

On the Role of Textual Connectives in Sentence Comprehension: A New Dataset for Italian

Giorgia Albertin[•], Alessio Miaschi^{*◊}, Dominique Brunato[◊]

[•] University of Padova

^{*}Department of Computer Science, University of Pisa

[◊]Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC-CNR)

ItaliaNLP Lab - www.italianlp.it

giorgia.albertin.2@studenti.unipd.it,

alessio.miaschi@phd.unipi.it, dominique.brunato@ilc.cnr.it

Abstract

In this paper we present a new evaluation resource for Italian aimed at assessing the role of textual connectives in the comprehension of the meaning of a sentence. The resource is arranged in two sections (acceptability assessment and cloze test), each one corresponding to a distinct challenge task conceived to test how subtle modifications involving connectives in real usage sentences influence the perceived acceptability of the sentence by native speakers and Neural Language Models (NLMs). Although the main focus is the presentation of the dataset, we also provide some preliminary data comparing human judgments and NLMs performance in the two tasks¹.

1 Introduction

The outstanding performance reached by recent Neural Language Models (NLMs) across a variety of NLP tasks that require extensive linguistic skills has stimulated an increased interest in the theoretical and computational linguistics community towards a better understanding of their inner mechanisms. In particular, the debate is focused on trying to understand what kind of linguistic knowledge these models are able to induce from the raw data they are exposed to and to what extent this knowledge resembles human-like generalization patterns (Linzen and Baroni, 2021; Manning, 2015). To pursue this investigation, it has become of pivotal importance the availability of challenging test sets, also called ‘diagnostic’ or ‘stress’ tests, built to probe the sensitivity of a model to specific language phenomena.

So far, most of the efforts have been focused on assessing the syntactic abilities encoded by NLMs by exploiting human curated benchmarks, which are usually proposed in the form of minimal sentence pairs, i.e. minimally different sentences exemplifying a wide array of linguistic contrasts. A well-known one is BLiMP (Benchmark of Linguistic Minimal Pairs) (Warstadt et al., 2020) which contains pairs that contrast in syntactic acceptability and isolating fine-grained phenomena in specific domains of the English grammar, such as subject-verb agreement, island effects, ellipsis and negative polarity items.

Differently from syntactic well-formedness, less explored is the sensitivity of these models to deeper linguistic dimensions involving semantics and discourse, such as textual cohesion, which are critical to language understanding. With this respect, one of the explicit devices that natural languages use to convey textual cohesion is represented by function words. As observed by Kim et al. (2019), although these words play a key role in compositional meaning as they introduce discourse referents or make explicit relations between them, they are still under-investigated in the literature on representation learning. To this end, the authors released a suite of nine challenge tasks for English aimed to test the NLMs’ understanding of specific types of function word, e.g. coordinating conjunctions, quantifiers, definite articles. Reasoning about conjuncts in conjunctive sentences, Saha et al. (2020), instead, introduced CONJNLI, a challenge stress-test for Natural Language Inference (NLI) over conjunctive sentences, where the premise differs from the hypothesis by conjuncts removed, added, or replaced.

Taking inspiration from this work, in this paper we focus the attention on the role of textual connectives in the comprehension of a sentence and we introduce a new evaluation resource for Italian which, to our knowledge, is the first one for this language. The resource is articulated into two sections

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

¹The resource is available at: <http://www.italianlp.it/resources/>.

(acceptability assessment and cloze test), each one corresponding to a distinct task aimed at probing, in a different format, to what extent current NLMs are able to properly encode the role of connectives in a sentence. A peculiarity of the dataset is that it contains sentences that were extracted and minimally modified from existing corpora so as to test the comprehension of connectives in the real use of language.

2 Corpus Collection

This section is divided into two parts. In the first one, we discuss the methodology implemented for the selection of connectives and the extraction of the sentences. Subsequently, we provide an overview of the two tasks defined to test the correct comprehension of connectives.

2.1 Selecting Connectives and Extracting Sentences

As a first step, we defined the linguistic criteria for the selection of connectives to include in the corpus. By *connective* we mean specific words that have the function of drawing a relation between two or more clauses (Sanders and Noordman, 2000; Graesser and McNamara, 2011). To this end, two resources were employed: the INVALSI reading comprehension and language reflection tests designed by the National Institute for the Evaluation of the Education System and the *Nuovo Vocabolario di Base di Italiano* (NVdB) (De Mauro and Chiari, 2016). Starting from the collection of the INVALSI tests proposed in the last six years for different grades, we extracted all words which were expressly called ‘connective’ in the tests or were involved in defining a logical relationship between two sentences. We thus obtained a first list of 46 elements, belonging to diverse morpho-syntactic categories (i.e. prepositions, conjunctions, adverbs), which was then integrated with other 19 connectives extracted from the NVdB. We then checked the distribution of the selected items in existing Italian treebanks and extracted the sentences in which these words were unambiguously used as sentence connectives. Three different sections of the Italian Universal Dependency Treebank (IUDT) (Zeman et al., 2020) were used: ISDT (Bosco et al., 2013), PoSTWITA (Sanguinetti et al., 2018) and TWITTIRÒ (Cignarella et al., 2019)², the first one representative of standard

language and the latter collecting Italian tweets. We employed PML TreeQuery³ to query the treebanks and filter the sentences containing the connectives we were interested in. In particular, to exclude occurrences which do not have the role of phrasal connectives (e.g. the conjunction *e* joining two nouns), only sentences in which the connective was headed by a verb or a copula were taken into account. We observed that the absolute frequency’s positions of the selected connectives in the three corpora above-mentioned mostly overlap, although their occurrences in PoSTWITA and TWITTIRÒ (jointly considered as sample of Italian social media language) were lower than in ISDT, also given the different corpora sizes (i.e. 289,343 words in ISDT vs 154,050 words in PoSTWITA and TWITTIRÒ). Given the partial overlapping of the frequency data and the potential non-standard use of connectives in treebanks representative of social media texts, also due to genre-specific features (e.g. hashtag, emoticons etc.), we decided to consider only the first 21 most frequent connectives occurring in ISDT. As the first Italian corpus for the comprehension on textual connectives, we prefer to focus in sentences as close as possible to standard Italian language. Further considerations on connectives’ distributions led us to the deletion of *per*, *così*, *ancora*, because of their ambiguous behavior as textual connectives (e.g. we noticed that the majority of the occurrences of *per* involves the presence of an infinite verb, a distribution which is far from the other connectives). The following 18 connectives were finally considered: *e*, *se*, *quando*, *come*, *ma*, *dove*, *o*, *anche*, *perché*, *poi*, *mentre*, *infatti*, *prima*, *però*, *invece*, *inoltre*, *tuttavia*, *quindi*. The distribution of the finally selected connectives from ISDT and from PoSTWITA and TWITTIRÒ is reported in Appendix A.

Once established the final list, those sentences which we consider more suitable to be involved in our tasks were manually extracted from ISDT and eventually modified following some patterns, to guarantee sentence comprehension. For example, in some cases two sentences occurring in the treebank in a subsequent order, but that were clearly extracted from the same text, were joined together to form a unique sentence, through the insertion of the appropriate punctuation. This happened e.g. when the connective appeared at the beginning of the second sentence joining this to the first one,

²<https://universaldependencies.org/treebanks/it-comparison.html>.

³<https://ufal.mff.cuni.cz/pmltq>.

which serves as the antecedent to comprehend the logical relationship. We tried to include in the dataset sentences with different degrees of syntactic and lexical complexity, considering the number of subordinate clauses and the variety of the lexicon as related proxies. All the original sentences, later arranged into the acceptability assessment and the cloze test task, are drawn from ISDT.

2.2 Definition of the Tasks

The collected sentences were grouped in two sections aimed at testing the correct comprehension of connectives in a different format, i.e. through an acceptability assessment task and a cloze test task. Table 1 provides an example of sentences/sentences pairs for each task.

2.2.1 Acceptability Assessment Section

To design the acceptability assessment task, we selected 15 sentences per connective from the whole dataset. For each sentence, an unacceptable counterpart was created by replacing the original connective with another of the list. The replacement strategy was meant to obtain unacceptable sentences with contradictory or nonsensical meaning but preserving their grammaticality. Indeed those sentences should be the most challenging one for NLMs, which have been shown to be capable of detecting sentence grammaticality (Jawahar et al., 2019), but still struggle to track down unacceptable meanings and contradictions. Nevertheless, we were not always able to guarantee this constraint as for some specific contexts none of the available connective could be substituted without affecting the resulting grammaticality. This happened in 98 cases, which we decided to keep in the dataset but we signaled with the label ‘no’ in the field ‘grammaticality’, as in:

Nei campi si sopravvive anche intorno tutto muore.

Although the assessment of grammaticality is not the main focus of this work, given the fact that it was unavoidably violated in the above-reported cases, we feel compelled to provide distinguished analysis for the group of ungrammatical sentences. A few sentences were also deleted due to ambiguity. The final section contains 518 sentence pairs, i.e. 259 acceptable and 259 unacceptable ones.

2.2.2 Cloze Test Section

The second section was designed as a cloze test task and contains 270 sentences, 15 for connec-

tive. For every sentence the original connective was replaced by a blank space and 5 alternatives were proposed for completion: the target, a plausible alternative and three implausible options. For ‘plausible alternative’ we mean another connective of the list that could occupy the same linguistic context of the target, yielding to an identical meaning or to a different, yet totally plausible, reading. As for the acceptability task, it turns out that for some connectives (e.g. *prima*) it was very challenging, if not impossible, to propose such a plausible connective. In those cases, that in truth are only a minority, it has been proposed an alternative that at least should guarantee the grammaticality.

3 Corpus Annotation

The two sections of the dataset were splitted into 9 surveys (5 for the acceptability assessment task and 4 for the cloze task) and submitted to human evaluation by recruiting Italian native speakers of different ages through the Prolific platform⁴.

In the **acceptability assessment task**, participants were asked to judge the acceptability of each sentence on a 5-grade Likert scale (from 1=‘totally unacceptable’ to 5=‘totally acceptable’). Although this makes the dataset more challenging, we assume that acceptability is a gradual rather than binary notion as it is affected by many factors (Sorace and Keller, 2005; Sprouse, 2007). To disambiguate the interpretation of sentence acceptability and orient annotators in giving their judgments, the survey guidelines encouraged them to think if they found the sentence natural in Italian and if they would have used it in a real conversation or any other communicative context.

For the **cloze test task**, participants were required to supply the missing element choosing among the proposed options plus the one “none of the previous options is suitable”.

Each survey was completed by 20 annotators on average. The number of annotations per sentence in the acceptability task ranges from 16 to 21 and for the cloze task from 18 to 21. To improve data quality, we discarded annotators who took less than 10 minutes to complete the test, considering the average threshold time for each survey. This led us to reject 5 annotators only for the acceptability task.

Table 2 reports the average human score and standard deviation obtained by the acceptable and

⁴<https://prolific.co>.

Section	Id	Sentence
Acceptability	e_11A	L’arte e la scienza sono libere e libero ne è l’insegnamento.
	e_11NA	L’arte e la scienza sono libere tuttavia libero ne è l’insegnamento.
	ma_64A	Paolo si muove con difficoltà, ma è sempre allegro e di buon umore.
	ma_64NA	Paolo si muove con difficoltà, perché è sempre allegro e di buon umore
Cloze test	se_23cl	Che cosa possiamo fare in estate ... vogliamo partire per le vacanze e abbiamo un cane o un gatto? [se <i>quando</i> perché dove come]
	mentre_162cl	Nelle botteghe artigianali della produzione di piastrelle la smaltatura è ancora tradizionale, ... i forni, come è naturale, oggi funzionano a gas. [mentre <i>invece</i> come dove perché]

Table 1: Examples from the dataset. Sentences are indicated with the last part of id, which gives information about the target connective, the position of the sentence in the section and the label in each section (A=‘acceptable’, NA=‘not acceptable’; cl=‘cloze test’). For the cloze task, the target connective is marked in bold and the plausible alternative in italics.

Acceptability label	AvgIntScore	(StDev)
Acceptable	4.286	0.519
NonAcceptable	1.822	0.451
NonAccept+NonGr	1.616	0.350

Table 2: Average scores assigned by humans (with standard deviation) to the acceptable, unacceptable and unacceptable+ungrammatical sentences.

unacceptable sentences. For the latter, we separately computed these scores for the subset of sentences which were also labeled as ungrammatical (see Section 2.2.1). As it can be seen, humans perform very well on the task assigning quite higher scores to the acceptable sentences with respect to the unacceptable ones, also with little variability. Within the unacceptable subset, the slightly smaller score received on average by ungrammatical sentences provides further evidence that humans are sensitive to this distinction.

Also for the **cloze test task** the human evaluation confirms the validity of the resource. Indeed, as shown in Table 3, the target connective was largely chosen by the majority of annotators as the most adequate one, although for $\sim 20\%$ of sentences humans preferred the plausible candidate or the two options got half annotations each. The percentage of sentences for which the majority label was given to an implausible choice is largely negligible.

4 Testing the Sensitivity of Neural Language Models to Connectives

We conclude by presenting some preliminary findings aimed at testing the performance of NLMs in the two tasks. Specifically, we performed two distinct evaluations. For the **acceptability assessment**

Cloze task choice	N. Items	(%)
Target	213	78.89
Plausible alt.	48	17.78
Implausible alt.	4	1.48
Target=Plausible alt.	5	1.85

Table 3: Number and % of sentences for which the majority label was assigned to the target connective, to the plausible alternative, to an implausible alternative or equally balanced between the target and the plausible alternative.

task, we computed the perplexity (*PPL*) score assigned by the GePpeTto model (De Mattei et al., 2020) to all sentences of the corresponding section. We relied on perplexity as it is a standard evaluation measure of the quality of a language model yielding a good approximation of how well a model recognises an unseen piece of text as a plausible one. Accordingly, we assumed that higher *PPL* scores should be assigned to sentences labeled as unacceptable with respect to their original version. GePpeTto was chosen as it is a traditional unidirectional model built using the GPT-2 architecture (Radford et al., 2019) and, differently from a bidirectional model such as BERT (Devlin et al., 2019), allows computing a well-formed probability distribution over sentences. The sentence-level *PPL* was calculated using the formula reported in Miaschi et al. (2020).

By inspecting the results in Table 4, we observed that the average *PPL* score assigned to the acceptable sentences is quite lower than the one assigned to the unacceptable ones (i.e. 42.512 vs 78.280).

As expected, for the subset of unacceptable sentences, perplexity was on average higher for the ones marked as ungrammatical (98.992), reflecting

AcceptabilityLabel	AvgPPL	minPPL	maxPPL
Acceptable	42.512	2.059	455.961
NonAcceptable	78.280	3.534	390.824
NonAccept+NonGr	98.992	9.933	1178.162

Table 4: Average, minimum and maximum perplexity value given by the model to the acceptable, unacceptable and unacceptable+ungrammatical sentences.

the model’s capability of encoding syntactic phenomena. Interestingly, among unacceptable sentences, those obtaining lower *PPL* scores were perfectly well-formed but with an implausible meaning, as in the case of:

*Il film 'Le chiavi di casa' ha partecipato al Festival del Cinema di Venezia di quest'anno, **perché** non ha vinto nessun premio (PPL = 13.892).*

To compare humans and model performance, we also computed the Spearman’s rank correlation (ρ) between the average acceptability score given by annotators and the *PPL* score assigned by the model to the same sentences. Although limited to this analysis, the resulting very weak correlation (i.e. $\rho = -0.120$, $p - value < 0.01$) suggests that connectives differently impact on the ability of humans and models to assess the plausibility of a sentence.

As for the **cloze task** test, we relied on the pre-trained Italian version of the BERT model developed by the MDZ Digital Library Team and available through the Huggingface’s *Transformers* library (Wolf et al., 2020)⁵. We extracted the first ten completions provided by the model through the Masked Language Modeling task (MLM) for each sentence, along with their probabilities. This allowed us to inspect whether and in how many cases either the target connective or the plausible alternative appear in the top-ranked predictions.

As shown in Table 5, for the large majority of cases BERT is able to infer in its first 10 predictions that the sentence should be completed with a correct connective. That happens in 86.29% of the sentences for the target, resulting from the sum of the cases where only the target occurs in the completions (31.48%) with the cases in which both the target and the plausible alternative were predicted (54.81%), and in 59.25% for the plausible

⁵<https://huggingface.co/dbmdz/bert-base-italian-xxl-cased>

Predict.	10_match	1st_match
Target	(85) 31.48%	(111) 41.11%
Pl. alt.	(12) 4.44%	(23) 8.52%
Target+Pl. alt.	(148) 54.81%	–
Other	(25) 9.26%	–

Table 5: (Number) and % of BERT’s completions in which only the target, only the plausible alternative, both of them or none of them (*Other*) occur in the first 10 predictions (*10_match*). (Number) and % of the completions in which the target and the plausible alternative were predicted with the highest probability are also reported (*1st_match*).

alternative (that is 4.44% plus 54.81%). Focusing instead on the first completion for each sentence, we observe that in almost half of the sentences BERT assigns the highest probability to the original connective (41.11%) or to the plausible one (8.52%).

We are currently performing a more qualitative analysis to better investigate the cases in which the correct connective hasn’t received a high probability score, as well as those in which neither of the two options appeared at all (i.e. *Other* cases in Table 5), in order to understand whether the other completions can still be considered as plausible ones. Preliminary findings showed that, among the *Other* cases, about 56 of the completions provided by BERT are unacceptable and 34 of them are dubious acceptable i.e. not clearly recognizable as acceptable⁶, as in the case of the following sentence⁷:

*Secondo gli esperti, in Italia i giovani leggono meno i giornali rispetto ai giovani di altri Paesi europei, ... rispetto agli anni passati i giovani tra i 14 e i 19 anni leggono più spesso i giornali. [**perché** anche però].*

Nevertheless, the majority of *Other*’s completions can be considered as acceptable ones. In fact, BERT predicted a word leading to the same meaning (or, at least, very similar) to the original sentence in more than 60 cases. Moreover, in most cases (i.e. 92) the completions provided are plausible ones, although in some of them the sentences acquire different meanings.

⁶Note that in order to assign the acceptability label of each completion we refer to the usage of the Italian language as standard as possible.

⁷the unacceptable completion is marked in bold, the dubious acceptable one is reported in block and the original connective is indicated in italics.

5 Conclusion

In the context of studies devoted to assess the linguistic knowledge implicitly encoded by Neural Language Models, we introduced a new evaluation dataset for Italian designed to test the understanding of textual connectives in real-usage sentences. At first, we verified the significance of a set of selected connectives through a frequency analysis on already existing Italian gold corpora. Then, we manually selected only those sentences in which occur a genuine connective. Finally, we grouped the sentences into two different tasks, differing for the format used to elicit sentence comprehension in humans and current state-of-the-art NLMs: acceptability assessment and cloze test tasks. Human evaluation was provided for both the section, to verify the robustness of the dataset, which indeed was confirmed from the judgements collected.

Preliminary findings on NLMs behaviour on textual connectives showed that in several cases the models are capable of distinguishing between acceptable and unacceptable sentences, thus suggesting their ability to encode sentence meaning within their internal mechanisms. However, it remains unclear to what extent these models rely on semantic acceptability features, since we observed cases in which they fail to recognize implausible meaning of perfectly grammatical sentences.

We are currently increasing the dataset with the introduction of a new section designed in the form of the traditional Natural Language Inference task, for which the understanding of a given connective will be fundamental to infer the correct entailment relation between a premise and a hypothesis. We also believe that expanding the dataset to further connectives and including sentences representative of non standard Italian language usage, i.e. social-media language, would be desirable to improve the robustness of the resource.

References

Cristina Bosco, Simonetta Montemagni, and Maria Simi. 2013. Converting Italian treebanks: Towards an Italian Stanford dependency treebank. In *Proceedings of the ACL Linguistic Annotation Workshop & Interoperability with Discourse*.

Alessandra Teresa Cignarella, Cristina Bosco, and Paolo Rosso. 2019. Presenting TWITTIRÒ-UD: An Italian Twitter treebank in universal dependencies. In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*.

Lorenzo De Mattei, Michele Cafagna, Felice Dell’Orletta, Malvina Nissim, and Marco Guerini. 2020. Geppetto carves Italian into a language model. In *CLiC-it*.

Tullio De Mauro and I Chiari. 2016. Il nuovo vocabolario di base della lingua italiana. *Internazionale*. [28/11/2020]. <https://www.internazionale.it/opinione/tullio-de-mauro/2016/12/23/il-nuovo-vocabolario-di-base-della-lingua-italiana>.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.

Arthur C Graesser and Danielle S McNamara. 2011. Computational analyses of multilevel discourse comprehension. *Topics in cognitive science*, 3(2):371–398.

Ganesh Jawahar, Benoît Sagot, and Djamel Seddah. 2019. What does BERT learn about the structure of language? In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3651–3657, Florence, Italy, July. Association for Computational Linguistics.

Najoung Kim, Roma Patel, Adam Poliak, Alex Wang, Patrick Xia, R. Thomas McCoy, Ian Tenney, Alexis Ross, Tal Linzen, Benjamin Van Durme, Samuel R. Bowman, and Ellie Pavlick. 2019. Probing what different NLP tasks teach machines about function word comprehension. In **SEMVAL*.

Tal Linzen and Marco Baroni. 2021. Syntactic structure from deep learning. *Annual Review of Linguistics*, 7:195–212.

Christopher D. Manning. 2015. Computational Linguistics and Deep Learning. *Computational Linguistics*, 41(4):701–707, 12.

Alessio Miaschi, Chiara Alzetta, Dominique Brunato, Felice Dell’Orletta, and Giulia Venturi. 2020. Is neural language model perplexity related to readability? In *CLiC-it*.

Alec Radford, Jeff Wu, R. Child, David Luan, Dario Amodei, and Ilya Sutskever. 2019. Language models are unsupervised multitask learners.

Swarnadeep Saha, Yixin Nie, and Mohit Bansal. 2020. ConjNLI: Natural language inference over conjunctive sentences. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8240–8252, Online, November. Association for Computational Linguistics.

Ted JM Sanders and Leo GM Noordman. 2000. The role of coherence relations and their linguistic markers in text processing. *Discourse processes*, 29(1):37–60.

- Manuela Sanguinetti, Cristina Bosco, Alberto Lavelli, Alessandro Mazzei, and Fabio Tamburini. 2018. PoSTWITA-UD: an Italian Twitter Treebank in universal dependencies. In *Proceedings of the Eleventh Language Resources and Evaluation Conference (LREC 2018)*.
- Antonella Sorace and Frank Keller. 2005. Gradience in linguistic data. *Lingua*, 115(11):1497–1524.
- Jon Sprouse. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics*, pages 1123–134.
- Alex Warstadt, Alicia Parrish, Haokun Liu, Anhad Mohananey, Wei Peng, Sheng-Fu Wang, and Samuel R. Bowman. 2020. BLiMP: The benchmark of linguistic minimal pairs for English. *Transactions of the Association for Computational Linguistics*, 8:377–392.
- Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. 2020. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, October. Association for Computational Linguistics.
- Daniel Zeman, Joakim Nivre, Mitchell Abrams, Elia Ackermann, Noëmi Aepli, Hamid Aghaei, Željko Agić, Amir Ahmadi, Lars Ahrenberg, Chika Kennedy Ajede, Gabrielė Aleksandravičiūtė, Ika Alfina, Lene Antonsen, Katya Aplonova, Angelina Aquino, Carolina Aragon, Maria Jesus Aranzabe, Hórunn Arnardóttir, Gashaw Arutie, Jessica Naraiswari Arwidarasti, Masayuki Asahara, Luma Ateyah, Furkan Atmaca, Mohammed Attia, Aitziber Atutxa, Liesbeth Augustinus, Elena Badmaeva, Keerthana Balasubramani, Miguel Ballesteros, Esha Banerjee, Sebastian Bank, Verginica Barbu Mititelu, Victoria Basmov, Colin Batchelor, John Bauer, Seyyit Talha Bedir, Kepa Bengoetxea, Gözde Berk, Yevgeni Berzak, Irshad Ahmad Bhat, Riyaz Ahmad Bhat, Erica Biagetti, Eckhard Bick, Agnė Bielinskienė, Kristín Bjarnadóttir, Rogier Blokland, Victoria Bobicev, Loïc Boizou, Emanuel Borges Völker, Carl Börstell, Cristina Bosco, Gosse Bouma, Sam Bowman, Adriane Boyd, Kristina Brokaitė, Aljoscha Burchardt, Marie Candito, Bernard Caron, Gauthier Caron, Tatiana Cavalcanti, Gülşen Cebiroğlu Eryiğit, Flavio Massimiliano Cecchini, Giuseppe G. A. Celano, Slavomír Čěplö, Savas Cetin, Özlem Çetinöğlü, Fabricio Chalub, Ethan Chi, Yongseok Cho, Jinho Choi, Jayeol Chun, Alessandra T. Cignarella, Silvie Cinková, Aurélie Collomb, Çağrı Çöltekin, Miriam Connor, Marine Courtin, Elizabeth Davidson, Marie-Catherine de Marneffe, Valeria de Paiva, Mehmet Oguz Derin, Elvis de Souza, Arantza Diaz de Ilarraza, Carly Dickerson, Arawinda Dinakaramani, Bamba Dione, Peter Dirix, Kaja Dobrovolska, Timothy Dozat, Kira Droganova, Puneet Dwivedi, Hanne Eckhoff, Marhaba Eli, Ali Elkahky, Binyam Ephrem, Olga Erina, Tomáš Erjavec, Aline Etienne, Wograine Evelyn, Sidney Facundes, Richárd Farkas, Marília Fernanda, Hector Fernandez Alcalde, Jennifer Foster, Cláudia Freitas, Kazunori Fujita, Katarína Gajdošová, Daniel Galbraith, Marcos Garcia, Moa Gärdenfors, Sebastian Garza, Fabrício Ferraz Gerardi, Kim Gerdes, Filip Ginter, Iakes Goenaga, Koldo Gojenola, Memduh Gökırmak, Yoav Goldberg, Xavier Gómez Guinovart, Berta González Saavedra, Bernadeta Griciūtė, Matias Groni, Loïc Grobol, Normunds Grūzītis, Bruno Guillaume, Céline Guillot-Barbance, Tunga Güngör, Nizar Habash, Hinrik Hafsteinsson, Jan Hajič, Jan Hajič jr., Mika Hämäläinen, Linh Hà Mỹ, Na-Rae Han, Muhammad Yudistira Hanifmuti, Sam Hardwick, Kim Harris, Dag Haug, Johannes Heinicke, Oliver Hellwig, Felix Hennig, Barbora Hladká, Jaroslava Hlaváčová, Florinel Hociung, Petter Hohle, Eva Huber, Jena Hwang, Takumi Ikeda, Anton Karl Ingason, Radu Ion, Elena Irimia, Olájidé Ishola, Tomáš Jelínek, Anders Johannsen, Hildur Jónsdóttir, Fredrik Jørgensen, Markus Juutinen, Sarveswaran K. Hüner Kaşıkara, Andre Kaasen, Nadezhda Kabaeva, Sylvain Kahane, Hiroshi Kanayama, Jenna Kanerva, Boris Katz, Tolga Kayadelen, Jessica Kenney, Václava Kettnerová, Jesse Kirchner, Elena Klementieva, Arne Köhn, Abdullatif Köksal, Kamil Kopacewicz, Timo Korkiakangas, Natalia Kotsyba, Jolanta Kovalevskaitė, Simon Krek, Parameswari Krishnamurthy, Sookyoung Kwak, Veronika Laippala, Lucia Lam, Lorenzo Lambertino, Tatiana Lando, Septina Dian Larasati, Alexei Lavrentiev, John Lee, Phng Lê H'ông, Alessandro Lenci, Saran Lertpradit, Herman Leung, Maria Levina, Cheuk Ying Li, Josie Li, Keying Li, Yuan Li, KyungTae Lim, Krister Lindén, Nikola Ljubešić, Olga Loginova, Andry Luthfi, Mikko Luukko, Olga Lyashevskaya, Teresa Lynn, Vivien Macketanz, Aibek Makazhanov, Michael Mandl, Christopher Manning, Ruli Manurung, Cătălina Mărănduc, David Mareček, Katrin Marheinecke, Héctor Martínez Alonso, André Martins, Jan Mašek, Hiroshi Matsuda, Yuji Matsumoto, Ryan McDonald, Sarah Mc Guinness, Gustavo Mendonça, Niko Miekka, Karina Mischenkova, Margarita Misirpashayeva, Anna Missilä, Cătălin Mititelu, Maria Mitrofan, Yusuke Miyao, AmirHossein Mojiri Foroushani, Amirsaeid Moloodi, Simonetta Montemagni, Amir More, Laura Moreno Romero, Keiko Sophie Mori, Shinsuke Mori, Tomohiko Morioka, Shigeki Moro, Bjartur Mortensen, Bohdan Moskalevskyi, Kadri Muischnek, Robert Munro, Yugo Murawaki, Kaili Müürisep, Pinkey Nainwani, Mariam Nakhlé, Juan Ignacio Navarro Horñiáček, Anna Nedoluzhko, Gunta Nešpore-Bērzkalne, Lng Nguy`ên Thị, Huy`ên Nguy`ên Thị Minh, Yoshihiro Nikaido, Vitaly Nikolaev, Rattima Nitisaroj, Alireza Nourian, Hanna Nurmi, Stina Ojala, Atul Kr. Ojha, Adédayo Olúòkun, Mai Omura, Emeka Onwuegbuzia, Petya Osenova, Robert Östling, Lilja Øvre-

lid, Şaziye Betül Özateş, Arzucan Özgür, Balkız Öztürk Başaran, Niko Partanen, Elena Pascual, Marco Passarotti, Agnieszka Patejuk, Guilherme Paulino-Passos, Angelika Peljak-Łapińska, Siyao Peng, Cenel-Augusto Perez, Natalia Perkova, Guy Perrier, Slav Petrov, Daria Petrova, Jason Phelan, Jussi Piitulainen, Tommi A Pirinen, Emily Pitler, Barbara Plank, Thierry Poibeau, Larisa Ponomareva, Martin Popel, Lauma Pretkalniņa, Sophie Prévost, Prokopis Prokopidis, Adam Przepiórkowski, Tiina Puolakainen, Sampo Pyysalo, Peng Qi, Andriela Rääbis, Alexandre Rademaker, Taraka Rama, Loganathan Ramasamy, Carlos Ramisch, Fam Rashel, Mohammad Sadegh Rasooli, Vinit Ravishankar, Livy Real, Petru Rebeja, Siva Reddy, Georg Rehm, Ivan Riabov, Michael Riebler, Erika Rimkutė, Larissa Rinaldi, Laura Rituma, Luisa Rocha, Eiríkur Rögnvaldsson, Mykhailo Romanenko, Rudolf Rosa, Valentin Roşca, Davide Rovati, Olga Rudina, Jack Rueter, Kristján Rúnarsson, Shoal Sadde, Pegah Safari, Benoît Sagot, Aleksí Sahala, Shadi Saleh, Alessio Salomoni, Tanja Samardžić, Stephanie Samson, Manuela Sanguinetti, Dage Särg, Baiba Saulīte, Yanin Sawanakunanon, Kevin Scannell, Salvatore Scarlata, Nathan Schneider, Sebastian Schuster, Djámé Seddah, Wolfgang Seeker, Mojgan Seraji, Mo Shen, Atsuko Shimada, Hiroyuki Shirasu, Muh Shohibussirri, Dmitry Sichinava, Einar Freyr Sigursson, Aline Silveira, Natalia Silveira, Maria Simi, Radu Simionescu, Katalin Simkó, Mária Šimková, Kiril Simov, Maria Skachedubova, Aaron Smith, Isabela Soares-Bastos, Carolyn Spadine, Steinhór Steingrímsson, Antonio Stella, Milan Straka, Emmett Strickland, Jana Strnadová, Alane Suhr, Yogi Lesmana Sulestio, Umut Sulubacak, Shingo Suzuki, Zsolt Szántó, Dima Taji, Yuta Takahashi, Fabio Tamburini, Mary Ann C. Tan, Takaaki Tanaka, Samson Tella, Isabelle Tellier, Guillaume Thomas, Lisi Torga, Marsida Toska, Trond Trosterud, Anna Trukhina, Reut Tsarfaty, Utku Türk, Francis Tyers, Sumire Uematsu, Roman Untilov, Zdeňka Urešová, Larraitz Uria, Hans Uszkoreit, Andrius Utka, Sowmya Vajjala, Daniel van Niekerk, Gertjan van Noord, Viktor Varga, Eric Villemonte de la Clergerie, Veronika Vincze, Aya Wakasa, Joel C. Wallenberg, Lars Wallin, Abigail Walsh, Jing Xian Wang, Jonathan North Washington, Maximilian Wendt, Paul Widmer, Seyi Williams, Mats Wirén, Christian Wittern, Tsegay Woldemariam, Tak-sum Wong, Alina Wróblewska, Mary Yako, Kayo Yamashita, Naoki Yamazaki, Chunxiao Yan, Koichi Yasuoka, Marat M. Yavrumyan, Zhuoran Yu, Zdeněk Žabokrtský, Shorouq Zahra, Amir Zeldes, Hanzhi Zhu, and Anna Zhuravleva. 2020. Universal dependencies 2.7. LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University.

A Appendix A

Conn.	ISDT	PoSTWITA+TWITTIRò
e	(1,906) 0.639%	(909) 0.590%
se	(575) 0.193%	(477) 0.309%
quando	(529) 0.177%	(141) 0.092%
come	(422) 0.141%	(226) 0.147%
ma	(312) 0.105%	(713) 0.463%
dove	(306) 0.103%	(60) 0.039%
o	(259) 0.087%	(89) 0.058%
anche	(253) 0.085%	(123) 0.080%
perché	(231) 0.077%	(255) 0.166%
poi	(138) 0.046%	(46) 0.030%
mentre	(126) 0.042%	(24) 0.016%
infatti	(109) 0.037%	(13) 0.008%
prima	(106) 0.036%	(49) 0.032%
però	(101) 0.034%	(46) 0.030%
invece	(98) 0.033%	(49) 0.032%
inoltre	(88) 0.029%	(1) 0.0006%
tuttavia	(80) 0.027%	(1) 0.0006%
quindi	(78) 0.026%	(28) 0.018%

Table 6: (Numbers) and % of the frequency of the 18 finally selected connectives in ISDT corpus and in PoSTWITA+TWITTIRò corpora.