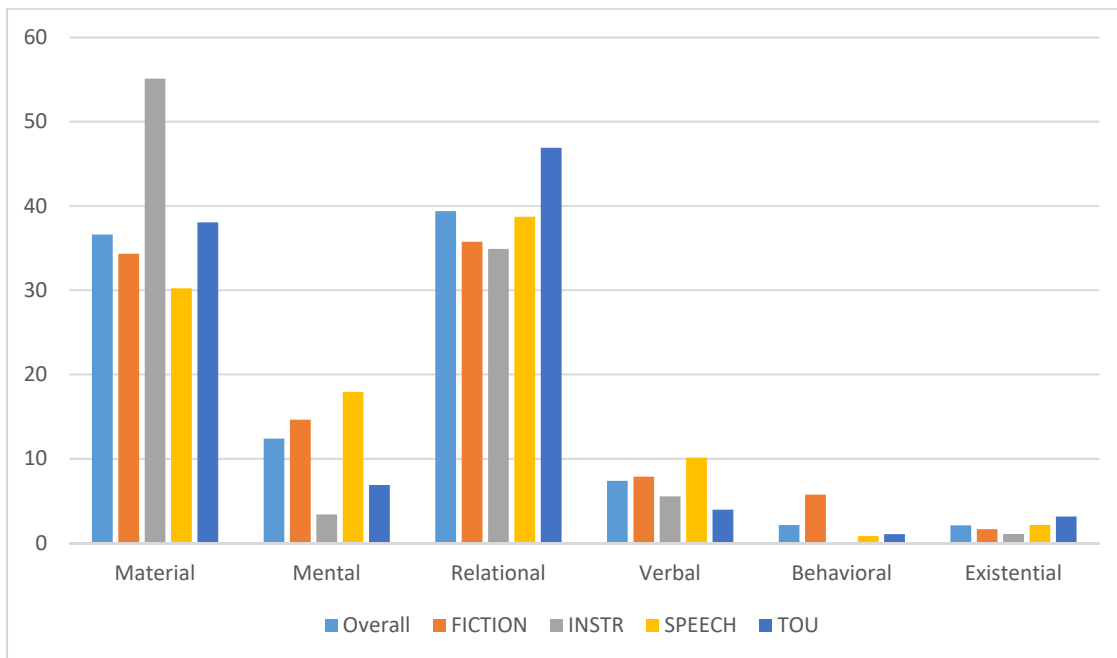


Figure A.2 Process type distribution in the first experiential element hypothesis

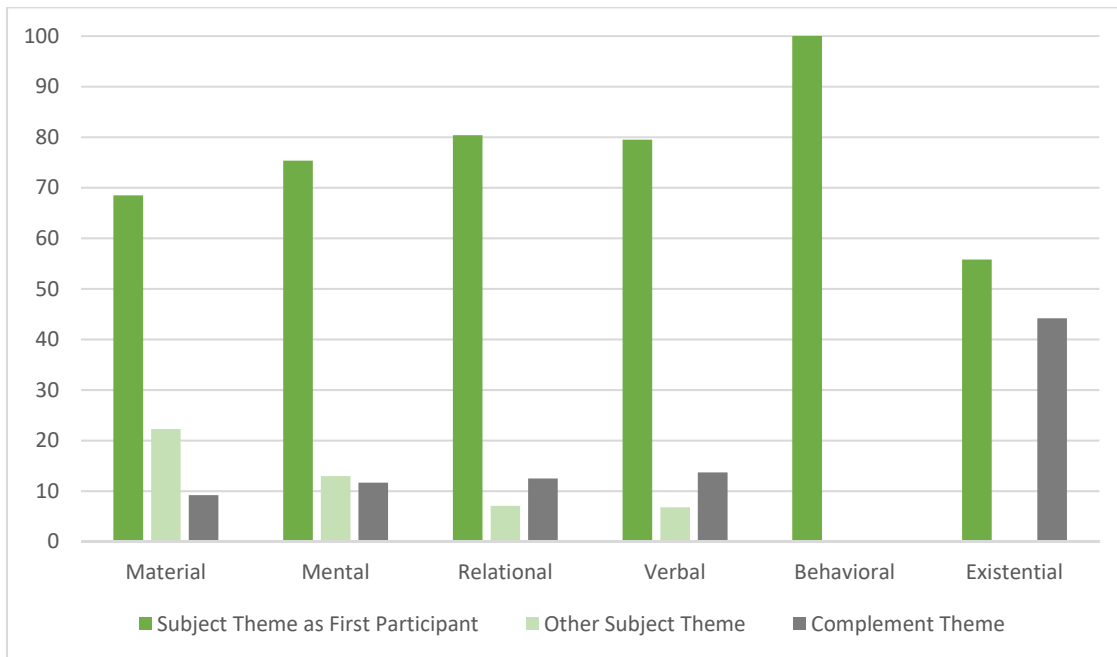


Figure A.3 Participants and MOOD elements in the Forefield hypothesis

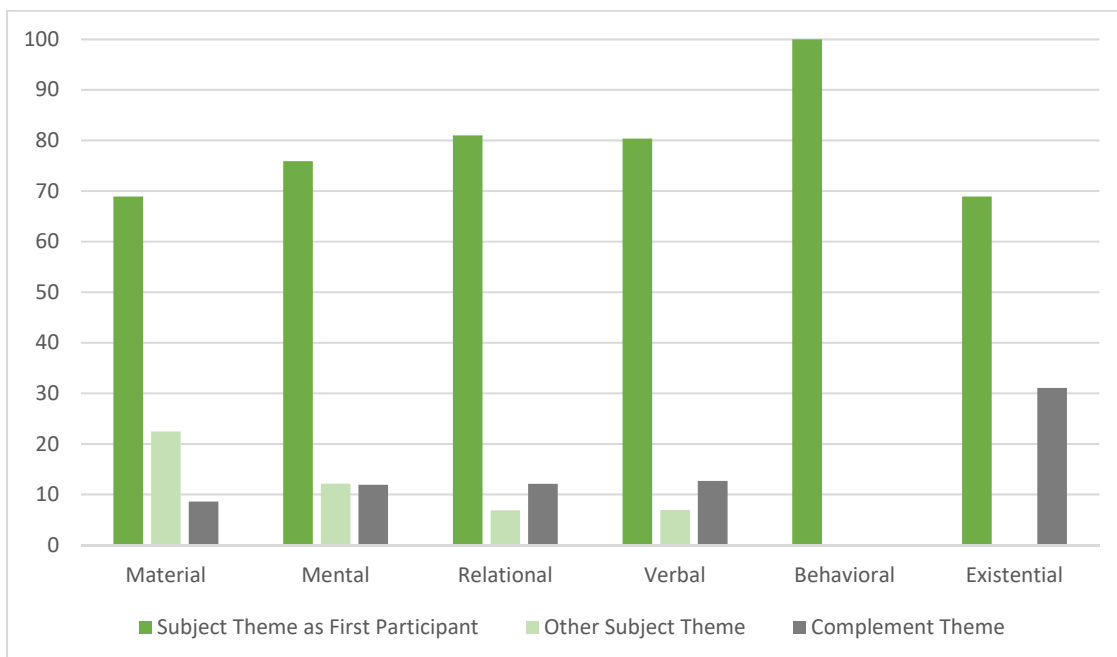


Figure A.4 Participants and MOOD elements in the first experiential element hypothesis

	Overall			FICTION			INSTR			SPEECH			TOU		
Total Clauses	4901			1306			755			1580			1261		
Total circumstances	1255	3235	38.9	282	716	39.4	279	625	44.6	313	920	34	386	974	39.6
Circumstances per clause	0.66			0.55			0.83			0.58			0.77		
Likelihood of Circ. Theme	25.6			21.1			36.2			19.4			30.8		
Additive	15	34	44.1	1	5	20.0	1	5	20.0	4	7	57.1	9	17	52.9
Behalf	38	85	44.7	3	5	60.0	4	5	80.0	10	41	24.4	21	34	61.8
Comitative	46	154	29.9	13	42	31.0	1	15	6.7	15	41	36.6	17	56	30.4
Concession	20	34	58.8	12	18	66.7	0	0	-	7	15	46.7	1	1	100
Condition	186	252	73.8	26	36	72.2	119	146	81.5	26	44	59.1	15	26	57.7
Comparison	16	55	29.1	4	30	13.3	1	1	100	7	18	38.9	4	6	66.7
Distance	2	16	12.5	0	7	0.0	0	0	-	0	1	0.0	2	8	25.0
Default	0	1	0.0	0	0	-	0	0	-	0	1	0.0	0	0	-
Degree	1	12	8.3	0	0	-	0	0	-	0	7	0.0	1	5	20.0
Duration	35	92	38.0	15	34	44.1	1	9	11.1	1	13	7.7	18	36	50.0
Frequency	17	48	35.4	9	18	50.0	1	5	20.0	1	7	14.3	6	18	33.3
Guise	19	89	21.3	6	24	25.0	1	15	6.7	0	20	0.0	12	30	40.0
Matter	24	51	47.1	3	6	50.0	1	4	25.0	15	32	46.9	5	9	55.6
Means	87	232	37.5	7	17	41.2	33	106	31.1	36	76	47.4	12	33	36.4
Place	0	16	0.0	0	1	0.0	0	4	0.0	0	3	0.0	0	8	0.0
Product	337	892	37.8	65	172	37.8	37	102	36.3	74	217	34.1	161	401	40.1
Purpose	42	134	31.3	3	16	18.8	23	44	52.3	11	46	23.9	5	28	17.9
Quality	46	311	14.8	27	103	26.2	0	55	0.0	6	90	6.7	13	63	20.6
Reason	41	109	37.6	9	27	33.3	6	16	37.5	14	44	31.8	12	22	54.5
Source	5	13	38.5	1	2	50.0	2	6	33.3	1	3	33.3	1	2	50.0
Time	274	596	46.0	82	151	54.3	41	86	47.7	77	188	41.0	74	171	43.3
Viewpoint	4	9	44.4	2	2	100	1	1	100	1	6	16.7	0	0	-

Table A.3 Circumstances and thematic potential in GO (Forefield hypothesis)

	Overall			FICTION			INSTR			SPEECH			TOU		
Total Clauses	4901			1306			755			1580			1261		
Total circumstances	1370	3235	42.3	304	716	42.5	295	625	47.2	340	920	37.0	432	974	44.4
Circumstances per clause	0.66			0.55			0.83			0.58			0.77		
Likelihood of Circ. Theme	25.7			21.6			37			19.8			30.6		
Additive	15	34	44.1	1	5	20.0	1	5	20.0	4	7	57.1	9	17	52.9
Behalf	42	85	49.4	3	5	60.0	5	5	100	13	41	31.7	21	34	61.8
Comitative	50	154	32.5	13	42	31.0	2	15	13.3	17	41	41.5	18	56	32.1
Concession	21	34	61.8	12	18	66.7	0	0	-	8	15	53.3	1	1	100
Condition	186	252	73.8	26	36	72.2	119	146	81.5	26	44	59.1	15	26	57.7
Comparison	16	55	29.1	4	30	13.3	1	1	100	7	18	38.9	4	6	66.7
Distance	2	16	12.5	0	7	0.0	0	0	-	0	1	0.0	2	8	25.0
Default	0	1	0.0	0	0	-	0	0	-	0	1	0.0	0	0	-
Degree	1	12	8.3	0	0	-	0	0	-	0	7	0.0	1	5	20.0
Duration	37	92	40.2	15	34	44.1	3	9	33.3	1	13	7.7	18	36	50.0
Frequency	22	48	45.8	10	18	55.6	3	5	60.0	2	7	28.6	7	18	38.9
Guise	20	89	22.5	7	24	29.2	1	15	6.7	0	20	0.0	12	30	40.0
Matter	27	51	52.9	4	6	66.7	1	4	25.0	16	32	50.0	6	9	66.7
Means	90	232	38.8	7	17	41.2	33	106	31.1	38	76	50.0	13	33	39.4
Place	391	892	43.8	70	172	40.7	48	102	47.1	86	217	39.6	187	401	46.6
Product	0	16	0.0	0	1	0.0	0	4	0.0	0	3	0.0	0	8	0.0
Purpose	44	134	32.8	4	16	25.0	24	44	54.5	11	46	23.9	5	28	17.9
Quality	54	311	17.4	29	103	28.2	1	55	1.8	8	90	8.9	16	63	25.4
Reason	42	109	38.5	9	27	33.3	7	16	43.8	14	44	31.8	12	22	54.5
Source	6	13	46.2	1	2	50.0	2	6	33.3	1	3	33.3	2	2	100
Time	299	596	50.2	87	151	57.6	43	86	50.0	86	188	45.7	83	171	48.5
Viewpoint	5	9	55.6	2	2	100	1	1	100	2	6	33.3	0	0	-

Table A.4 Circumstances and thematic potential in GO (Subject hypothesis)

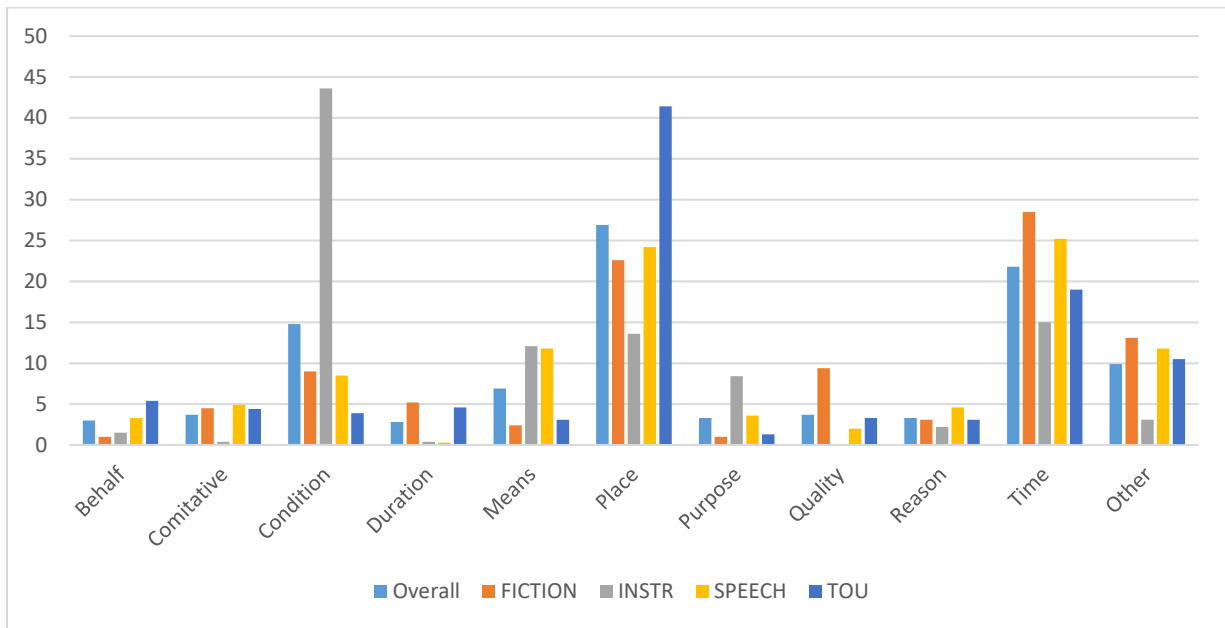


Figure A.5 Circumstance Theme distribution in GO (Forefield hypothesis)

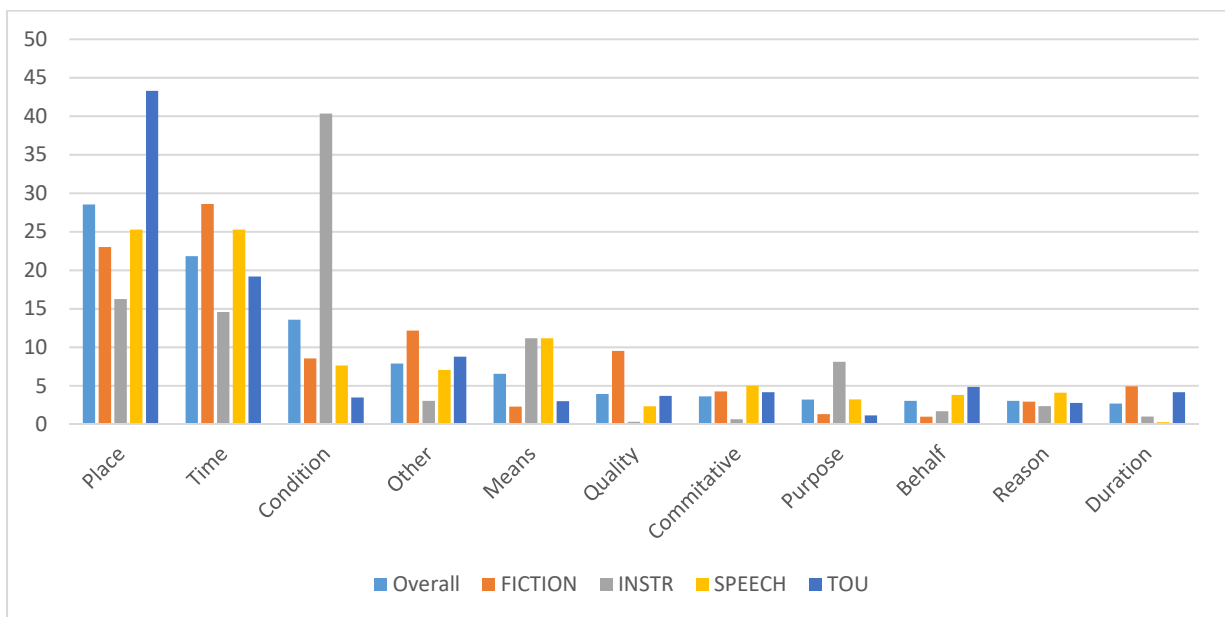


Figure A.6 Circumstance Theme distribution in GO (Subject hypothesis)

	Overall			FICTION			INSTR			SPEECH			TOU		
Total clauses	4901			1306			755			1580			1261		
Relational	1118			393			211			528			485		
Not analyzed	2403			148			259			259			225		
Analyzed Subjects	1380			765			285			793			551		
Human	795			632			106			462			231		
Animal	10	10	100	10	10	100	0	0	-	0	0	-	0	0	-
Organization	58	88	65.9	1	1	100	0	0	-	49	70	70.0	8	17	47.1
Machine	17	33	51.5	0	0	-	16	32	50.0	1	1	100	0	0	-
Vehicle	2	4	50.0	1	2	50.0	0	0	-	0	0	-	1	2	50.0
Concrete In- animate	42	135	31.1	9	31	29.0	5	51	9.8	4	9	44.4	24	44	54.5
Nonconcrete Inanimate	72	223	32.3	8	27	29.6	11	17	64.7	28	120	23.3	25	59	42.4
Place	22	59	37.3	0	0	-	0	0	-	0	1	0.0	22	58	37.9
Time	2	3	66.7	1	1	100	0	0	-	0	0	-	1	2	50.0
Process	5	30	16.7	3	6	50.0	0	8	0.0	1	12	8.3	1	4	25.0

Table A.5 Subject Theme Sentience and Animacy in GO (Forefield hypothesis)

	Overall			FICTION			INSTR			SPEECH			TOU		
Total clauses	4901			1306			755			1580			1261		
Relational	1202			310			157			407			328		
Not analyzed	2132			470			441			587			635		
Analyzed Subjects	1567			526			157			586			298		
Human	917			434			39			337			107		
Animal	11	11	100	11	11	100	0	0	-	0	0	-	0	0	-
Organization	67	103	65.0	2	2	100	0	0	-	57	83	68.7	8	18	44.4
Machine	20	38	52.6	0	0	-	19	37	51.4	1	1	100	0	0	-
Vehicle	2	5	40.0	1	3	33.3	0	0	-	0	0	-	1	2	50.0
Concrete In- animate	45	145	31.0	10	38	26.3	5	52	9.6	5	10	50.0	25	45	55.6
Nonconcrete Inanimate	73	249	29.3	9	31	29.0	11	19	57.9	30	140	21.4	23	59	39.0
Place	24	62	38.7	0	0	-	0	0	-	0	1	0.0	24	61	39.3
Time	2	5	40.0	1	1	100	0	1	0	0	1	0	1	2	50.0
Process	5	32	15.6	3	6	50.0	0	9	0	1	13	7.7	1	4	25.0

Table A.6 Subject Sentience and Animacy in GO (first experiential element hypothesis)

		Overall		FICTION		INSTR		SPEECH		TOU	
Subjects in total		4901		1292		752		1554		1258	
	Identifiable	2106	81.3	650	92.2	280	79.5	799	86.8	377	61.4
	Non-Identifiable	485	18.7	55	7.8	72	20.5	121	13.2	237	38.6
Other		261		96		27		93		45	
No Subject		2049		505		376		567		602	
Pre-verbal Subjects		2852		801		379		1013		659	
	Identifiable	2106	81.3	650	92.2	280	79.5	799	86.8	377	61.4
	Non-Identifiable	485	18.7	55	7.8	72	20.5	121	13.2	237	38.6
Post-verbal Subjects		0		0		0		0		0	
	Identifiable	0	-	0	-	0	-	0	-	0	-
	Non-Identifiable	0	-	0	-	0	-	0	-	0	-
Subjects in immediate post-verbal position		0		0		0		0		0	
	Identifiable	0	-	0	-	0	-	0	-	0	-
	Non-Identifiable	0	-	0	-	0	-	0	-	0	-

Table A.7 Subject Theme Identifiability in GO (Forefield hypothesis)

		Overall		FICTION		INSTR		SPEECH		TOU	
Subjects in total		4901		1292		752		1554		1258	
	Identifiable	2358	81.5	743	92.3	326	79.7	887	86.4	402	61.6
	Non-Identifiable	536	18.5	62	7.7	83	20.3	140	13.6	251	38.4
Other		340		124		33		120		63	
No Subject		1667		377		313		433		545	
Pre-verbal Subjects		2852		2875		806		382		1023	
	Identifiable	2105	81.2	648	92.2	280	79.5	800	86.8	377	61.3
	Non-Identifiable	487	18.8	55	7.8	72	20.5	122	13.2	238	38.7
Post-verbal Subjects		359		123		60		124		52	
	Identifiable	253	83.8	95	93.1	46	80.7	87	82.9	25	65.8
	Non-Identifiable	49	16.2	7	6.9	11	19.3	18	17.1	13	34.2
Subjects in immediate post-verbal position		327		115		53		115		44	
	Identifiable	243	85.3	93	93.9	42	82.4	86	83.5	22	68.8
	Non-Identifiable	42	14.7	6	6.1	9	17.6	17	16.5	10	31.3

Table A.8 Subject Theme Identifiability in GO (first experiential element hypothesis)

Chapter 8: Theme in English Originals

	Overall	FICTION	INSTR	SPEECH	TOU
Total Clauses	4496	1241	845	1430	980
Total processes	3168	830	612	985	741
Material	1203	253	369	379	202
Subject Actor	959	234	261	315	149
Subject Goal	139	13	53	33	40
Subject Initiator	92	3	54	25	10
Complement Goal	3	2	0	0	1
Other	10	1	1	6	2
Mental	348	144	29	131	44
Subject Senser	312	139	23	124	26
Subject Phenomenon	35	4	6	7	18
Complement Phenomenon	1	1	0	0	0
Relational	1270	295	189	359	427
Attributive	1008	232	149	277	350
Subject Carrier	975	222	149	273	331
Complement Attribute	33	10	0	4	19
Identifying	262	63	40	82	77
Subject Token	194	32	35	67	60
Subject Value	63	28	5	14	16
Other	5	3	0	1	1
Verbal	195	70	15	100	10
Subject Sayer	176	64	13	92	7
Subject Verbiage	8	2	2	1	3
Complement Verbiage	1	1	0	0	0
Other	10	3	0	7	0
Behavioral	55	45	0	3	7
Subject Behavior	55	45	0	3	7
Existential	97	23	10	13	51
Process	97	23	10	13	51
Empty Subjects	16	15	0	1	0

Table A.9 Participants and process types in the first element hypothesis

	Overall	FICTION	INSTR	SPEECH	TOU
Total Clauses	4496	1241	845	1430	980
Total processes	3589	1034	641	1126	788
Material	1351	317	385	437	212
Subject Actor	1087	291	272	366	158
Subject Goal	149	17	55	36	41
Subject Initiator	99	5	56	28	10
Complement Goal	3	2	0	0	1
Other	13	2	2	7	2
Mental	409	188	33	144	44
Subject Senser	370	180	27	137	26
Subject Phenomenon	38	7	6	7	18
Complement Phenomenon	1	1	0	0	0
Relational	1414	354	196	410	454
Attributive	1117	278	155	315	369
Subject Carrier	1080	266	155	311	348
Complement Attribute	37	12	0	4	21
Identifying	297	76	41	95	85
Subject Token	217	40	36	76	65
Subject Value	75	33	5	18	19
Other	5	3	0	1	1
Verbal	231	88	15	117	11
Subject Sayer	212	82	13	109	8
Subject Verbiage	8	2	2	1	3
Complement Verbiage	1	1	0	0	0
Other	10	3	0	7	0
Behavioral	68	55	2	3	8
Subject Behaver	68	55	2	3	8
Existential	116	32	10	15	59
Process	116	32	10	15	59
Empty Subjects	25	23	0	2	0

Table A.10 Participants and process types in the first experiential element hypothesis

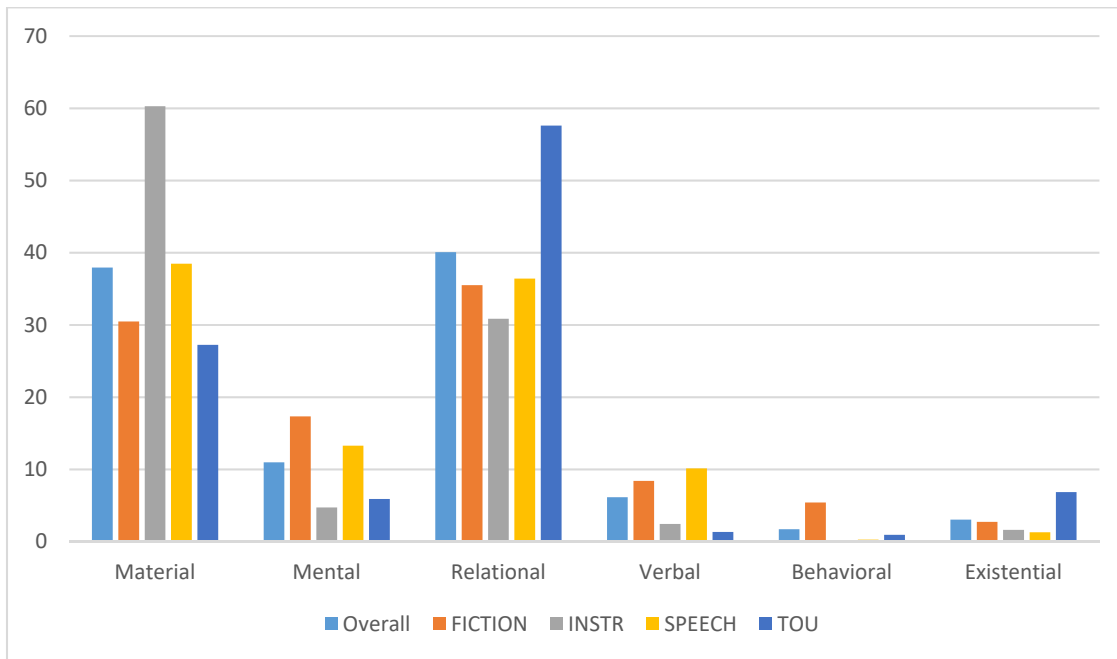


Figure A.7 Process type distribution in the first element hypothesis

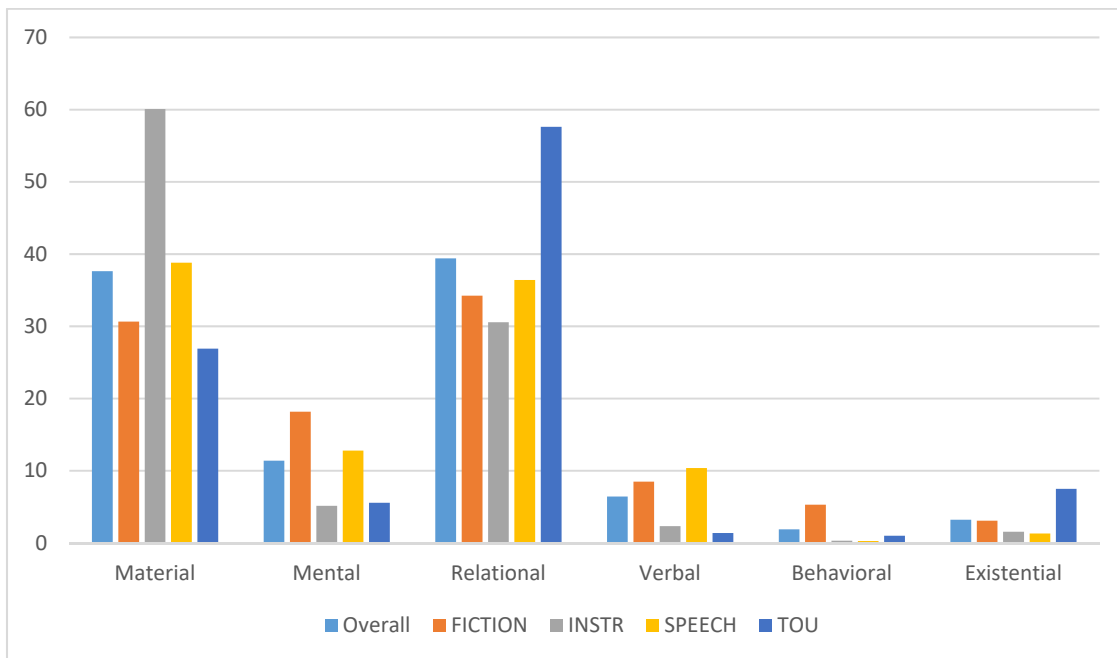


Figure A.8 Process type distribution in the first experiential element hypothesis

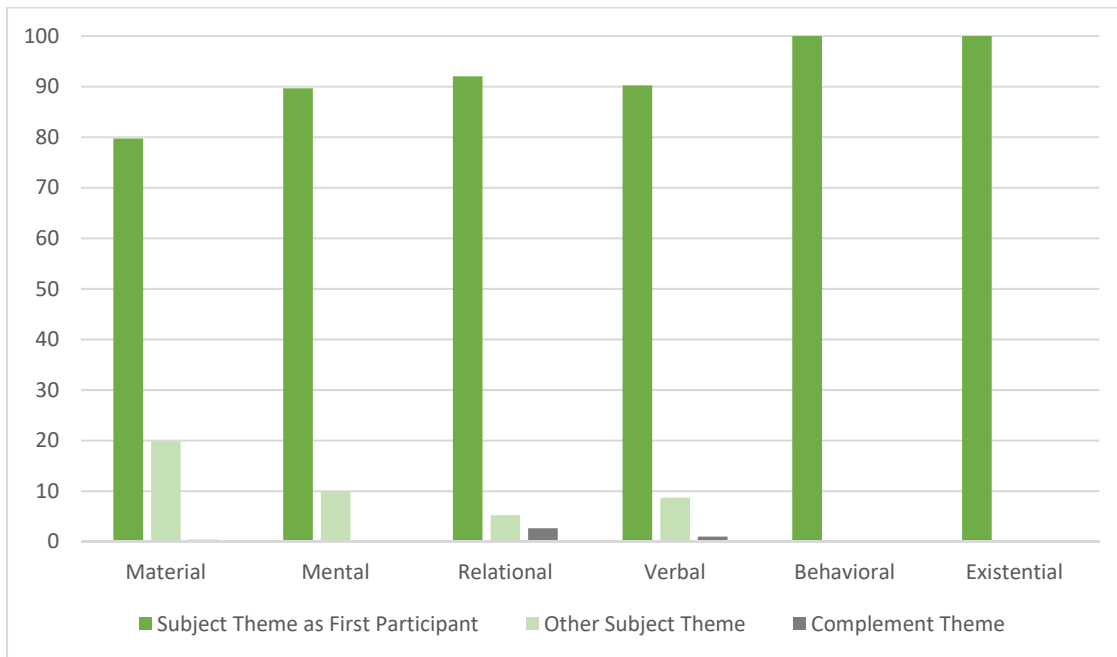


Figure A.9 Participants and MOOD elements in the first element hypothesis

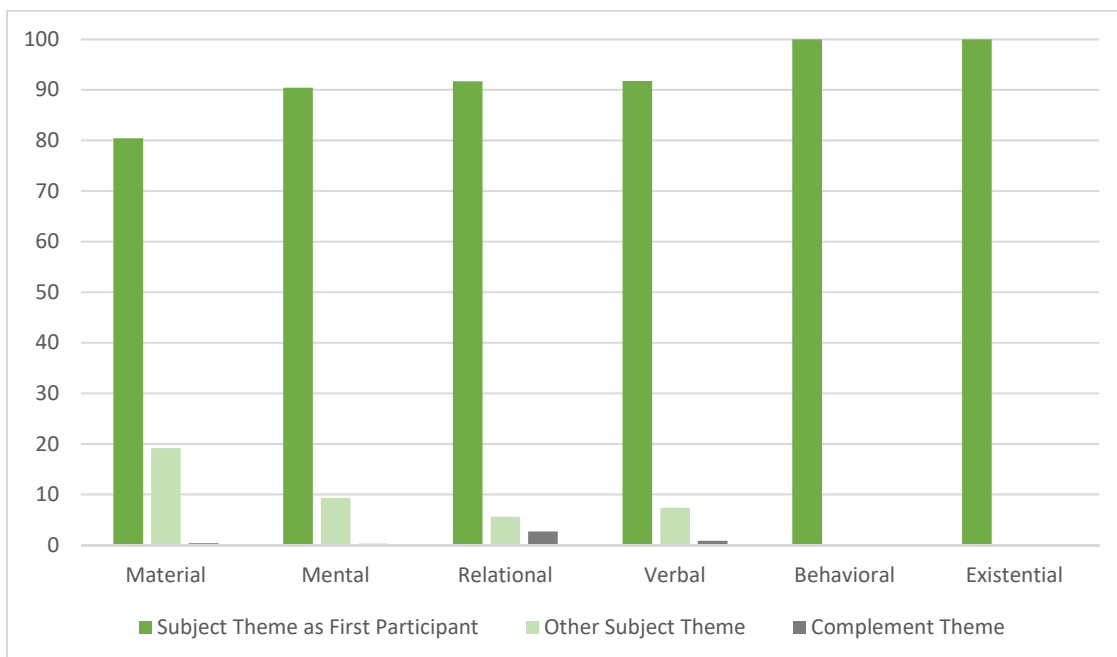


Figure A.10 Participants and MOOD elements in the first experiential element hypothesis

	Overall			FICTION			INSTR			SPEECH			TOU		
Total Clauses	4496			1241			845			1430			980		
Total circumstances	731	2443	29.9	127	540	23.5	187	562	33.3	244	800	30.5	173	541	32.0
Circumstances per clause	0.54			0.44			0.67			0.56			0.55		
Likelihood of Circ. Theme	16.3			10.2			22.1			17.1			17.7		
Additive	13	18	72.2	0	0	-	2	3	66.7	8	8	100	3	7	42.9
Behalf	16	50	32.0	4	10	40.0	7	14	50.0	0	2	0	5	24	20.8
Comitative	14	131	10.7	2	29	6.9	1	16	6.3	4	40	10.0	7	46	15.2
Concession	21	39	53.8	1	11	9.1	0	2	0	14	18	77.8	6	8	75.0
Condition	173	282	61.3	16	25	64.0	121	187	64.7	21	50	42.0	15	20	75.0
Comparison	13	51	25.5	4	29	13.8	2	4	50.0	5	15	33.3	2	3	66.7
Distance	0	6	0	0	2	0	0	2	0	0	1	0	0	1	0
Default	0	1	0	0	0	-	0	1	0	0	0	-	0	0	-
Degree	0	5	0	0	1	0	0	0	-	0	3	0	0	1	0
Duration	27	76	35.5	8	31	25.8	0	5	0	11	19	57.9	8	21	38.1
Frequency	6	44	13.6	3	9	33.3	1	11	9.1	0	13	0	2	11	18.2
Guise	28	61	45.9	3	5	60.0	0	10	0.0	11	21	52.4	14	25	56.0
Matter	7	21	33.3	1	5	20.0	1	2	50.0	5	12	41.7	0	2	0.0
Means	39	128	30.5	3	11	27.3	9	56	16.1	24	52	46.2	3	9	33.3
Place	124	520	23.8	19	105	18.1	17	107	15.9	31	115	27.0	57	193	29.5
Product	0	9	0	0	1	0	0	3	0	0	2	0	0	3	0
Purpose	21	253	8.3	0	39	0	11	68	16.2	9	115	7.8	1	31	3.2
Quality	12	184	6.5	7	58	12.1	0	29	0	4	82	4.9	1	15	6.7
Reason	21	104	20.2	4	40	10.0	4	11	36.4	4	27	14.8	9	26	34.6
Source	7	11	63.6	0	1	0	0	0	-	7	10	70.0	0	0	-
Time	187	444	42.1	52	127	40.9	11	31	35.5	84	191	44.0	40	95	42.1
Viewpoint	2	5	40.0	0	1	0	0	0	-	2	4	50.0	0	0	-

Table A.11 circumstances and thematic potential in the first element hypothesis

	Overall			FICTION			INSTR			SPEECH			TOU		
Total Clauses	4496			1241			845			1430			980		
Total circumstances	865	2443	35.4	179	540	33.1	201	562	35.8	296	800	37.0	189	541	34.9
Circumstances per clause	0.54			0.44			0.67			0.56			0.55		
Likelihood of Circ. Theme	19.2			14.4			23.8			20.7			19.3		
Additive	14	18	77.8	0	0	-	2	3	66.7	8	8	100.0	4	7	57.1
Behalf	19	50	38.0	4	10	40.0	8	14	57.1	1	2	50.0	6	24	25.0
Comitative	15	131	11.5	2	29	6.9	1	16	6.3	4	40	10.0	8	46	17.4
Concession	24	39	61.5	3	11	27.3	0	2	0.0	15	18	83.3	6	8	75.0
Condition	198	282	70.2	20	25	80.0	133	187	71.1	29	50	58.0	16	20	80.0
Comparison	20	51	39.2	6	29	20.7	2	4	50.0	10	15	66.7	2	3	66.7
Distance	0	6	0.0	0	2	0.0	0	2	0.0	0	1	0.0	0	1	0.0
Default	0	1	0.0	0	0	-	0	1	0.0	0	0	-	0	0	-
Degree	0	5	0.0	0	1	0.0	0	0	-	0	3	0.0	0	1	0.0
Duration	29	76	38.2	10	31	32.3	0	5	0.0	11	19	57.9	8	21	38.1
Frequency	6	44	13.6	3	9	33.3	1	11	9.1	0	13	0.0	2	11	18.2
Guise	30	61	49.2	4	5	80.0	0	10	0.0	12	21	57.1	14	25	56.0
Matter	10	21	47.6	1	5	20.0	1	2	50.0	7	12	58.3	1	2	50.0
Means	42	128	32.8	3	11	27.3	9	56	16.1	26	52	50.0	4	9	44.4
Place	142	520	27.3	27	105	25.7	18	107	16.8	37	115	32.2	60	193	31.1
Product	0	9	0.0	0	1	0.0	0	3	0.0	0	2	0.0	0	3	0.0
Purpose	22	253	8.7	0	39	0.0	11	68	16.2	10	115	8.7	1	31	3.2
Quality	15	184	8.2	10	58	17.2	0	29	0.0	4	82	4.9	1	15	6.7
Reason	31	104	29.8	7	40	17.5	4	11	36.4	9	27	33.3	11	26	42.3
Source	8	11	72.7	0	1	0.0	0	0	-	8	10	80.0	0	0	-
Time	237	444	53.4	79	127	62.2	11	31	35.5	102	191	53.4	45	95	47.4
Viewpoint	3	5	60.0	0	1	0.0	0	0	-	3	4	75.0	0	0	-

Table A.12 circumstances and thematic potential in the Subject hypothesis

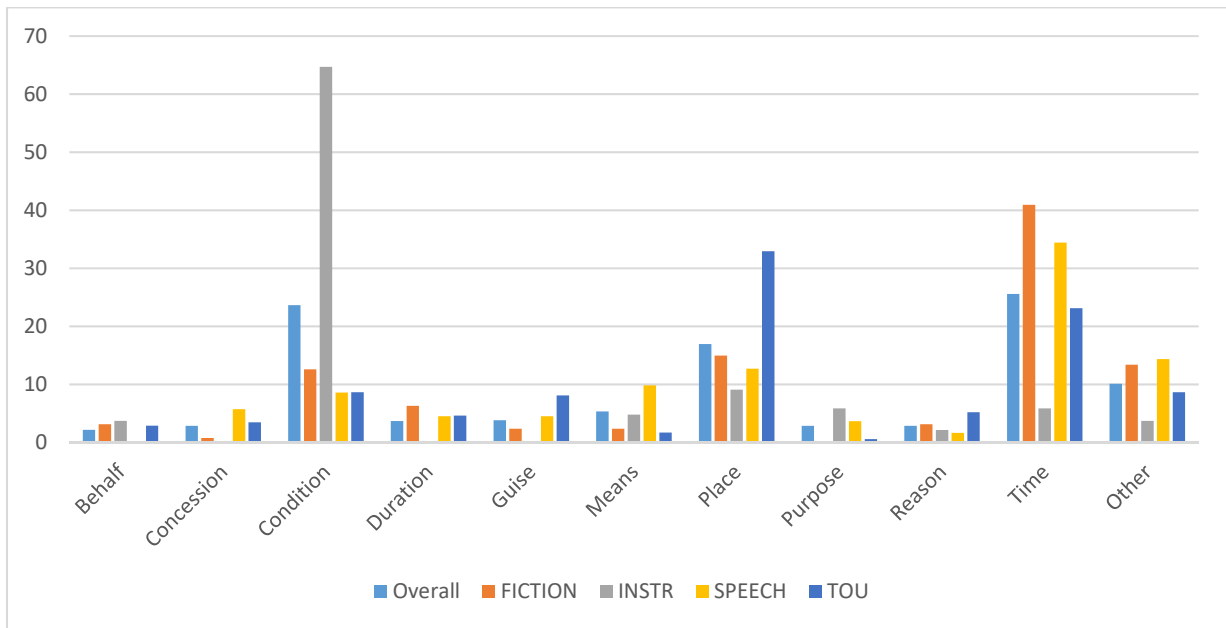


Figure A.11 circumstance Theme distribution in the first element hypothesis

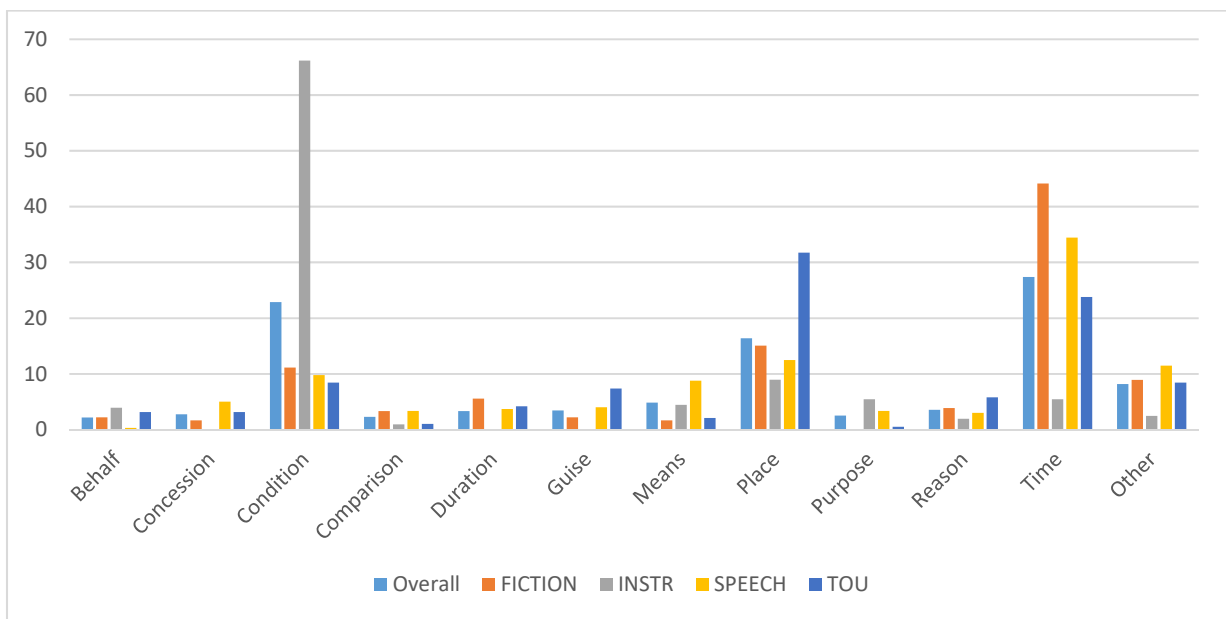


Figure A.12 circumstance Theme distribution in the Subject hypothesis

	Overall			FICTION			INSTR			SPEECH			TOU		
Total clauses	4496			1241			845			1430			980		
Relational	1218			272			189			351			406		
Not analyzed	1684			484			305			520			375		
Analyzed Subjects	1594			485			351			559			199		
Human	928			423			121			322			62		
Animal	7	7	100	7	7	100	0	0	-	0	0	-	0	0	-
Organization	87	116	75.0	6	6	100	0	0	-	76	96	79.2	5	14	35.7
Machine	16	38	42.1	0	0	-	16	38	42.1	0	0	-	0	0	-
Vehicle	2	6	33.3	2	3	66.7	0	1	0	0	0	-	0	2	0
Concrete In-animate	28	112	25.0	9	19	47.4	11	60	18.3	3	8	37.5	5	25	20.0
Nonconcrete Inanimate	86	262	32.8	8	21	38.1	35	97	36.1	33	114	28.9	10	30	33.3
Place	20	65	30.8	1	1	100	0	0	-	0	2	0	19	62	30.6
Time	1	8	12.5	0	0	-	0	1	0.0	0	5	0	1	2	50.0
Process	14	52	26.9	1	5	20.0	9	33	27.3	2	12	16.7	2	2	100

Table A.13 Subject Sentience and Animacy in the first element hypothesis

	Overall			FICTION			INSTR			SPEECH			TOU		
Total clauses	4496			1241			845			1430			980		
Relational	1356			327			196			402			431		
Not analyzed	1304			301			279			385			339		
Analyzed Subjects	1836			613			370			643			210		
Human	1116			537			134			378			67		
Animal	9	9	100	9	9	100	0	0	-	0	0	-	0	0	-
Organization	92	127	72.4	6	6	100	0	0	-	81	107	75.7	5	14	35.7
Machine	21	43	48.8	0	0	-	21	43	48.8	0	0	-	0	0	-
Vehicle	2	6	33.3	2	3	66.7	0	1	0	0	0	-	0	2	0
Concrete In-animate	31	121	25.6	12	24	50.0	11	60	18.3	3	10	30.0	5	27	18.5
Nonconcrete Inanimate	95	282	33.7	12	28	42.9	35	98	35.7	38	126	30.2	10	30	33.3
Place	20	67	29.9	1	1	100	0	0	-	0	2	0	19	64	29.7
Time	1	10	10.0	0	0	-	0	1	0	0	6	0	1	3	33.3
Process	14	55	25.5	1	5	20.0	9	33	27.3	2	14	14.3	2	3	66.7

Table A.14 Subject Sentience and Animacy in the first experiential element hypothesis

Chapter 9: Contrastive Analysis of Theme in English and German

	FICTION				INSTR			
	GO		EO		GO		EO	
Total Clauses	1306		1241		755		845	
Total Themes	1421		1241		758		845	
Single Themes	1195	91.5	1241	100	752	99.6	845	100
Multiple Themes	111	8.5	0	0	3	0.4	0	0
Average # of Themes	1.09		1.00		1.00		1.00	
Experiential Themes	1169		985		682		803	
Subject Themes	796	68.1	829	84.2	379	55.6	612	76.2
Circumstance Themes	288	24.6	127	12.9	273	40.0	187	23.3
Complement Themes	81	6.9	16	1.6	30	4.4	0	0.0
Predicator Themes	4	0.3	1	0.1	0	0.0	0	0.0
Textual Themes	165	11.6	215	17.3	51	6.7	30	3.6
Interpersonal Themes	74	5.2	41	3.3	22	2.9	12	1.4
Cleft	13	0.9	12	1.2	3	0.4	4	0.5

Table A.15 Differences in general Theme distribution between GO (Forefield hypothesis) and EO (first element hypothesis) in FICTION and INSTR

	SPEECH				TOU			
	GO		EO		GO		EO	
Total Clauses	1580		1430		1261		980	
Total Themes	1772		1430		1327		980	
Single Themes	1401	88.7	1430	100	1196	94.8	980	100
Multiple Themes	179	11.3	0	0	65	5.2	0	0
Average # of Themes	1.12		1.00		1.05		1.00	
Experiential Themes	1415		1248		1210		919	
Subject Themes	1002	70.8	982	78.7	665	55.0	719	78.2
Circumstance Themes	306	21.6	244	19.6	389	32.1	173	18.8
Complement Themes	106	7.5	4	0.3	147	12.1	22	2.4
Predicator Themes	0	0	0	0	9	0.7	3	0.3
Textual Themes	218	12.3	134	9.4	94	7.1	32	3.3
Interpersonal Themes	117	6.6	48	3.4	22	1.7	29	3.0
Cleft	22	1.2	18	1.4	1	0.1	2	0.2

Table A.16 Differences in general Theme distribution between GO (Forefield hypothesis) and EO (first element hypothesis) in SPEECH and TOU

	FICTION				INSTR			
	GO		EO		GO		EO	
Total Clauses	1306		1241		755		845	
Total Themes	1421		1543		758		887	
Single Themes	1195	91.5	985	79.4	752	99.6	803	95.0
Multiple Themes	111	8.5	256	20.6	3	0.4	42	5.0
Average # of Themes	1.09		1.24		1.00		1.05	
Experiential Themes	1169		1241		682		845	
Subject Themes	796	68.1	1039	83.7	379	55.6	641	75.9
Circumstance Themes	288	24.6	166	13.4	273	40.0	200	23.7
Complement Themes	81	6.9	18	1.5	30	4.4	0	0.0
Predicator Themes	4	0.3	1	0.1	0	0.0	0	0.0
Textual Themes	165	11.6	233	15.1	51	6.7	30	3.4
Interpersonal Themes	74	5.2	69	4.5	22	2.9	12	1.4
Cleft	13	0.9	17	1.4	3	0.4	4	0.5

Table A.17 Differences in general Theme distribution between GO (Forefield hypothesis) and EO (first experiential element hypothesis) in FICTION and INSTR

	SPEECH				TOU			
	GO		EO		GO		EO	
Total Clauses	1580		1430		1261		980	
Total Themes	1772		1625		1327		1042	
Single Themes	1401	88.7	1248	87.3	1196	94.8	919	93.8
Multiple Themes	179	11.3	182	12.7	65	5.2	61	6.2
Average # of Themes	1.12		1.14		1.05		1.06	
Experiential Themes	1415		1430		1210		980	
Subject Themes	1002	70.8	1124	78.6	665	55.0	764	78.0
Circumstance Themes	306	21.6	277	19.4	389	32.1	186	19.0
Complement Themes	106	7.5	4	0.3	147	12.1	24	2.4
Predicator Themes	0	0	0	0	9	0.7	3	0.3
Textual Themes	218	12.3	136	8.4	94	7.1	32	3.1
Interpersonal Themes	117	6.6	59	3.6	22	1.7	30	2.9
Cleft	22	1.2	25	1.7	1	0.1	3	0.3

Table A.18 Differences in general Theme distribution between GO (Forefield hypothesis) and EO (first experiential element hypothesis) in SPEECH and TOU

	FICTION				INSTR			
	GO		EO		GO		EO	
Total Clauses	1306		1241		755		845	
Total Themes	1680		1543		904		887	
Single Themes	1093	83.7	985	79.4	683	90.5	803	95.0
Multiple Themes	213	16.3	256	20.6	72	9.5	42	5.0
Average # of Themes	1.29		1.24		1.20		1.05	
Avg. # of Themes w/o Finite	1.18		1.24		1.10		1.05	
Experiential Themes	1294		1241		752		845	
Subject Themes	918	70.9	1039	83.7	439	58.4	641	75.9
Circumstance Themes	282	21.8	166	13.4	279	37.1	200	23.7
Complement Themes	89	6.9	18	1.5	32	4.3	0	0.0
Predicator Themes	5	0.4	1	0.1	2	0.3	0	0.0
Textual Themes	163	9.7	233	15.1	55	6.1	30	3.4
Interpersonal Themes	211	12.6	69	4.5	94	10.4	12	1.4
Finite Themes	139	65.9	0	0	70	74.5	0	0
Modal Adjuncts	72	34.1	69	100	24	25.5	12	100
Cleft	12	0.7	17	1.4	3	0.3	4	0.5

Table A.19 Differences in general Theme distribution between GO (first experiential element hypothesis) and EO (first experiential element hypothesis) in FICTION and INSTR

	SPEECH				TOU			
	GO		EO		GO		EO	
Total Clauses	1580		1430		1261		980	
Total Themes	2063		1625		1438		1042	
Single Themes	1271	80.4	1248	87.3	1157	91.8	919	93.8
Multiple Themes	309	19.6	182	12.7	104	8.2	61	6.2
Average # of Themes	1.31		1.14		1.14		1.06	
Avg. # of Themes w/o Finite	1.21		1.14		1.09		1.06	
Experiential Themes	1558		1430		1260		980	
Subject Themes	1128	72.4	1124	78.6	715	56.7	764	78.0
Circumstance Themes	313	20.1	277	19.4	386	30.6	186	19.0
Complement Themes	116	7.4	4	0.3	150	11.9	24	2.4
Predicator Themes	0	0	0	0	9	0.7	3	0.3
Textual Themes	217	10.5	136	8.4	92	6.4	32	3.1
Interpersonal Themes	266	12.9	59	3.6	85	5.9	30	2.9
Finite Themes	149	56.0	0	0	58	68.2	0	0
Modal Adjuncts	117	44.0	59	100	27	31.8	30	100
Cleft	22	1.1	25	1.7	1	0.1	3	0.3

Table A.20 Differences in general Theme distribution between GO (first experiential element hypothesis) and EO (first experiential element hypothesis) in SPEECH and TOU

	FICTION				INSTR			
	GO		EO		GO		EO	
Total Clauses	1306		1241		755		845	
Total Themes	2463		1752		1546		1088	
Single Themes	764	58.5	842	67.8	381	50.5	616	72.9
Multiple Themes	542	41.5	399	32.2	374	49.5	229	27.1
Average # of Themes	1.89		1.41		2.05		1.29	
Avg. # of Themes w/o Finite	1.51		1.41		1.56		1.29	
Experiential Themes	1717		1439		1086		1046	
Subject Themes	1292	75.2	1222	84.9	752	69.2	839	80.2
Circumstance Themes	304	17.7	179	12.4	295	27.2	201	19.2
Complement Themes	100	5.8	18	1.3	33	3.0	0	0
Predicator Themes	7	0.4	2	0.1	3	0.3	0	0
Textual Themes	168	6.8	233	13.3	58	3.8	30	2.8
Interpersonal Themes	578	23.5	80	4.6	402	26.0	12	1.1
Finite Themes	494	85.5	0	0	367	91.3	0	0
Modal Adjuncts	84	14.5	80	100	35	8.7	12	100
Cleft	14	0.8	18	1.3	3	0.3	6	0.6

Table A.21 Differences in general Theme distribution between GO (Subject hypothesis) and EO (Subject hypothesis) in FICTION and INSTR

	SPEECH				TOU			
	GO		EO		GO		EO	
Total Clauses	1580		1430		1261		980	
Total Themes	2997		1936		2623		1291	
Single Themes	909	57.5	1000	69.9	639	50.7	720	73.5
Multiple Themes	671	42.5	430	30.1	622	49.3	259	26.5
Average # of Themes	1.90		1.35		2.08		1.32	
Avg. # of Themes w/o Finite	1.55		1.35		1.61		1.32	
Experiential Themes	2066		1731		1862		1197	
Subject Themes	1554	75.2	1400	80.9	1258	67.6	976	81.5
Circumstance Themes	340	16.5	296	17.1	432	23.2	189	15.8
Complement Themes	133	6.4	4	0.2	159	8.5	24	2.0
Predicator Themes	13	0.6	1	0.1	10	0.5	5	0.4
Textual Themes	239	8.0	137	7.1	101	3.9	34	2.6
Interpersonal Themes	692	23.1	68	3.5	660	25.2	60	4.6
Finite Themes	549	79.3	0	0	592	89.7	0	0
Modal Adjuncts	143	20.7	68	100	68	10.3	60	100
Cleft	26	1.3	30	1.7	3	0.2	3	0.3

Table A.22 Differences in general Theme distribution between GO (Subject hypothesis) and EO (Subject hypothesis) in SPEECH and TOU

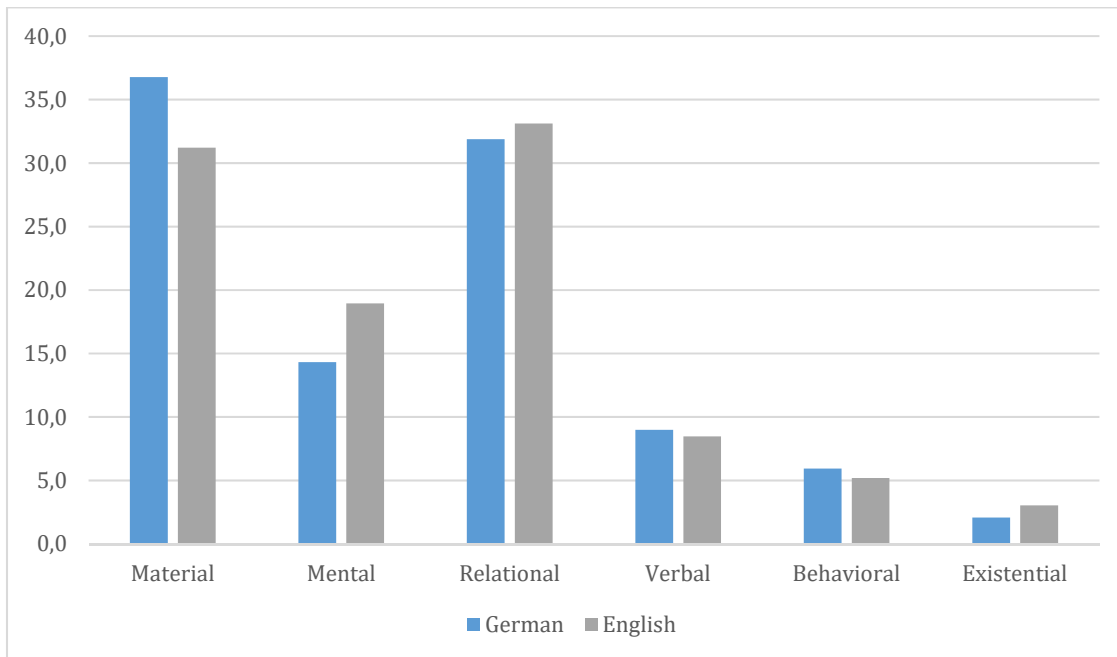


Figure A.13 Process type distribution between GO and EO in FICTION

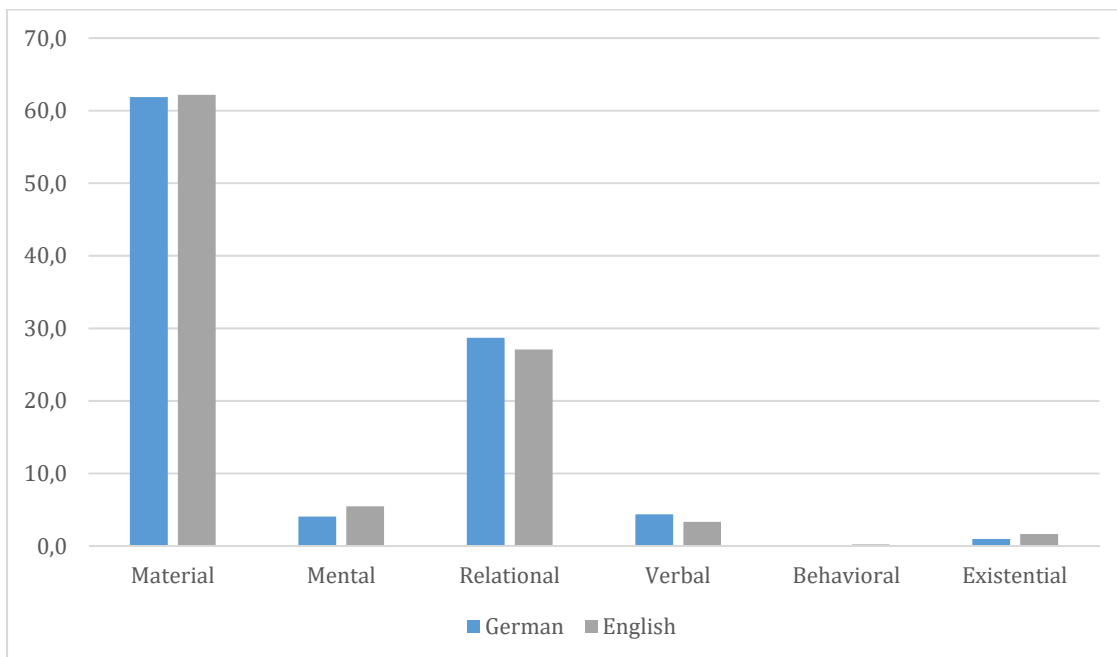


Figure A.14 Process type distribution between GO and EO in INSTR

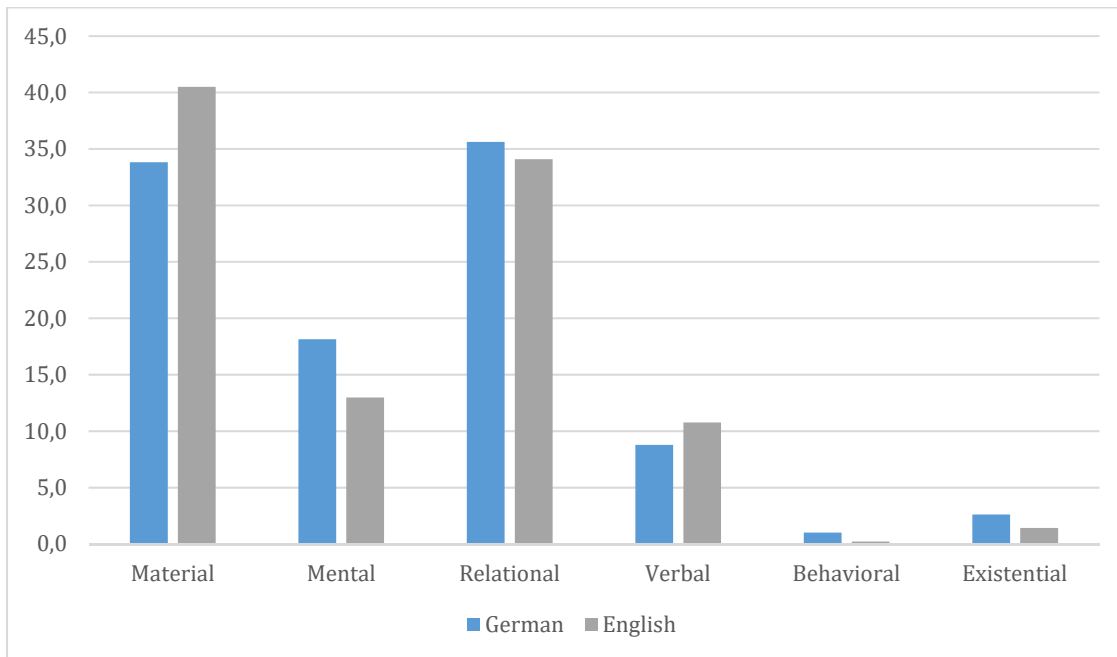


Figure A.15 Process type distribution between GO and EO in SPEECH

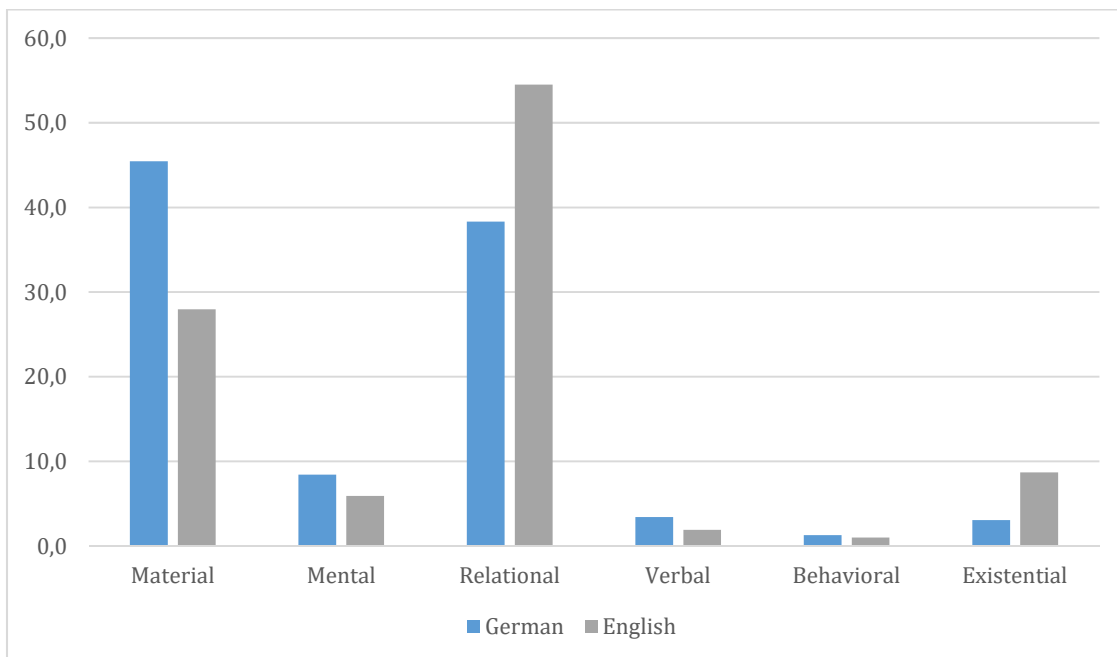


Figure A.16 Process type distribution between GO and EO in TOU

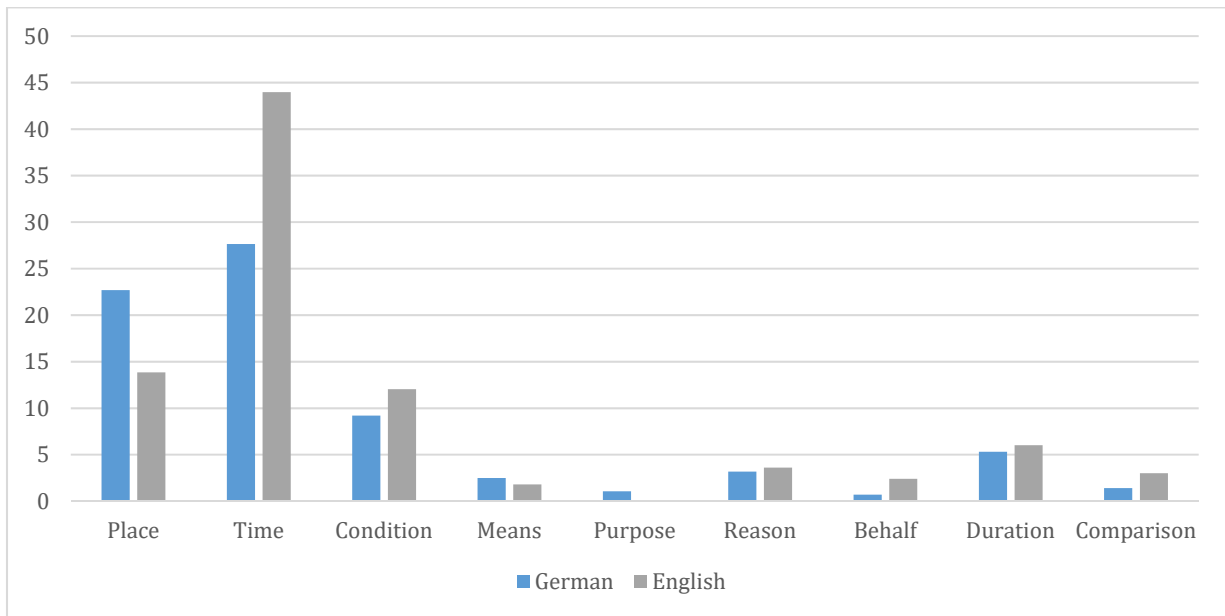


Figure A.17 Circumstance Theme distribution between GO and EO in FICTION

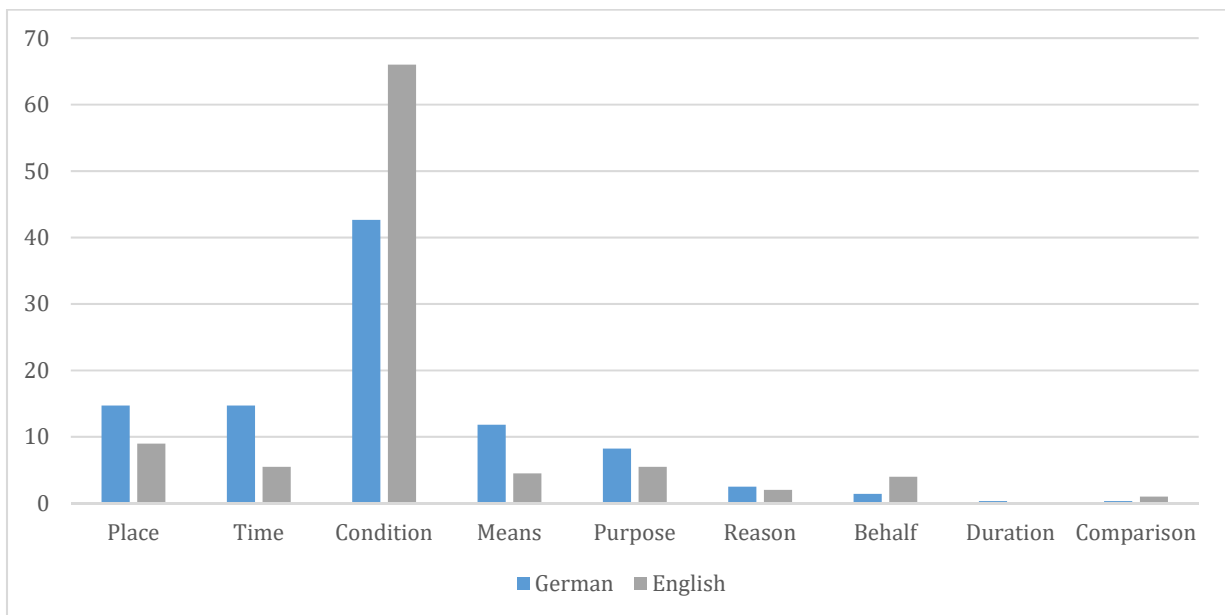


Figure A.18 Circumstance Theme distribution between GO and EO in INSTR

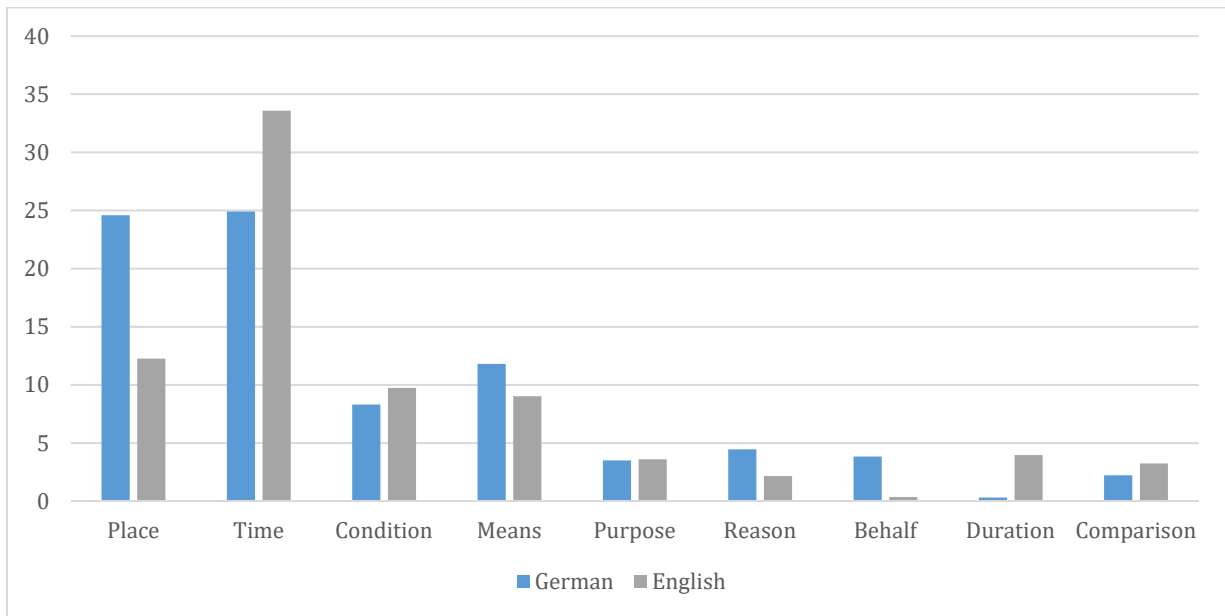


Figure A.19 Circumstance Theme distribution between GO and EO in SPEECH

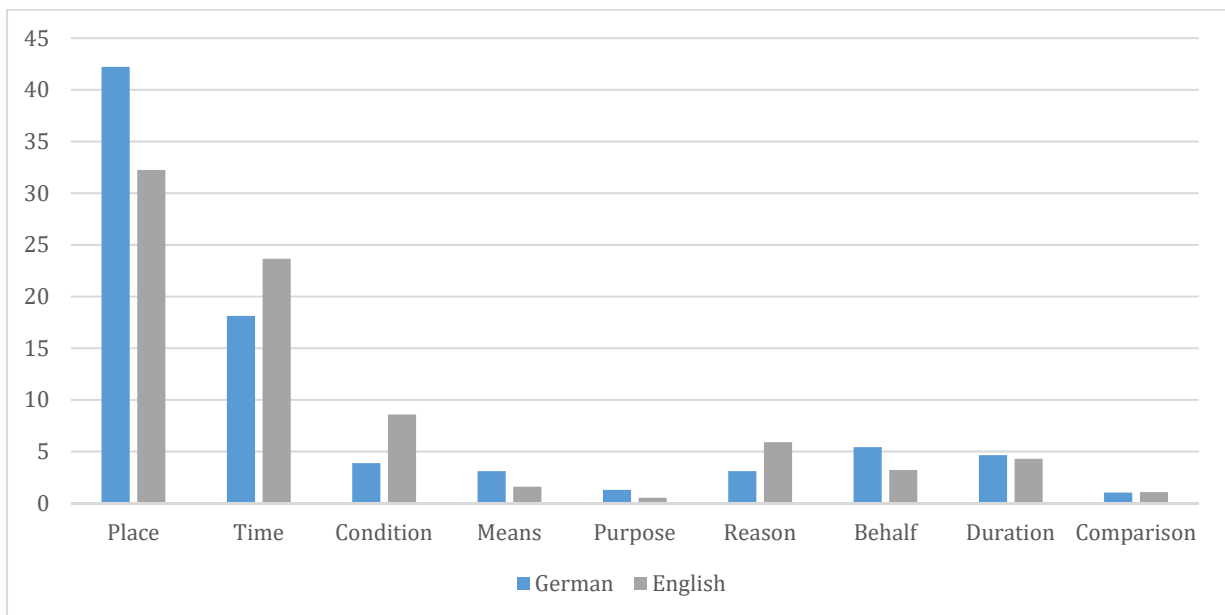


Figure A.19 Circumstance Theme distribution between GO and EO in TOU

	FICTION				INSTR			
	GO		EO		GO		EO	
Subjects in total	1292		1241		752		845	
Identifiable	1017	90.6	1019	90.9	567	80.8	603	77.6
Non-Identifiable	105	9.4	102	9.1	135	19.2	174	22.4
Other	184		120		53		68	
Pre-verbal Subjects	812		1241		388		845	
Identifiable	650	92.2	1019	90.9	280	79.5	603	77.6
Non-Identifiable	55	7.8	102	9.1	72	20.5	174	22.4
Post-verbal Subjects	494		0		367		0	
Identifiable	367	88.0	0	-	287	82.0	0	-
Non-Identifiable	50	12.0	0	-	63	18.0	0	-
Subjects in immediate post-verbal position	438		0		322		0	
Identifiable	345	90.8	0	-	265	84.9	0	-
Non-Identifiable	35	9.2	0	-	47	15.1	0	-

Table A.23 Differences in Subject Theme identifiability between GO and EO in FICTION and INSTR

	SPEECH				TOU			
	GO		EO		GO		EO	
Subjects in total	1554		1430		1258		980	
Identifiable	1169	84.6	1034	78.6	686	63.3	591	71.2
Non-Identifiable	213	15.4	282	21.4	397	36.7	239	28.8
Other	198		114		178		150	
Pre-verbal Subjects	1031		1430		669		980	
Identifiable	800	86.8	1034	78.6	379	61.3	591	71.2
Non-Identifiable	122	13.2	282	21.4	239	38.7	239	28.8
Post-verbal Subjects	549		0		592		0	
Identifiable	369	80.2	0	-	307	66.0	0	-
Non-Identifiable	91	19.8	0	-	158	34.0	0	-
Subjects in immediate post-verbal position	431		0		459		0	
Identifiable	331	84.9	0	-	255	68.2	0	-
Non-Identifiable	59	15.1	0	-	119	31.8	0	-

Table A.24 Differences in Subject Theme identifiability between GO and EO in SPEECH and TOU

Chapter 11: Inferential Analysis of Theme in English-German translations

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.32316	1.6771	0.789	0.430137	
scale(Theme.number)	5,226,007	4.743905	1.102	0.270625	
LengthCatLong	-0.007975	0.092193	-0.087	0.931068	
LengthCatShort	0.335893	0.099103	3.389	0.000701	***
LengthCatVery long	-0.142748	0.15808	-0.903	0.366521	
TextualYes	-12.064812	12.456641	-0.969	0.332772	
InterpersonalYes	-10.983615	12.458634	-0.882	0.37799	
ExpTheme.typeAdditive	0.299197	0.59525	0.503	0.615217	
ExpTheme.typeBehalf	1.493237	0.651416	2,292	0.021889	*
ExpTheme.typeComitative	0.957103	0.576789	1.659	0.097042	.
ExpTheme.typeComparison	-0.197362	0.524965	-0.376	0.706953	
ExpTheme.typeComplement	0.170746	0.476543	0.358	0.720117	
ExpTheme.typeConcession	0.44546	0.454866	0.979	0.327421	
ExpTheme.typeCondition	-0.382756	0.190243	-2.012	0.044227	*
ExpTheme.typeDuration	-0.02492	0.415829	-0.06	0.952213	
ExpTheme.typeGuise	1.027039	0.406212	2.528	0.011461	*
ExpTheme.typeMatter	-0.182862	0.717932	-0.255	0.798949	
ExpTheme.typeMeans	0.18758	0.349554	0.537	0.591526	
ExpTheme.typeOther	0.656464	0.525878	1.248	0.211914	
ExpTheme.typePlace	0.167126	0.197334	0.847	0.397039	
ExpTheme.typePurpose	0.722289	0.46689	1.547	0.121858	
ExpTheme.typeQuality	-0.418606	0.609055	-0.687	0.491891	
ExpTheme.typeReason	0.020982	0.444878	0.047	0.962383	
ExpTheme.typeSpecial	1.005722	0.337641	2.979	0.002895	**
ExpTheme.typeTime	-0.051089	0.167956	-0.304	0.760993	
Participant.ThemeActor	-0.569673	0.122613	-4.646	3.38E-06	***
Participant.ThemeAttribute	0.22821	0.441843	0.516	0.605509	
Participant.ThemeBehaver	-0.058841	0.291323	-0.202	0.839932	
Participant.ThemeEmpty	-1.883542	0.511879	-3.68	0.000234	***
Participant.ThemeExistential	2.080047	0.553874	3.755	0.000173	***
Participant.ThemeGoal	0.139585	0.189435	0.737	0.461215	
Participant.ThemeIdentified	-0.270417	0.164548	-1.643	0.100302	
Participant.ThemeIdentifier	-0.298999	0.297709	-1.004	0.315218	
Participant.ThemeInitiator	0.50745	0.298485	1.7	0.089115	.
Participant.ThemeOther	0.391464	0.434982	0.9	0.368144	
Participant.ThemePhenomenon	0.732376	0.376513	1.945	0.051757	.
Participant.ThemeReceiver	0.330997	0.711554	0.465	0.641806	
Participant.ThemeSayer	-0.958461	0.187772	-5.104	3.32E-07	***
Participant.ThemeSenser	-0.3682	0.144507	-2.548	0.010835	*

RegisterINSTR	0.400772	0.204813	1.957	0.050374	.
RegisterSPEECH	-0.438275	0.182995	-2.395	0.01662	*
RegisterTOU	0.455445	0.196885	2.313	0.020708	*
IdentifiabilityNon-Identifiable	0.878881	0.104796	8.387	< 2e-16	***
IdentifiabilityOther	1.389937	0.209861	6.623	3.52E-11	***
SentienceAnimal Sentience	1.735817	0.883851	1.964	0.049539	*
Sentienceconcrete Inanimate No Sentience	1.154036	0.273902	4.213	2.52E-05	***
Sentienceconcrete Inanimate Sentience	2.051576	0.486578	4.216	2.48E-05	***
Sentiencenonconcrete Inanimate No Sentience	0.910489	0.200964	4.531	5.88E-06	***
Sentiencenonconcrete Inanimate Sentience	1.623644	0.262853	6.177	6.53E-10	***
SentienceMachine No Sentience	1.798167	0.688454	2.612	0.009004	**
SentienceMachine Sentience	0.942824	0.530351	1.778	0.075447	.
SentienceOrganization No Sentience	1.020932	0.380271	2.685	0.007259	**
SentienceOrganization Sentience	0.794639	0.243439	3.264	0.001098	**
SentienceOther	-0.054461	0.573989	-0.095	0.924409	
SentiencePlace No Sentience	0.614456	0.340945	1.802	0.071512	.
SentiencePlace Sentience	0.9349	0.507847	1.841	0.065635	.
Sentienceprocess No Sentience	0.879995	0.402908	2.184	0.028954	*
Sentienceprocess Sentience	1.329758	0.698916	1.903	0.057093	.

Regression A.1 Regression analysis measuring the effects of different thematic aspects on binary Change between EO (first experiential element hypothesis) and GT (Forefield hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.421	0.08455	-4.979	6.38E-07	***
scale(Theme.number)	0.02126	0.04853	0.438	0.661316	
LengthCatLong	-0.05159	0.04452	-1.159	0.246546	
LengthCatShort	0.13632	0.04527	3.011	0.002602	**
LengthCatVery long	-0.09652	0.07767	-1.243	0.213954	
TextualYes	0.77052	0.14094	5.467	4.57E-08	***
InterpersonalYes	0.87566	0.14686	5.963	2.48E-09	***
ExpTheme.typeAdditive	-0.06195	0.38171	-0.162	0.871076	
ExpTheme.typeBehalf	0.21402	0.21067	1.016	0.309679	
ExpTheme.typeComitative	0.12137	0.2713	0.447	0.654618	
ExpTheme.typeComparison	-0.04561	0.26246	-0.174	0.86204	
ExpTheme.typeComplement	0.26632	0.21799	1.222	0.22182	
ExpTheme.typeConcession	0.29683	0.21774	1.363	0.17281	
ExpTheme.typeCondition	-0.47504	0.1104	-4.303	1.69E-05	***
ExpTheme.typeDuration	-0.35064	0.27045	-1.296	0.194806	
ExpTheme.typeGuise	0.26364	0.20153	1.308	0.190816	
ExpTheme.typeMatter	-0.04147	0.35853	-0.116	0.907924	
ExpTheme.typeMeans	-0.3002	0.23403	-1.283	0.199587	
ExpTheme.typeOther	-0.11828	0.29266	-0.404	0.686106	
ExpTheme.typePlace	-0.05324	0.10468	-0.509	0.611074	
ExpTheme.typePurpose	0.16076	0.24091	0.667	0.50459	
ExpTheme.typeQuality	0.14425	0.25861	0.558	0.576999	
ExpTheme.typeReason	-0.14194	0.21876	-0.649	0.516451	
ExpTheme.typeSpecial	0.33534	0.13588	2.468	0.013592	*
ExpTheme.typeTime	-0.23164	0.0914	-2.535	0.01126	*
Participant.ThemeActor	-0.20607	0.06047	-3.408	0.000655	***
Participant.ThemeAttribute	0.0124	0.207	0.06	0.952228	
Participant.ThemeBehaver	-0.15373	0.13775	-1.116	0.264427	
Participant.ThemeEmpty	-0.33815	0.20575	-1.643	0.100285	
Participant.ThemeExistential	0.30388	0.10506	2.892	0.003823	**
Participant.ThemeGoal	0.0669	0.08909	0.751	0.452709	
Participant.ThemeIdentified	-0.22368	0.08593	-2.603	0.009236	**
Participant.ThemeIdentifier	-0.04406	0.14403	-0.306	0.759683	
Participant.ThemeInitiator	0.08044	0.1138	0.707	0.479654	
Participant.ThemeOther	0.12851	0.19488	0.659	0.509609	
Participant.ThemePhenomenon	0.35749	0.14559	2.455	0.014069	*
Participant.ThemeReceiver	0.68821	0.31228	2.204	0.027538	*
Participant.ThemeSayer	-0.53629	0.09968	-5.38	7.45E-08	***
Participant.ThemeSenser	-0.08329	0.07018	-1.187	0.235347	
RegisterINSTR	0.22317	0.10582	2.109	0.034941	*
RegisterSPEECH	-0.22347	0.0963	-2.32	0.020315	*
RegisterTOU	0.27787	0.10128	2.744	0.006076	**

IdentifiabilityNon-Identifiable	0.36878	0.04633	7.959	1.73E-15	***
IdentifiabilityOther	0.33888	0.07972	4.251	2.13E-05	***
SentienceAnimal Sentience	0.27033	0.33358	0.81	0.417722	
Sentienceconcrete Inanimate No Sentience	0.4736	0.10621	4.459	8.24E-06	***
Sentienceconcrete Inanimate Sentience	0.67675	0.15648	4.325	1.53E-05	***
Sentiencenonconcrete Inanimate No Sentience	0.42351	0.0908	4.664	3.09E-06	***
Sentiencenonconcrete Inanimate Sentience	0.39642	0.11365	3.488	0.000487	***
SentienceMachine No Sentience	0.15461	0.20709	0.747	0.455297	
SentienceMachine Sentience	0.01875	0.21582	0.087	0.930767	
SentienceOrganization No Sentience	0.3121	0.18866	1.654	0.098069	.
SentienceOrganization Sentience	0.24745	0.13266	1.865	0.062142	.
SentienceOther	0.0324	0.28358	0.114	0.909036	
SentiencePlace No Sentience	0.23878	0.15293	1.561	0.118449	
SentiencePlace Sentience	0.51543	0.19928	2.586	0.009697	**
Sentienceprocess No Sentience	0.3171	0.1606	1.974	0.048329	*
Sentienceprocess Sentience	0.54586	0.24189	2.257	0.024029	*

Regression A.2 Regression analysis measuring the effects of different thematic aspects on numerical Change between EO (first experiential element hypothesis) and GT (Forefield hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.391774	0.184483	-2.124	0.033701	*
scale(Theme.number)	0.123708	0.251784	0.491	0.623195	
LengthCatLong	0.190867	0.091592	2.084	0.037171	*
LengthCatShort	-0.190732	0.097703	-1.952	0.05092	.
LengthCatVery long	-0.05308	0.157783	-0.336	0.73656	
TextualYes	1.497345	0.69426	2.157	0.031025	*
InterpersonalYes	2.223083	0.735008	3.025	0.00249	**
ExpTheme.typeAdditive	-0.272536	0.621034	-0.439	0.660775	
ExpTheme.typeBehalf	1.002523	0.589141	1.702	0.088817	.
ExpTheme.typeComitative	0.832825	0.575964	1.446	0.148187	
ExpTheme.typeComparison	-0.972076	0.551043	-1.764	0.077721	.
ExpTheme.typeComplement	0.06561	0.478798	0.137	0.891006	
ExpTheme.typeConcession	-0.365385	0.46613	-0.784	0.433117	
ExpTheme.typeCondition	-0.70173	0.192159	-3.652	0.00026	***
ExpTheme.typeDuration	-0.112627	0.415587	-0.271	0.786385	
ExpTheme.typeGuise	0.651232	0.39823	1.635	0.101983	
ExpTheme.typeMatter	-0.553945	0.726218	-0.763	0.445594	
ExpTheme.typeMeans	0.052985	0.346712	0.153	0.87854	
ExpTheme.typeOther	-0.0703	0.54556	-0.129	0.89747	
ExpTheme.typePlace	0.033807	0.196733	0.172	0.863562	
ExpTheme.typePurpose	-0.340737	0.484137	-0.704	0.481556	
ExpTheme.typeQuality	-0.12285	0.579035	-0.212	0.83198	
ExpTheme.typeReason	-0.335271	0.441642	-0.759	0.447764	
ExpTheme.typeSpecial	1.038787	0.348838	2.978	0.002903	**
ExpTheme.typeTime	-0.431734	0.173271	-2.492	0.012715	*
Participant.ThemeActor	-0.331326	0.12275	-2.699	0.006951	**
Participant.ThemeAttribute	0.541894	0.44235	1.225	0.220562	
Participant.ThemeBehaver	-0.283385	0.296985	-0.954	0.339979	
Participant.ThemeEmpty	-1.353374	0.499612	-2.709	0.006752	**
Participant.ThemeExistential	0.900208	0.360611	2.496	0.012548	*
Participant.ThemeGoal	0.173382	0.187941	0.923	0.356249	
Participant.ThemeIdentified	-0.192648	0.163832	-1.176	0.239638	
Participant.ThemeIdentifier	-0.489498	0.301759	-1.622	0.104771	
Participant.ThemeInitiator	0.741781	0.292946	2.532	0.011337	*
Participant.ThemeOther	0.341505	0.432254	0.79	0.429495	
Participant.ThemePhenomenon	0.489805	0.358321	1.367	0.171642	
Participant.ThemeReceiver	0.512766	0.715381	0.717	0.473514	
Participant.ThemeSayer	-0.841593	0.190776	-4.411	1.03E-05	***
Participant.ThemeSenser	-0.196152	0.145945	-1.344	0.178946	
RegisterINSTR	0.394841	0.206562	1.911	0.055941	.
RegisterSPEECH	-0.376939	0.185162	-2.036	0.041778	*
RegisterTOU	0.330948	0.198188	1.67	0.094945	.

IdentifiabilityNon-Identifiable	0.654897	0.10285	6.367	1.92E-10	***
IdentifiabilityOther	1.380246	0.200372	6.888	5.64E-12	***
SentienceAnimal Sentience	0.584153	0.761151	0.767	0.442809	
Sentienceconcrete Inanimate No Sentience	0.787086	0.268523	2.931	0.003377	**
Sentienceconcrete Inanimate Sentience	1.757975	0.484633	3.627	0.000286	***
Sentiencenonconcrete Inanimate No Sentience	0.701	0.199275	3.518	0.000435	***
Sentiencenonconcrete Inanimate Sentience	1.071615	0.250449	4.279	1.88E-05	***
SentienceMachine No Sentience	0.422554	0.55161	0.766	0.443654	
SentienceMachine Sentience	-0.196359	0.534788	-0.367	0.713491	
SentienceOrganization No Sentience	0.573316	0.380637	1.506	0.132016	
SentienceOrganization Sentience	0.953057	0.241803	3.941	8.10E-05	***
SentienceOther	-0.008351	0.571442	-0.015	0.988341	
SentiencePlace No Sentience	0.433399	0.337529	1.284	0.199129	
SentiencePlace Sentience	1.111656	0.521262	2.133	0.032956	*
Sentienceprocess No Sentience	0.923963	0.410043	2.253	0.024238	*
Sentienceprocess Sentience	1.3278	0.714327	1.859	0.063054	.

Regression A.3 Regression analysis measuring the effects of different thematic aspects on binary Change between EO (first experiential element hypothesis) and GT (first experiential element hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.47647	0.16713	-2.851	0.00436	**
Theme.number	-0.006051	0.136961	-0.044	0.964759	
LengthCatLong	0.00452	0.045144	0.1	0.920243	
LengthCatShort	0.119197	0.046657	2.555	0.010626	*
LengthCatVery long	-0.034861	0.079046	-0.441	0.659196	
TextualYes	0.790684	0.150619	5.25	1.52E-07	***
InterpersonalYes	0.875652	0.156852	5.583	2.37E-08	***
ExpTheme.typeAdditive	-0.364782	0.450199	-0.81	0.417786	
ExpTheme.typeBehalf	0.182332	0.215301	0.847	0.397068	
ExpTheme.typeComitative	0.02782	0.292502	0.095	0.924228	
ExpTheme.typeComparison	-0.234228	0.292531	-0.801	0.423308	
ExpTheme.typeComplement	0.067641	0.216809	0.312	0.755053	
ExpTheme.typeConcession	-0.332455	0.29553	-1.125	0.260612	
ExpTheme.typeCondition	-0.482025	0.110418	-4.365	1.27E-05	***
ExpTheme.typeDuration	-0.150332	0.253676	-0.593	0.55344	
ExpTheme.typeGuise	0.14778	0.218404	0.677	0.498637	
ExpTheme.typeMatter	-0.168587	0.382702	-0.441	0.659563	
ExpTheme.typeMeans	-0.153301	0.218423	-0.702	0.48277	
ExpTheme.typeOther	-0.346679	0.336792	-1.029	0.303312	
ExpTheme.typePlace	-0.098459	0.10978	-0.897	0.369787	
ExpTheme.typePurpose	0.212544	0.235001	0.904	0.365763	
ExpTheme.typeQuality	0.176598	0.259318	0.681	0.495867	
ExpTheme.typeReason	-0.324358	0.240843	-1.347	0.178057	
ExpTheme.typeSpecial	0.523597	0.128662	4.07	4.71E-05	***
ExpTheme.typeTime	-0.326551	0.098031	-3.331	0.000865	***
Participant.ThemeActor	-0.206788	0.062796	-3.293	0.000991	***
Participant.ThemeAttribute	0.258547	0.199562	1.296	0.195123	
Participant.ThemeBehaver	-0.322593	0.151781	-2.125	0.033555	*
Participant.ThemeEmpty	-0.246626	0.199611	-1.236	0.216633	
Participant.ThemeExistential	0.300504	0.105265	2.855	0.004307	**
Participant.ThemeGoal	0.15419	0.088595	1.74	0.081791	.
Participant.ThemeIdentified	-0.126867	0.085043	-1.492	0.135749	
Participant.ThemeIdentifier	0.033361	0.143092	0.233	0.815648	
Participant.ThemeInitiator	0.115262	0.114097	1.01	0.312394	
Participant.ThemeOther	0.385737	0.179872	2.145	0.031992	*
Participant.ThemePhenomenon	0.598424	0.134455	4.451	8.56E-06	***
Participant.ThemeReceiver	0.463155	0.363038	1.276	0.202035	
Participant.ThemeSayer	-0.629832	0.107846	-5.84	5.22E-09	***
Participant.ThemeSenser	-0.060171	0.072338	-0.832	0.405515	
RegisterINSTR	0.259443	0.115894	2.239	0.025181	*
RegisterSPEECH	-0.246341	0.106259	-2.318	0.020433	*
RegisterTOU	0.208578	0.111797	1.866	0.062086	.

IdentifiabilityNon-Identifiable	0.364878	0.047215	7.728	1.09E-14	***
IdentifiabilityOther	0.425462	0.078022	5.453	4.95E-08	***
SentienceAnimal Sentience	0.23459	0.370068	0.634	0.526138	
Sentienceconcrete Inanimate No Sentience	0.486154	0.108253	4.491	7.09E-06	***
Sentienceconcrete Inanimate Sentience	0.761083	0.152729	4.983	6.25E-07	***
Sentiencenonconcrete Inanimate No Sentience	0.400612	0.093636	4.278	1.88E-05	***
Sentiencenonconcrete Inanimate Sentience	0.386629	0.116946	3.306	0.000946	***
SentienceMachine No Sentience	0.066407	0.210434	0.316	0.752326	
SentienceMachine Sentience	0.333864	0.190636	1.751	0.079891	.
SentienceOrganization No Sentience	0.331168	0.195062	1.698	0.089554	.
SentienceOrganization Sentience	0.305931	0.133407	2.293	0.021836	*
SentienceOther	0.157111	0.274351	0.573	0.566872	
SentiencePlace No Sentience	0.276642	0.159102	1.739	0.082075	.
SentiencePlace Sentience	0.636597	0.200393	3.177	0.001489	**
Sentienceprocess No Sentience	0.556495	0.1467	3.793	0.000149	***
Sentienceprocess Sentience	0.510943	0.242638	2.106	0.035223	*

Regression A.4 Regression analysis measuring the effects of different thematic aspects on numerical Change between EO (first experiential element hypothesis) and GT (first experiential element hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.85505	0.25624	-3.337	0.000847	***
scale(Theme.number)	-0.33077	0.28568	-1.158	0.246938	
LengthCatLong	0.49587	0.1024	4.842	1.28E-06	***
LengthCatShort	0.43711	0.11336	3.856	0.000115	***
LengthCatVery long	0.74435	0.15212	4.893	9.92E-07	***
TextualYes	2.31715	0.51657	4.486	7.27E-06	***
InterpersonalYes	3.30197	0.59928	5.51	3.59E-08	***
Marked.Theme2Time	4.89966	0.69564	7.043	1.88E-12	***
Marked.ThemeAdditive	3.44463	1.16584	2.955	0.00313	**
Marked.ThemeBehalf	1.78048	0.82182	2.167	0.030272	*
Marked.ThemeComitative	18.09515	26.22874	0.69	0.490258	
Marked.ThemeComparison	17.93796	29.85396	0.601	0.547935	
Marked.ThemeComplement	16.9381	32.6162	0.519	0.603541	
Marked.ThemeConcession	3.08184	0.89882	3.429	0.000606	***
Marked.ThemeCondition	5.36961	0.8696	6.175	6.62E-10	***
Marked.ThemeDuration	18.58034	21.20061	0.876	0.380809	
Marked.ThemeGuise	18.44615	16.77834	1.099	0.271593	
Marked.ThemeMatter	18.10714	209.73776	0.086	0.931202	
Marked.ThemeMeans	4.76661	1.12814	4.225	2.39E-05	***
Marked.ThemeOther	18.36065	25.63004	0.716	0.473761	
Marked.ThemePlace	3.30693	0.60143	5.498	3.83E-08	***
Marked.ThemePurpose	4.49561	1.15016	3.909	9.28E-05	***
Marked.ThemeQuality	18.55865	60.95399	0.304	0.76077	
Marked.ThemeReason	4.22079	1.17031	3.607	0.00031	***
Participant.ThemeActor	-0.60511	0.12493	-4.843	1.28E-06	***
Participant.ThemeAttribute	0.15952	0.50719	0.315	0.753133	
Participant.ThemeBehaver	-0.49793	0.29479	-1.689	0.091205	.
Participant.ThemeEmpty	-1.43279	0.49233	-2.91	0.003612	**
Participant.ThemeExistential	-0.69517	0.29129	-2.387	0.017008	*
Participant.ThemeGoal	0.21026	0.19353	1.086	0.277283	
Participant.ThemeIdentified	-0.22323	0.16428	-1.359	0.174198	
Participant.ThemeIdentifier	-0.1755	0.30347	-0.578	0.563047	
Participant.ThemeInitiator	0.38661	0.29921	1.292	0.196325	
Participant.ThemeOther	0.17939	0.55063	0.326	0.744588	
Participant.ThemePhenomenon	0.45466	0.36833	1.234	0.217068	
Participant.ThemeReceiver	-0.43867	0.7889	-0.556	0.578172	
Participant.ThemeSayer	-0.93866	0.1845	-5.088	3.63E-07	***
Participant.ThemeSenser	-0.49428	0.14721	-3.358	0.000786	***
Participant.ThemeSpecial	-0.36257	0.38983	-0.93	0.352339	
Participant.ThemeVerbiage	1.77309	1.08988	1.627	0.103764	
RegisterINSTR	0.47126	0.25983	1.814	0.069724	.
RegisterSPEECH	-0.24993	0.23359	-1.07	0.284653	

RegisterTOU	0.22087	0.25141	0.879	0.379652	
IdentifiabilityNon-Identifiable	0.86952	0.10603	8.201	2.38E-16	***
IdentifiabilityOther	1.07452	0.20342	5.282	1.28E-07	***
SentienceAnimal Sentience	-0.32931	0.7863	-0.419	0.675356	
Sentienceconcrete Inanimate No Sentience	1.17197	0.27985	4.188	2.82E-05	***
Sentienceconcrete Inanimate Sentience	2.09708	0.51874	4.043	5.28E-05	***
Sentiencenonconcrete Inanimate No Sentience	0.72455	0.20217	3.584	0.000339	***
Sentiencenonconcrete Inanimate Sentience	1.03864	0.25286	4.108	4.00E-05	***
SentienceMachine No Sentience	1.23627	0.60997	2.027	0.042686	*
SentienceMachine Sentience	0.48204	0.51034	0.945	0.344883	
SentienceOrganization No Sentience	0.65718	0.37859	1.736	0.082584	.
SentienceOrganization Sentience	0.88446	0.24706	3.58	0.000344	***
SentienceOther	0.08448	0.5775	0.146	0.883693	
SentiencePlace No Sentience	0.75382	0.35418	2.128	0.033309	*
SentiencePlace Sentience	1.62258	0.55772	2.909	0.003622	**
Sentienceprocess No Sentience	2.02478	0.51554	3.927	8.58E-05	***
Sentienceprocess Sentience	1.99688	0.81194	2.459	0.013918	*

Regression A.5 Regression analysis measuring the effects of different thematic aspects on binary Change between EO (Subject hypothesis) and GT (Subject hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.21391	0.12319	-1.736	0.082483	.
Theme.number	-0.02623	0.08622	-0.304	0.760938	
LengthCatLong	0.11887	0.03782	3.143	0.001671	**
LengthCatShort	0.24871	0.04376	5.683	1.32E-08	***
LengthCatVery long	0.13436	0.05501	2.442	0.014592	*
TextualYes	0.43193	0.10039	4.303	1.69E-05	***
InterpersonalYes	0.64121	0.10639	6.027	1.67E-09	***
Marked.Theme2Time	0.59922	0.10741	5.579	2.42E-08	***
Marked.ThemeAdditive	0.6843	0.22598	3.028	0.002461	**
Marked.ThemeBehalf	0.29017	0.17741	1.636	0.101928	
Marked.ThemeComitative	0.72341	0.1905	3.797	0.000146	***
Marked.ThemeComparison	0.59671	0.19027	3.136	0.001712	**
Marked.ThemeComplement	0.21651	0.13971	1.55	0.121213	
Marked.ThemeConcession	0.79684	0.171	4.66	3.17E-06	***
Marked.ThemeCondition	0.55114	0.10414	5.292	1.21E-07	***
Marked.ThemeDuration	0.72212	0.16303	4.429	9.45E-06	***
Marked.ThemeGuise	0.87285	0.15774	5.534	3.14E-08	***
Marked.ThemeMatter	0.24494	0.29308	0.836	0.4033	
Marked.ThemeMeans	0.85687	0.14474	5.92	3.22E-09	***
Marked.ThemeOther	1.00192	0.18333	5.465	4.62E-08	***
Marked.ThemePlace	0.70486	0.10969	6.426	1.31E-10	***
Marked.ThemePurpose	0.797	0.18281	4.36	1.30E-05	***
Marked.ThemeQuality	0.49045	0.24641	1.99	0.046545	*
Marked.ThemeReason	0.40566	0.16434	2.468	0.013572	*
Participant.ThemeActor	-0.33698	0.04934	-6.83	8.48E-12	***
Participant.ThemeAttribute	0.03561	0.19321	0.184	0.853774	
Participant.ThemeBehaver	-0.35951	0.12511	-2.874	0.004059	**
Participant.ThemeEmpty	-0.33868	0.16609	-2.039	0.041441	*
Participant.ThemeExistential	0.24618	0.08887	2.77	0.005606	**
Participant.ThemeGoal	0.13837	0.06523	2.121	0.033908	*
Participant.ThemeIdentified	-0.12951	0.07089	-1.827	0.067695	.
Participant.ThemeIdentifier	0.16113	0.119	1.354	0.175729	
Participant.ThemeInitiator	0.11431	0.08757	1.305	0.191792	
Participant.ThemeOther	0.18346	0.20306	0.903	0.366262	
Participant.ThemePhenomenon	0.39259	0.10607	3.701	0.000215	***
Participant.ThemeReceiver	-0.06081	0.23492	-0.259	0.795739	
Participant.ThemeSayer	-0.56209	0.07878	-7.135	9.68E-13	***
Participant.ThemeSenser	-0.16398	0.05768	-2.843	0.004473	**
Participant.ThemeSpecial	0.07523	0.12835	0.586	0.55779	
Participant.ThemeVerbiage	0.25463	0.22214	1.146	0.251696	
RegisterINSTR	0.24233	0.1149	2.109	0.034938	*
RegisterSPEECH	-0.28169	0.10605	-2.656	0.007903	**

RegisterTOU	0.09941	0.11157	0.891	0.372948	
IdentifiabilityNon-Identifiable	0.36339	0.03622	10.034	< 2e-16	***
IdentifiabilityOther	0.30068	0.06054	4.967	6.80E-07	***
SentienceAnimal Sentience	-0.0656	0.3155	-0.208	0.835299	
Sentienceconcrete Inanimate No Sentience	0.51209	0.08322	6.153	7.58E-10	***
Sentienceconcrete Inanimate Sentience	0.74934	0.11661	6.426	1.31E-10	***
Sentiencenonconcrete Inanimate No Sentience	0.48583	0.07439	6.531	6.53E-11	***
Sentiencenonconcrete Inanimate Sentience	0.63837	0.089	7.173	7.35E-13	***
SentienceMachine No Sentience	0.34676	0.14455	2.399	0.016441	*
SentienceMachine Sentience	0.39506	0.11771	3.356	0.00079	***
SentienceOrganization No Sentience	0.37159	0.14453	2.571	0.010139	*
SentienceOrganization Sentience	0.44314	0.09526	4.652	3.29E-06	***
SentienceOther	0.09338	0.23304	0.401	0.688647	
SentiencePlace No Sentience	0.41191	0.12097	3.405	0.000662	***
SentiencePlace Sentience	0.69776	0.15908	4.386	1.15E-05	***
Sentienceprocess No Sentience	0.7595	0.11579	6.559	5.41E-11	***
Sentienceprocess Sentience	1.00955	0.16011	6.305	2.88E-10	***

Regression A.6 Regression analysis measuring the effects of different thematic aspects on numerical Change between EO (Subject hypothesis) and GT (Subject hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.06286	0.14608	-0.43	0.666951	
scale(Theme.number)	0.78188	0.2082	3.755	0.000173	***
LengthCatLong	-0.68615	0.0905	-7.582	3.40E-14	***
LengthCatShort	-0.57975	0.0916	-6.329	2.46E-10	***
LengthCatVery long	-1.09285	0.15516	-7.044	1.87E-12	***
TextualYes	-2.49697	0.76338	-3.271	0.001072	**
InterpersonalYes	-1.88248	0.80879	-2.328	0.019937	*
ExpTheme.typeAdditive	0.31481	0.55602	0.566	0.571268	
ExpTheme.typeBehalf	0.48874	0.36418	1.342	0.179587	
ExpTheme.typeComitative	0.70899	0.33062	2.144	0.031997	*
ExpTheme.typeComparison	-0.53844	0.5894	-0.914	0.360959	
ExpTheme.typeComplement	3.21939	0.34428	9.351	< 2e-16	***
ExpTheme.typeConcession	-1.34015	0.76857	-1.744	0.081213	.
ExpTheme.typeCondition	-0.28865	0.19046	-1.516	0.129631	
ExpTheme.typeDuration	0.98807	0.3945	2.505	0.012259	*
ExpTheme.typeFrequency	-0.07897	0.52692	-0.15	0.88086	
ExpTheme.typeGuise	1.18495	0.55202	2.147	0.031826	*
ExpTheme.typeMatter	1.20256	0.53149	2.263	0.023659	*
ExpTheme.typeMeans	1.54975	0.29061	5.333	9.67E-08	***
ExpTheme.typeOther	0.61874	0.66564	0.93	0.352605	
ExpTheme.typePlace	0.74855	0.14264	5.248	1.54E-07	***
ExpTheme.typeprocess	18.55698	65.59858	0.283	0.777264	
ExpTheme.typePurpose	0.46693	0.33541	1.392	0.163889	
ExpTheme.typeQuality	2.92301	0.6093	4.797	1.61E-06	***
ExpTheme.typeReason	0.51346	0.34058	1.508	0.131651	
ExpTheme.typeSpecial	-0.16344	0.37453	-0.436	0.662553	
ExpTheme.typeTime	-0.17737	0.15423	-1.15	0.250122	
Participant.ThemeActor	0.12464	0.10469	1.191	0.233798	
Participant.ThemeAttribute	-0.1145	0.38174	-0.3	0.764221	
Participant.ThemeBehaver	-0.5801	0.29932	-1.938	0.052611	.
Participant.ThemeBeneficiary	0.80539	0.73719	1.093	0.274608	
Participant.ThemeEmpty	-0.05056	0.39914	-0.127	0.899195	
Participant.ThemeExistent	15.30439	82.15752	0.186	0.852224	
Participant.ThemeExistential	1.04693	0.57662	1.816	0.069426	.
Participant.ThemeGoal	0.14243	0.17324	0.822	0.410995	
Participant.ThemeIdentified	-0.28263	0.18002	-1.57	0.11641	
Participant.ThemeIdentifier	0.57287	0.26651	2.149	0.031596	*
Participant.ThemeInitiator	0.62071	0.28958	2.143	0.032073	*
Participant.ThemeOther	0.85521	1.1007	0.777	0.437176	
Participant.ThemePhenomenon	0.54802	0.30321	1.807	0.070702	.
Participant.ThemeSayer	-0.25504	0.17219	-1.481	0.138566	
Participant.ThemeSenser	-0.33841	0.14872	-2.275	0.02288	*
Participant.ThemeVerbiage	0.01976	0.49545	0.04	0.968183	

RegisterINSTR	0.29519	0.16681	1.77	0.076793	.
RegisterSPEECH	0.8694	0.13929	6.242	4.32E-10	***
RegisterTOU	0.77083	0.14403	5.352	8.70E-08	***
IdentifiabilityNon-Identifiable	0.12754	0.11245	1.134	0.256729	
IdentifiabilityOther	1.61614	0.18882	8.559	< 2e-16	***

Regression A.7 Regression analysis measuring the effects of different thematic aspects on binary Change between GO (Forefield hypothesis) and ET (first experiential element hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.62624	0.21419	-2.924	0.003459	**
scale(Theme.number)	0.05896	0.0249	2.368	0.017884	*
LengthCatLong	-0.17705	0.03959	-4.472	7.76E-06	***
LengthCatShort	-0.12274	0.04155	-2.954	0.003138	**
LengthCatVery long	-0.35794	0.07532	-4.752	2.01E-06	***
TextualYes	0.08731	0.10463	0.835	0.403981	
InterpersonalYes	0.18355	0.10795	1.7	0.089072	.
ExpTheme.typeAdditive	0.63162	0.32841	1.923	0.054442	.
ExpTheme.typeBehalf	0.61573	0.26008	2.368	0.017908	*
ExpTheme.typeComitative	0.50603	0.26157	1.935	0.053045	.
ExpTheme.typeComparison	-0.18299	0.43354	-0.422	0.67296	
ExpTheme.typeComplement	1.24802	0.19408	6.43	1.27E-10	***
ExpTheme.typeConcession	-1.03493	0.61598	-1.68	0.092931	.
ExpTheme.typeCondition	-0.1329	0.23968	-0.555	0.579236	
ExpTheme.typeDuration	0.62853	0.26549	2.367	0.017911	*
ExpTheme.typeFrequency	0.09608	0.36828	0.261	0.794188	
ExpTheme.typeGuise	0.80749	0.30108	2.682	0.007319	**
ExpTheme.typeMatter	0.65778	0.28632	2.297	0.021597	*
ExpTheme.typeMeans	0.6551	0.23402	2.799	0.00512	**
ExpTheme.typeOther	0.57168	0.34166	1.673	0.094282	.
ExpTheme.typePlace	0.50543	0.21774	2.321	0.020273	*
ExpTheme.typeprocess	1.02262	0.30016	3.407	0.000657	***
ExpTheme.typePurpose	0.4438	0.27211	1.631	0.102898	
ExpTheme.typeQuality	1.04061	0.24185	4.303	1.69E-05	***
ExpTheme.typeReason	0.53481	0.26311	2.033	0.042086	*
ExpTheme.typeSpecial	0.3549	0.27942	1.27	0.204036	
ExpTheme.typeTime	0.14278	0.22321	0.64	0.522378	
Participant.ThemeActor	0.27786	0.2132	1.303	0.192471	
Participant.ThemeAttribute	0.07652	0.09459	0.809	0.418534	
Participant.ThemeBehaver	-0.2068	0.27507	-0.752	0.452175	
Participant.ThemeBeneficiary	0.16807	0.19896	0.845	0.398274	
Participant.ThemeEmpty	0.31732	0.12153	2.611	0.009025	**
Participant.ThemeExistent	0.09662	0.15951	0.606	0.544715	
Participant.ThemeExistential	-0.03893	0.17624	-0.221	0.825166	
Participant.ThemeGoal	0.15983	0.07529	2.123	0.033768	*
Participant.ThemeIdentified	-0.07206	0.09506	-0.758	0.448388	
Participant.ThemeIdentifier	0.25913	0.11688	2.217	0.026621	*
Participant.ThemeInitiator	0.35703	0.25011	1.428	0.153432	
Participant.ThemeOther	0.09312	0.15447	0.603	0.546598	
Participant.ThemePhenomenon	0.36565	0.09958	3.672	0.000241	***
Participant.ThemeSayer	0.07938	0.22619	0.351	0.725616	
Participant.ThemeSenser	0.24332	0.20391	1.193	0.232759	
Participant.ThemeVerbiage	0.08062	0.14433	0.559	0.576462	

RegisterINSTR	0.07736	0.06764	1.144	0.252757	
RegisterSPEECH	0.31883	0.05351	5.959	2.54E-09	***
RegisterTOU	0.22987	0.05582	4.118	3.82E-05	***
IdentifiabilityNon-Identifiable	0.0689	0.05706	1.207	0.227266	
IdentifiabilityOther	0.72491	0.06559	11.052	< 2e-16	***
SentienceAnimal Sentience	-0.59213	0.58182	-1.018	0.308811	
Sentienceconcrete Inanimate No Sentience	0.14255	0.12232	1.165	0.243855	
Sentienceconcrete Inanimate Sentience	-0.10803	0.19715	-0.548	0.583705	
Sentiencenonconcrete Inanimate No Sentience	0.24326	0.09896	2.458	0.013962	*
Sentiencenonconcrete Inanimate Sentience	0.46524	0.11578	4.018	5.86E-05	***
SentienceMachine No Sentience	0.04233	0.28711	0.147	0.882785	
SentienceMachine Sentience	0.04553	0.27571	0.165	0.868837	
Sentience Not analyzed	0.25371	0.20751	1.223	0.22146	
SentienceOrganization No Sentience	0.03356	0.19877	0.169	0.865942	
SentienceOrganization Sentience	0.25092	0.13357	1.879	0.060301	.
SentienceOther	-0.07682	0.33962	-0.226	0.821053	
SentiencePlace No Sentience	0.29188	0.17265	1.691	0.090914	.
SentiencePlace Sentience	0.48114	0.20167	2.386	0.017042	*
Sentienceprocess No Sentience	0.2532	0.18884	1.341	0.17997	

Regression A.8 Regression analysis measuring the effects of different thematic aspects on numerical Change between GO (Forefield hypothesis) and ET (first experiential element hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	0.670377	0.16955	3.954	7.69E-05	***
scale(Theme.number)	2.524176	0.280447	9.001	< 2e-16	***
LengthCatLong	-0.66822	0.089602	-7.458	8.81E-14	***
LengthCatShort	-0.563063	0.090877	-6.196	5.80E-10	***
LengthCatVery long	-1.064793	0.15353	-6.935	4.05E-12	***
TextualYes	-3.664901	0.507637	-7.22	5.22E-13	***
InterpersonalYes	-3.303738	0.563637	-5.861	4.59E-09	***
ExpTheme.typeAdditive	0.08122	0.543385	0.149	0.8812	
ExpTheme.typeBehalf	0.484358	0.363136	1.334	0.1823	
ExpTheme.typeComitative	0.703229	0.329999	2.131	0.0331	*
ExpTheme.typeComparison	-0.434154	0.541927	-0.801	0.4231	
ExpTheme.typeComplement	3.083737	0.326665	9.44	< 2e-16	***
ExpTheme.typeConcession	-1.112857	0.649048	-1.715	0.0864	.
ExpTheme.typeCondition	-0.325215	0.188534	-1.725	0.0845	.
ExpTheme.typeDuration	1.015842	0.395279	2.57	0.0102	*
ExpTheme.typeFrequency	-0.235292	0.515882	-0.456	0.6483	
ExpTheme.typeGuise	1.274549	0.546505	2.332	0.0197	*
ExpTheme.typeMatter	0.964112	0.468137	2.059	0.0394	*
ExpTheme.typeMeans	1.548214	0.289736	5.344	9.12E-08	***
ExpTheme.typeOther	0.693699	0.649547	1.068	0.2855	
ExpTheme.typePlace	0.73387	0.141846	5.174	2.30E-07	***
ExpTheme.typeprocess	16.044212	60.487216	0.265	0.7908	
ExpTheme.typePurpose	0.443311	0.334855	1.324	0.1855	
ExpTheme.typeQuality	2.923518	0.609131	4.799	1.59E-06	***
ExpTheme.typeReason	0.439473	0.336953	1.304	0.1921	
ExpTheme.typeSpecial	-0.147787	0.371663	-0.398	0.6909	
ExpTheme.typeTime	-0.292184	0.152205	-1.92	0.0549	.
Participant.ThemeActor	0.12107	0.10404	1.164	0.2446	
Participant.ThemeAttribute	-0.117793	0.370123	-0.318	0.7503	
Participant.ThemeBehaver	-0.574934	0.295703	-1.944	0.0519	.
Participant.ThemeBeneficiary	0.800417	0.732043	1.093	0.2742	
Participant.ThemeEmpty	-0.006459	0.396063	-0.016	0.987	
Participant.ThemeExistent	13.431476	29.434716	0.456	0.6482	
Participant.ThemeExistential	0.25613	0.448404	0.571	0.5679	
Participant.ThemeGoal	0.151688	0.171777	0.883	0.3772	
Participant.ThemeIdentified	-0.282688	0.179061	-1.579	0.1144	
Participant.ThemeIdentifier	0.561812	0.265888	2.113	0.0346	*
Participant.ThemeInitiator	0.609672	0.289012	2.11	0.0349	*
Participant.ThemeOther	1.025336	1.095831	0.936	0.3494	
Participant.ThemePhenomenon	0.486061	0.29816	1.63	0.1031	
Participant.ThemeSayer	-0.266307	0.170939	-1.558	0.1193	
Participant.ThemeSenser	-0.321815	0.147338	-2.184	0.0289	*
Participant.ThemeVerbiage	0.046494	0.489442	0.095	0.9243	

RegisterINSTR	0.31972	0.163297	1.958	0.0502	.
RegisterSPEECH	0.865339	0.136181	6.354	2.09E-10	***
RegisterTOU	0.784242	0.141108	5.558	2.73E-08	***
IdentifiabilityNon-Identifiable	0.129438	0.111842	1.157	0.2471	
IdentifiabilityOther	1.570315	0.186224	8.432	< 2e-16	***

Regression A.9 Regression analysis measuring the effects of different thematic aspects on binary Change between GO (first experiential element hypothesis) and ET (first experiential element hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.793625	0.180922	-4.387	1.15E-05	***
Theme.number	0.463854	0.047832	9.698	< 2e-16	***
LengthCatLong	-0.143104	0.036847	-3.884	0.000103	***
LengthCatShort	-0.128064	0.038394	-3.336	0.000851	***
LengthCatVery long	-0.231367	0.068625	-3.371	0.000748	***
TextualYes	-0.081873	0.08682	-0.943	0.345671	
InterpersonalYes	-0.026114	0.093848	-0.278	0.780816	
ExpTheme.typeAdditive	0.056301	0.31686	0.178	0.858971	
ExpTheme.typeBehalf	0.198794	0.220441	0.902	0.367162	
ExpTheme.typeComitative	0.095362	0.226347	0.421	0.673528	
ExpTheme.typeComparison	-0.588926	0.391854	-1.503	0.13286	
ExpTheme.typeComplement	0.810344	0.151498	5.349	8.85E-08	***
ExpTheme.typeConcession	-1.338728	0.528634	-2.532	0.011328	*
ExpTheme.typeCondition	-0.635996	0.204316	-3.113	0.001853	**
ExpTheme.typeDuration	0.211811	0.23634	0.896	0.370139	
ExpTheme.typeFrequency	-0.300019	0.334683	-0.896	0.370024	
ExpTheme.typeGuise	0.307697	0.273634	1.124	0.260808	
ExpTheme.typeMatter	0.139333	0.258861	0.538	0.590401	
ExpTheme.typeMeans	0.223701	0.195604	1.144	0.252773	
ExpTheme.typeOther	0.136774	0.325801	0.42	0.674625	
ExpTheme.typePlace	0.10102	0.176435	0.573	0.566941	
ExpTheme.typeprocess	0.490844	0.23951	2.049	0.040426	*
ExpTheme.typePurpose	0.005723	0.240549	0.024	0.981017	
ExpTheme.typeQuality	0.679858	0.205913	3.302	0.000961	***
ExpTheme.typeReason	0.038989	0.23053	0.169	0.865695	
ExpTheme.typeSpecial	0.067695	0.227314	0.298	0.765853	
ExpTheme.typeTime	-0.391285	0.185342	-2.111	0.034759	*
Participant.ThemeActor	-0.022061	0.172492	-0.128	0.898232	
Participant.ThemeAttribute	0.074536	0.087112	0.856	0.392195	
Participant.ThemeBehaver	-0.413268	0.234326	-1.764	0.077792	.
Participant.ThemeBeneficiary	0.040714	0.154871	0.263	0.792635	
Participant.ThemeEmpty	0.199408	0.099705	2.000	0.045502	*
Participant.ThemeExistent	0.126857	0.155623	0.815	0.414985	
Participant.ThemeExistential	-0.130208	0.121377	-1.073	0.283381	
Participant.ThemeGoal	0.156186	0.064825	2.409	0.01598	*
Participant.ThemeIdentified	-0.125953	0.086813	-1.451	0.146822	
Participant.ThemeIdentifier	0.206563	0.112759	1.832	0.066966	.
Participant.ThemeInitiator	-0.119487	0.214373	-0.557	0.577266	
Participant.ThemeOther	0.09052	0.138074	0.656	0.512087	
Participant.ThemePhenomenon	0.324464	0.093268	3.479	0.000504	***
Participant.ThemeSayer	-0.169369	0.183207	-0.924	0.355241	
Participant.ThemeSenser	-0.057655	0.160885	-0.358	0.720074	
Participant.ThemeVerbiage	0.077967	0.140909	0.553	0.580049	

RegisterINSTR	0.186885	0.077981	2.397	0.01655 *
RegisterSPEECH	0.394586	0.063169	6.246	4.20E-10 ***
RegisterTOU	0.311109	0.064437	4.828	1.38E-06 ***
IdentifiabilityNon-Identifiable	0.05963	0.04953	1.204	0.228624
IdentifiabilityOther	0.587218	0.060932	9.637	< 2e-16 ***
SentienceAnimal Sentience	-0.616357	0.50559	-1.219	0.222812
Sentienceconcrete Inanimate No Sentience	0.121384	0.110076	1.103	0.270145
Sentienceconcrete Inanimate Sentience	-0.150363	0.178384	-0.843	0.399274
Sentiencenonconcrete Inanimate No Sentience	0.193142	0.083401	2.316	0.020568 *
Sentiencenonconcrete Inanimate Sentience	0.277527	0.11241	2.469	0.013554 *
SentienceMachine No Sentience	-0.181789	0.267747	-0.679	0.497164
SentienceMachine Sentience	-0.138326	0.24445	-0.566	0.571484
Sentience Not analyzed	-0.081734	0.1658	-0.493	0.622037
SentienceOrganization No Sentience	-0.004277	0.163906	-0.026	0.979183
SentienceOrganization Sentience	0.073611	0.115628	0.637	0.524371
SentienceOther	0.074163	0.236451	0.314	0.753787
SentiencePlace No Sentience	0.150753	0.165871	0.909	0.363425
SentiencePlace Sentience	0.369471	0.179792	2.055	0.03988 *
Sentienceprocess No Sentience	0.131302	0.175479	0.748	0.45431

Regression A.10 Regression analysis measuring the effects of different thematic aspects on numerical Change between GO (first experiential element hypothesis) and ET (first experiential element hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.85614	0.2171	13.156	< 2e-16	***
scale(Theme.number)	3.44304	0.21325	16.146	< 2e-16	***
LengthCatLong	-0.072	0.10809	-0.666	0.505349	
LengthCatShort	-0.50132	0.11808	-4.246	2.18E-05	***
LengthCatVery long	-0.7612	0.16776	-4.537	5.70E-06	***
TextualYes	-2.4018	0.28679	-8.375	< 2e-16	***
InterpersonalYes	-1.89818	0.40255	-4.715	2.41E-06	***
Participant.ThemeActor	-0.049	0.1363	-0.359	0.719228	
Participant.ThemeAttribute	1.03571	0.56321	1.839	0.065922	.
Participant.ThemeBehaver	-0.66771	0.31283	-2.134	0.03281	*
Participant.ThemeBeneficiary	0.30095	0.74292	0.405	0.685409	
Participant.ThemeDummy	12.04161	56.0575	0.215	0.829917	
Participant.ThemeEmpty	0.3614	0.39864	0.907	0.364627	
Participant.ThemeExistent	9.46367	29.02917	0.326	0.74442	
Participant.ThemeExistential	-1.31492	0.39119	-3.361	0.000776	***
Participant.ThemeGoal	-0.1094	0.17656	-0.62	0.535501	
Participant.ThemeIdentified	0.02454	0.18216	0.135	0.892831	
Participant.ThemeIdentifier	0.2274	0.26303	0.865	0.387281	
Participant.ThemeInitiator	0.03187	0.32395	0.098	0.92164	
Participant.ThemeOther	14.22628	129.67743	0.11	0.912643	
Participant.ThemePhenomenon	0.57558	0.32171	1.789	0.073592	.
Participant.ThemeSayer	-0.05621	0.18479	-0.304	0.760989	
Participant.ThemeSenser	-0.10202	0.15648	-0.652	0.514437	
Participant.ThemeSpecial	-2.04926	0.41424	-4.947	7.54E-07	***
Participant.ThemeVerbiage	0.6945	0.5898	1.178	0.238995	
IdentifiabilityNon-Identifiable	0.1992	0.11702	1.702	0.088712	.
IdentifiabilityOther	1.51453	0.18497	8.188	2.66E-16	***
SentienceAnimal Sentience	-0.90172	0.99754	-0.904	0.366026	
Sentienceconcrete Inanimate No Sentience	0.21754	0.25133	0.866	0.386722	
Sentienceconcrete Inanimate Sentience	0.20302	0.34309	0.592	0.554023	
Sentiencenonconcrete Inanimate No Sentience	0.40226	0.21621	1.861	0.062811	.
Sentiencenonconcrete Inanimate Sentience	0.72938	0.28216	2.585	0.009737	**
SentienceMachine No Sentience	-0.04418	0.54952	-0.08	0.935914	
SentienceMachine Sentience	0.37284	0.5261	0.709	0.478516	
SentienceOrganization No Sentience	-0.20172	0.39926	-0.505	0.613386	

SentienceOrganization Sentience	0.53038	0.3035	1.748	0.080536 .
SentienceOther	-0.34584	0.63548	-0.544	0.586298
SentiencePlace No Sentience	0.36444	0.38023	0.958	0.337818
SentiencePlace Sentience	0.10034	0.46243	0.217	0.828214
Sentienceprocess No Sentience	0.77983	0.47635	1.637	0.101612

Regression A.11 Regression analysis measuring the effects of different thematic aspects on binary Change between GO (Subject hypothesis) and ET (Subject hypothesis)

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.723994	0.070053	10.335	< 2e-16	***
Theme.number	0.299533	0.027635	10.839	< 2e-16	***
LengthCatLong	-0.014051	0.032105	-0.438	0.661637	
LengthCatShort	-0.091264	0.03613	-2.526	0.011537	*
LengthCatVery long	-0.089997	0.046025	-1.955	0.050538	.
TextualYes	0.081398	0.051137	1.592	0.111438	
InterpersonalYes	0.040417	0.055353	0.73	0.465282	
Marked.Theme2Time	0.009413	0.072454	0.13	0.896632	
Marked.ThemeAdditive	-0.100155	0.186484	-0.537	0.591217	
Marked.ThemeBehalf	0.152151	0.116695	1.304	0.19229	
Marked.ThemeComitative	-0.035439	0.12351	-0.287	0.774168	
Marked.ThemeComparison	-0.304031	0.238821	-1.273	0.203	
Marked.ThemeComplement	0.226939	0.066255	3.425	0.000614	***
Marked.ThemeConcession	0.259542	0.155749	1.666	0.095632	
Marked.ThemeCondition	0.103945	0.078356	1.327	0.184648	
Marked.ThemeDuration	0.019191	0.137871	0.139	0.889295	
Marked.ThemeFrequency	-0.127521	0.213158	-0.598	0.549676	
Marked.ThemeGuise	0.200755	0.168682	1.19	0.233992	
Marked.ThemeMatter	0.123605	0.146381	0.844	0.39844	
Marked.ThemeMeans	0.275559	0.089337	3.084	0.002039	**
Marked.ThemeOther	0.213041	0.193016	1.104	0.269703	
Marked.ThemePlace	0.105702	0.06606	1.6	0.109575	
Marked.ThemePurpose	0.111684	0.124994	0.894	0.371583	
Marked.ThemeQuality	0.259114	0.119349	2.171	0.029927	*
Marked.ThemeReason	0.120963	0.11807	1.025	0.305596	
Participant.ThemeActor	0.050806	0.041965	1.211	0.226014	
Participant.ThemeAttribute	0.163334	0.123903	1.318	0.187424	
Participant.ThemeBehaver	-0.035992	0.11229	-0.321	0.74857	
Participant.ThemeBeneficiary	0.129449	0.162286	0.798	0.425069	
Participant.ThemeDummy	0.478244	0.152319	3.14	0.001691	
Participant.ThemeEmpty	0.371856	0.071732	5.184	2.17E-07	***
Participant.ThemeExistent	0.18457	0.122762	1.503	0.132716	
Participant.ThemeExistential	0.121198	0.076579	1.583	0.1135	
Participant.ThemeGoal	0.087179	0.049357	1.766	0.077344	.
Participant.ThemeIdentified	-0.10753	0.075443	-1.425	0.154064	
Participant.ThemeIdentifier	0.078522	0.108984	0.72	0.471221	
Participant.ThemeInitiator	0.037301	0.114253	0.326	0.744063	
Participant.ThemeOther	0.50413	0.16263	3.1	0.001936	**
Participant.ThemePhenomenon	0.19595	0.071048	2.758	0.005816	**
Participant.ThemeSayer	0.007033	0.062388	0.113	0.910247	
Participant.ThemeSenser	0.070221	0.050021	1.404	0.160366	
Participant.ThemeSpecial	-0.534618	0.159114	-3.36	0.00078	***

Participant.ThemeVerbiage	0.180319	0.120092	1.502	0.133222	
RegisterINSTR	0.216228	0.083059	2.603	0.009233	**
RegisterSPEECH	0.326586	0.069656	4.689	2.75E-06	***
RegisterTOU	0.384804	0.069101	5.569	2.57E-08	***
IdentifiabilityNon-Identifiable	0.134896	0.03332	4.048	5.16E-05	***
IdentifiabilityOther	0.409679	0.041047	9.981	< 2e-16	***
SentienceAnimal Sentience	-0.093638	0.229662	-0.408	0.683479	
Sentienceconcrete Inanimate No Sentience	0.061557	0.074111	0.831	0.406197	
Sentienceconcrete Inanimate Sentience	-0.028193	0.137706	-0.205	0.837779	
Sentiencenonconcrete Inanimate No Sentience	0.226766	0.060799	3.73	0.000192	***
Sentiencenonconcrete Inanimate Sentience	0.169931	0.092682	1.833	0.066732	.
SentienceMachine No Sentience	-0.334426	0.225349	-1.484	0.137799	
SentienceMachine Sentience	-0.158248	0.200165	-0.791	0.429185	
SentienceOrganization No Sentience	-0.082716	0.117776	-0.702	0.482481	
SentienceOrganization Sentience	0.10323	0.089076	1.159	0.246497	
SentienceOther	0.019493	0.156483	0.125	0.900865	
SentiencePlace No Sentience	0.247901	0.084279	2.941	0.003267	**
SentiencePlace Sentience	0.242976	0.116864	2.079	0.037606	*
Sentienceprocess No Sentience	0.215654	0.143282	1.505	0.132298	

Regression A.12 Regression analysis measuring the effects of different thematic aspects on numerical Change between GO (Subject hypothesis) and ET (Subject hypothesis)