OntoVerbal-M: a Multilingual Verbaliser for SNOMED CT

Shao Fen Liang, Robert Stevens and Alan Rector

School of Computer Science, The University of Manchester, Oxford Road, Manchester, UK M13 9PL
{Fennie.Liang, Robert.Stevens, Rector}@cs.man.ac.uk

Abstract. OntoVerbal-M is an ontology verbaliser that transforms OWL into fluent natural language paragraphs in multiple languages. We describe the application of OntoVerbal-M to SNOMED CT, whereby SNOMED CT classes are presented as textual paragraphs in both English and Mandarin through the use of natural language generation. SNOMED CT is a large description logic based terminology for recording in electronic health records. Often, neither the labels nor the description logic definitions in SNOMED CT are easy for users to understand. Furthermore, information is increasingly being recorded, not just using individual SNOMED CT concepts, but using dynamically created description logic expressions ("post-coordinated" concepts). Such post-coordinated expressions can have no pre-assigned labels. In this context automatic verbalisation into multiple languages will be useful both for understanding and quality assurance of SNOMED CT definitions, and for helping different language-speaking-users to understand and share post-coordinated expressions.

Keywords: Multilingual Generation, Ontology Verbalisation, Ontology verbaliser, SNOMED verbalisation.

1. Introduction

We present OntoVerbal-M, a multi-lingual verbaliser for ontologies tailored to be used with SNOMED CT, a large medical terminology. Such ontologies and terminologies are increasingly authored in description logics, such as the W3C recommendation, the Web Ontology Language, OWL [2]. Expressions in Description Logics and OWL are often difficult for domain experts to understand [17]. Even using the human readable Manchester Syntax [10], expressions can have multiple levels of nesting and many inter-related axioms.

Verbalising these expressions in natural language is therefore attractive as a means to communicate with users [3; 4; 6]. Verbalisation has the added advantage that it should be possible to re-use some of the same language generation components in the generation of verbalisations in multiple languages.

SNOMED CT [19; 21] (Systematized Nomenclature of Medicine Clinical Terms) is big and potentially widely used OWL based terminology in any field. It attempts to provide a comprehensive terminology for use in medical records across all of medicine, including diseases, diagnoses, procedures, anatomy, microorganisms and pharmaceuticals. It is maintained by the International Health Terminology Standards Development Organisation (IHTSDO)[1], and has been mandated or advocated for use in more than 50 countries. Today SNOMED CT is available in US English, UK English and Spanish. Translations to several other languages are currently taking place.

We have taken SNOMED CT as an example to demonstrate our techniques for verbalisation. In its OWL form, SNOMED CT is often awkward and even obscure. For example, the rendering of even just the definition of a simple concept such as *heart disease* in the raw OWL version of SNOMED CT is several lines long:

Class: Heart disease
EquivalentTo: Disorder of cardiovascular system
and RoleGroup *some* (Finding site *some* Heart structure)

By contrast, an English "verbalisation" of this definition in natural language as shown below will be easier for domain experts to understand, although it still seems somewhat stilted:

A heart disease is a disorder of the cardiovascular system that is found in the structure of the heart.

The verbalisation also omits the technically necessary, but to the domain expert mysterious, expression "RoleGroup".

When we attempt to present, not just the definition, but the information present in the ontology about a concept – *e.g.* Heart disease – the OWL expressions become more complex. Worse, they may not all be located together in the ontology. Hence the advantage of a verbaliser that presents the entire description of a concept in a single natural language paragraph, according to the discourse rules expressed in Rhetorical Structure Theory [14].

Using Rhetorical Structure Theory, furthermore, gives us a major component that appears to be re-usable across languages. The same mechanisms that produced the English above can generate Mandarin as:

心臟 病 是 由於 心臟 結構 異常 導致的 心血管
heart disease is from heart structure disorder caused cardiovascular
系統 失調
system disorder

Such verbalisations could be produced manually, but this is time consuming and, as mentioned, not possible for the dynamically created "post-coordinated" expressions for concepts.

OntoVerbal-M provides natural text descriptions with the aim of helping non-ontology experts understand the concepts in SNOMED CT. Currently, we have produced an English version using the official SNOMED CT labels and an experimental Mandarin version using ad hoc translations by a native speaker. The Mandarin must be taken with caution, as the translations of the individual labels are ad hoc and the validation has so far been only opportunistic. Nonetheless, the results have been sufficiently well received that we are strongly encouraged to extend the

[1] http://www.ihtsdo.com

study to a more formal analysis. In future, we hope to extend this to other languages and to compare verbalisations from OntoVerbal-M with manual translations.

It must be emphasised that OntoVerbal-M is not a machine translation system from one string to another. Rather it generates texts in multiple languages from the same underlying conceptual structure – ultimately a set of expressions in a description logic and the lexicon associated with those concepts in a particular language, as in other multilingual Natural Language Generation Systems [13; 18].

2. The OntoVerbal-M system

OntoVerbal-M is an extension of OntoVerbal. OntoVerbal was initially built for verbalising ontologies into English text [11], and has motivated us to test its top level rhetoric structure schema as a multilingual generator. Although there is no official SNOMED CT mandarin labels, we have tried our best using a mandarin native speaker's medical knowledge and consulting with an English SNOMED CT expert to produce mandarin labels as a test bed.

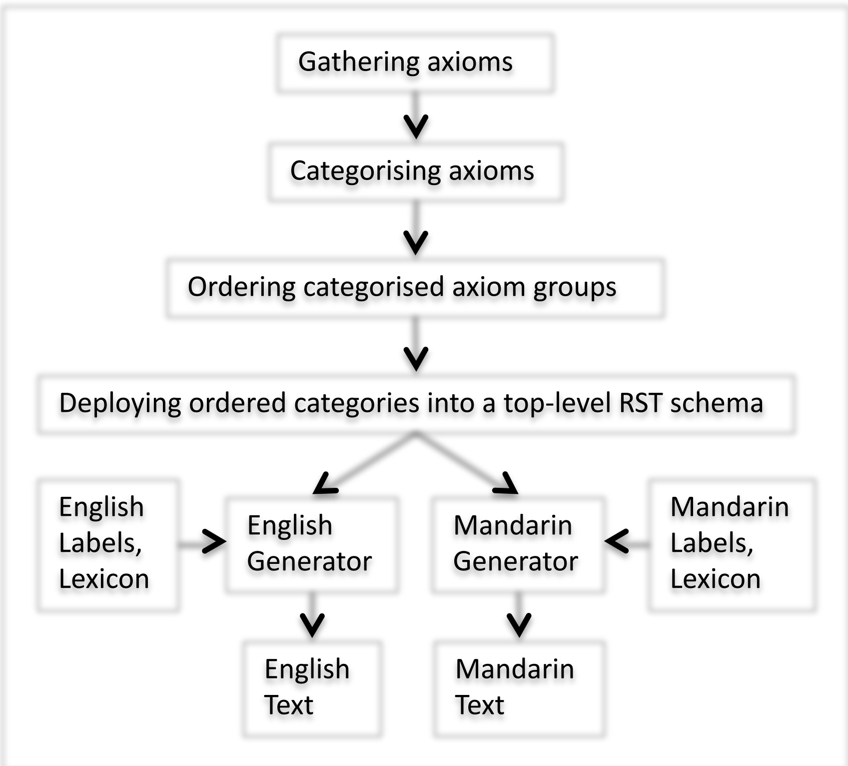

Fig. 1 The system architecture of OntoVerbal-M

OntoVerbal-M utilises the intuitive correlation between axioms and sentences to produce paragraphs that are more than simple collections of individual sentences. Instead, the sentences are structured and ordered [12]. This is achieved through five main operations: (a) gathering axioms together based on a shared focus class; (b) categorising the gathered axioms into different groups; (c) ordering categorised axiom groups; (d) deploying the ordered categories into a top-level discourse structure using Rhetorical Structure Theory (RST) [14]; and (e) using language generators to make the text hang together in a meaningful and organised manner.

OntoVerbal-M currently has two language generators as shown in Fig. 1, but they share a single discourse structure. Each generator has its own input labels as well as its own lexicons. Table 1 shows some examples of class labels in both languages.

Table 1 Example of class labels in both languages

SCT ID	English	Mandarin
302215000	thrombocytopenic disorder	血小板減少失調
107671003	vascular sclerosis	血管硬化
206596003	neonatal hypertension	新生兒的高血壓
10725009	benign hypertension	良性高血壓
113331007	structure of endocrine system	內分泌系統結構

Axioms in different notions are also transformed into sentences respectively according to the role of the focused class in the axiom as shown in Table 2. So, for example, a focus class “X” is to be expressed as a sub class of Y in an axiom, then this axiom is to be transformed into English as an X is a kind of Y, and in Mandarin as X屬於Y.

Table 2 Axiom transformation templates

Axiom notion	English template	Mandarin template
X sub class of …	An X is a kind of ..	X 屬於 …
X super class of …	A more specialised kind of X is..	X 包含了…

2.1. Applying natural language techniques

There are several natural language (NL) techniques that have been embedded in OntoVerbal-M. The first one is aggregation [8; 16]. For example “an X is a kind of an O”, “an X is a kind of a P” and “an X is a kind of a Q”. The three sentences are aggregated as “an X is a kind of an O, a P and a Q”. The same technique is also applied to Mandarin to have the sentence as “X屬於O 、P和Q”.

The second NL technique used is topic-maintenance-device [15]; This is used to avoid introducing a disfluency through the sudden shift of topic from one to another [22], and thus placing an additional cognitive load on the reader [7]. In general, axioms are expressed in one direction – from-child-to-parent – such as X sub class of Y, Y sub class of Z. However, there is often the chance that the focus class is in a parent position in an axiom. Therefore, in order to keep a topic consistent in a generated text, instead of saying “an X is a kind of Y”, we need to say “a more specialised kind of Y is X” in English and “Y包含了X” in Mandarin to maintain a consistent topic for Y.

The third NL technique is the use of discourse markers [5; 20]. Discourse markers are applied when a focus class contains several axioms to be verbalised. Using discourse markers ensures the maintenance of fluency and coherence in a paragraph. For example, to connect an additional sentence from the above example, we use “additionally” in English and “而且” in Mandarin to produce the following paragraphs: “an X is a Y. A more specialised kind of X is Z. Additionally, an X is defined as a P that …”, and “X屬於Y 。它也包含了Z。而且X被定義為 P…中…”.

The fourth NL technique uses a set of key phrases to signal a change of topic in the generated text. Without such signalling, the text will lack coherence and fluency and be harder to understand. In cases where extra information should be given to the focus class, we introduce “Another relevant aspect of” or “Other relevant aspects of ” as key phrases in English and “其他與...相關的資訊” in mandarin to signal the topic change.

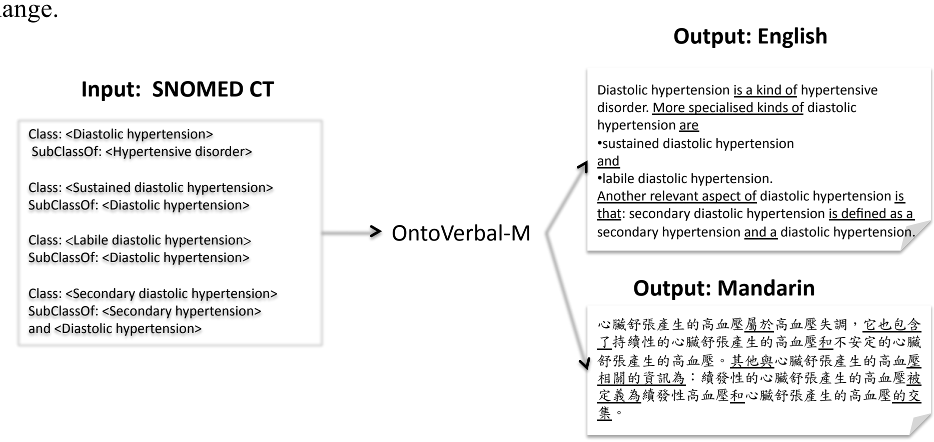

Fig. 2 Input and Output of OntoVerBal-M

Fig. 2 shows an example of an actual SNOMED CT concept input, and its English and Mandarin outputs in natural language. The non-underlined words are SNOMED CT labels, and the underlined words are system-selected words for text fluency purposes.

2.2. Results

Our primary goal is to provide text that not only has the structure of SNOMED CT concepts, but also to have them made clear. The textual output of OntoVerbal-M is thus faithful to the ontological input; a more idiomatic verbalisation is not our current goal. We show here some typical output in two languages; the input is the module extracted[2] using as a signature *Hypertension* and all of its inferred subclasses from SNOMED CT full. The underlined words are system-generated words, and the non-underlined words are formal SNOMED CT terms. The following are five outputs from OntoVerbal-M. Each of them can be a verbalisation from one simple axiom to several complex axioms.

(a) Goldblatt hypertension is a kind of renovascular hypertension.
腎血管阻塞性高血壓屬於腎血管性高血壓。

(b) Disorder of the pelvis is defined as a disorder of the trunk that has a finding site in the pelvis.
骨盆失調被定義為軀幹失調中在骨盆的結構上有病灶。

(c) The cell is a kind of anatomical structure. More specialised kinds of the cell are

- entire cell

and

- subcellular structure.

細胞結構屬於解剖的結構，它也包含了全部的細胞和亞細胞組成的結構。

(d) Disorder of pregnancy is defined as a finding related to pregnancy and a disease. Another relevant aspect of a disorder of pregnancy is that: complication related to pregnancy is defined as a complication and a disorder of pregnancy.
妊娠失調被定義為妊娠相關的發現和疾病的交集。其他與妊娠失調相關的資訊為：
妊娠相關的併發症被定義為併發症和妊娠失調的交集。

(e) Kidney disease is defined as a disorder of the genitourinary system that has a finding site in a kidney.
Other relevant aspects of a kidney disease include the following:

- renal impairment is a kind of kidney disease that has a finding site in a kidney;
- uremia is a kind of metabolic disease that is a kidney disease, and has a finding site in a kidney;
- renal vascular disorder is defined as a vascular disease of the abdomen that is a kidney disease, and has a finding site in a vessel of kidney;
- toxic nephropathy is defined as a kidney disease that has a causative agent in a substance;
- renal hypertension is defined as a secondary hypertension that is associated with a kidney disease;
- hypertensive renal disease is defined as a hypertensive disorder that is a kidney disease, and has a finding site in a kidney.

腎臟病被定義為生殖泌尿系統失調中在腎臟結構上有病灶。其他與腎臟病相關的資訊為：

一、腎臟損傷是一種腎臟病中在腎臟結構上有病灶。
二、尿毒症是一種新陳代謝疾病中的腎臟病和在腎臟結構上有病灶。
三、腎血管失調被定義為腹部血管的疾病中的腎臟病和在腎臟的血管結構上有病灶。
四、腎毒症被定義為腎臟病中在物質上有導致的藥物。
五、腎臟高血壓被定義為續發性高血壓中在腎臟病上有關聯。
六、腎性高血壓疾病被定義為高血壓失調中的腎臟病和在腎臟結構上有病灶。

[2] http://owl.cs.manchester.ac.uk/snomed/

3. SNOMED CT Challenges for English and Mandarin Text Generation

Most SNOMED CT class IDs have several associated terms [1], such as a "preferred term" (that is expected to be used most commonly in medical records and interfaces), and a "fully specified term" (that is intended to be completely unique and self explanatory). We have tried different SNOMED CT labels with OntoVerbal-M. So we get, for example, "Disorder of pelvic region is defined as a disorder of trunk that has a finding site in pelvic structure" from fully specified terms. We also get "Disorder of pelvis is defined as a disorder of trunk that has a finding site in pelvis" from SNOMED CT preferred terms. Neither of them has articles in the sentences as we have in the result section 2.2 (b).

3.1. The difference in using articles and plurality

The use of SNOMED CT supplied terms directly causes articles to be missed in the generated text, especially when definite articles are needed in the text. For example, if we used only the SNOMED CT supplied terms, we would have just "pelvis" and "trunk" in the output rather than "the pelvis" and "the trunk". The ontology alone does not provide sufficient information to deal with articles and plurals completely. For example, the use of singular or plurals in anatomy depends on whether the body normally has just one or more than one of a particular kind of part. The naming convention is, however, to take an unadorned singular form, such as "heart", rather than "the heart". Therefore, instead of naming a class "the heart structure", SNOMED CT actually names this class "heart structure". Our approach on dealing with the definite article and plurals is to look up online resources that can provide examples of usage of articles in human anatomical terms, so that we could replace the SNOMED CT labels with English expressions; Table 3 shows some examples from this approach.

Table 3 Relabelling anatomical terms

Original SNOMED CT label	New label
Procedure on thorax	Procedure on the thorax
Procedure on mediastinum	Procedure on the mediastinum
Procedure on abdomen	Procedure on the abdomen
Procedure on pelvis	Procedure on the pelvis
Procedure on heart	Procedure on the heart
Disorder of soft tissue of thoracic cavity	Disorder of soft tissue of the thoracic cavity
Branch of abdominal aorta	Branch of the abdominal aorta
Finding of cellular component of blood	Finding of cellular component of the blood

This approach is, however, not perfect, as some terms in SNOMED CT are either missing or in different phrasing order from our looked-up resources. For example, we are able to change "pelvis" to "the pelvis" but unable to change "pelvic structure" to "the pelvic structure". A further process, such as the use of the Unified Medical

Language System (UMLS)[3], would be needed to improve this problem to change the adjectival form for this case.

In Mandarin, articles and plurality are not as a concern, as they are in generating English text. The difference between these two languages is that English has count nouns in a singular or plural form, or even to have mass nouns, while Mandarin can only modify a noun by adding an adjective or numeral in front of the noun. For example: "你的(your)心臟(heart)", "我的(my)心臟(heart)", "一個(one)心臟(heart)" or "二個(two)心臟(heart)". Wherever the "心臟" appears in a text, it never becomes "the心臟" or "心臟s". This means the fixes applied for articles in English NLG of SNOMED CT are not needed in the Mandarin version.

3.2. The difference in generality

Logical conjunction is one of the important methods used in designing SNOMED CT terms. It reveals the semantic relationship between the clinical medical concepts and their logical triple structure in order to present clinical information. For example:

Class: Acute metabolic disorder
EquivalentTo: Acute disease *and* Metabolic disease

The *Acute metabolic disorder* is an intersection between *Acute disease* and *Metabolic disease*. This axiom is transformed into English as "Acute metabolic disorder is defined as both an acute disease and a metabolic disease", where the intersection in this axiom is not shown directly in the English verbalisation. This verbalisation can be ambiguous for non-native English speakers in understanding that *Acute metabolic disorder = Acute disease ∪ Metabolic disease* rather than *Acute metabolic disorder = Acute disease ∩ Metabolic disease*.

Why not just simply transformed the above axiom as "Acute metabolic disorder is defined as an intersection between an acute disease and a metabolic disease"? The reason is that the word "intersection" is a mathematical word, which is not commonly used outside Mathematics. Also, translating "intersection" unambiguously into English is awkward at best.

In comparison, the same axiom is transformed into Mandarin as "急性(Acute)新陳代謝(metabolic)失調(disorder)是(is)急性(acute)疾病(disease)和(and)新陳代謝(metabolic)疾病(disease)的(apostrophe)交集(intersection)。", where 交集(intersection) is the word – intersection, and is just simply to be used to express its role. In fact, the phrase 交集 is not awkward in Mandarin's daily conversation. For example, if A decides to break a relationship with B. A can say to B "your life has no 交集(intersection) with me", or A complains about B and says "my conversation with B has no 交集(intersection)".

[3] http://www.nlm.nih.gov/research/umls/

3.3. The different role of properties in the translation

Properties are one of the problems in ontology verbalisation due to the lack of NLG orientated guidelines for labelling properties [9]. Morphological features of the first word in a property have an impact on producing fluent text automatically. For example, the first word of a property could be a noun, an adjective, a verb or a preposition as shown in Table 4.

Table 4 Morphological features of SNOMED CT properties

Morphological feature of the first word	Frequency	Example
Noun	33	Procedure site – direct
Adjective	12	Clinical course
Verb in its present tense and 3rd person singular form	8	Has focus
Verb in its present participle	4	Using device
Verb in its past participle	4	Associated with
Preposition	1	After

With each different initial morphology, the verbaliser needs a Part Of Speech (POS) checking in order to choose a correct verb for generating a sentence such as using "is" or "has" or without adding verbs. Our experience has suggested that a standardised term modelling approach will represent ontologies well, and will also save much time when building ontology verbalisers. The quality of the verbalisation will improve dramatically if ontology classes and properties are well phrased or annotated such that the linguistic behaviour of the concept or property is apparent. For example: class labels are noun phrases; property labels start with a third personal singular verb and end with a preposition. In this case, a property would act as a nice predicate between its subject and object classes. This way would free a verbaliser from concerning itself with or without a verb and the text fluency while transforming properties without appropriate prepositions.

In SNOMED CT the property "after" is a good example. If it were lexicalised as "has an after effect in" then the *postoperative complication* concept can be verbalised as "postoperative complication is defined as a complication of a surgical procedure that has an after affect in a surgical procedure". In this case the verbaliser only needs to concern itself about the article in this sentence.

The issues on phrasing class and property labels in verbalising English SNOMED CT has led us to be more careful in translating Mandarin labels. We adapt the standardised term modelling approach suggested from English labels to translate Mandarin labels. So every class is translated into a noun phrase, and every property starts with a verb in the Mandarin labels.

Table 5 shows examples of our manually annotated property labels in both English and Mandarin according to the suggested standardised term modelling approach. However, because of the different sentence structure between English and Mandarin, instead of ending each property with a preposition, Mandarin properties are ended with nouns. Therefore each axiom can be transformed to start with a noun subject class, and what ever happens in between, then it ends with a property in a Mandarin sentence.

Table 5 Example of property labels in suggested standardised form

SCT ID	English	Mandarin
255234002	has an after affect in	有(has)後遺症(after affect)
246454002	has an occurrence in	出現(appears)異常(difference)
116676008	has an associated morphology in	有(has)關聯的(associated)形態(morphology)
363698007	has a finding site in	有(has)病灶(finding site)
263502005	has a clinical course in	有(has)臨床的(clinical)療程(course)

The following text is an example that shows English text ending with a verbalised class but Mandarin ends with a verbalised property:

English: Renal arterial hypertension is a kind of renovascular hypertension that has a finding site in a kidney

Mandarin: 腎臟(renal)動脈(artery)的(apostrophe)高血壓(hypertension)是(is)一種(a kind of)腎血管性(renovascular)高血壓(hypertension)中(among)在(at)腎臟(kidney)結構(structure)上(upon)有(has)病灶(finding site)。

4. Discussion

OntoVerbal-M currently produces well-structured English and Mandarin natural languages for the fragment of SNOMED CT so far studied. Natural language texts are easier to understand than DL based terminologies and ontologies, especially for non-DL users. This motivates the need for automatically generated verbalisations. When users in multiple languages for which there are no full translations want at least limited access to the content, then multilingual generation becomes important.

It is striking that the same rhetorical structure schema appears to be applicable across two such different languages, despite marked differences in grammar and syntax. This significantly reduces the effort required to produce verbalisers in different languages as significant portions of the verbalisation machinery can be re-used.

Clearly, there are dangers of erroneous verbalisations, particularly in "new" languages such as Mandarin. Our Mandarin labels were not developed by a team of Mandarin speaking medical experts, but purely from our own team knowledge. At this time they must be regarded as experimental and used for OntoVerbal-M's purpose only. However, a small survey from Mandarin speaking doctors without SNOMED CT knowledge has indicated that the non-underlined medical terms of OntoVerbal's output are appropriate to express the meaning of the output texts. The survey also indicates that the underlined words we have chosen are generally suitable for text fluency purposes.

Although OntoVerbal-M's output needs some linguistic polish, especially in plurality and articles in English, its design in, first: organising information into super, sub and equivalent classes, second: transforming OWL classes and properties into text, and third: the use of discourse structure for generating text, would apply to ontologies from any domain. Specific to SNOMED CT, OntoVerbal-M has annotated classes and properties for text fluency purposes, and particularly to deal with human anatomical phrasing. Its experiences have raised the issues in phrasing ontology terms in English and Mandarin.

In the future, we plan to evaluate the text generated by OntoVerbal-M in two aspects: a) whether the text is faithful to the ontology, so the subjects need to be ontologist to be able to read text and regenerate ontology axioms; b) whether the translation into Mandarin is equally faithful. The participants in evaluation a), and ideally b), need to be SNOMED CT experts so that they understand both the axiomatic descriptions and a natural language text. Failing this, we expect to ask a broad team of domain experts to identify definitions that appear questionable and then consult a more limited team of SNOMED CT experts about those identified as questionable. We will also explore if OntoVerbal-M's system architecture can adopt more language generators such as French, Spanish and German.

Acknowledgments. This work is part of the Semantic Web Authoring Tool (SWAT) project (see www.swatproject.org), which is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grant EP/G032459/1, to the University of Manchester, the University of Sussex, and the Open University. We are extremely grateful to Professor Donia Scott for her critical feedback on an earlier draft of this article. The Mandarin survey has been supported by doctors from Chang Gung Medical Foundation LinKo Taiwan. We also thank Dr. Wu from Kaohsiung to give his useful comments on the medical terms translation.

References

1. SNOMED-CT User Guide, http://www.ihtsdo.org/fileadmin/user_upload/Docs_01/SNOMED_CT/About_SNOMED_CT/Use_of_SNOMED_CT/SNOMED_CT_User_Guide_20090731.pdf
2. Baader, F., Horrocks, I., and Sattler, U.: Description logics as ontology languages for the semantic web. Lecture Notes in Artificial Intelligence. 2605, 228-248. (2005)
3. Baud, R., Lovis, C., Alpay, L. *et al*.: Modelling for natural language understanding. In: 17th Annual Symposium on Computer Applications in Medical Care (SCAMC-93), pp. 289-293. McGraw Hill, (1993)
4. Baud, R.H., Rodrigues, J.-M., Wagner, J.C. *et al*.: Validation of concept representation using natural language generation. Journal of the American Medical Informatics Association, 841. (1997)
5. Callaway, C.B.: Integrating discourse markers into a pipelined natural language generation architecture. 41st Annual Meeting on Association for Computational Linguistics. 1, 264-271. (2003)
6. Ceusters, W., and Spyns, P.: From natural language to formal language: when MultiTALE meets GALEN. In: Medical Informatics Europe '97, pp. 396-400. (1997)
7. Clark, H.H.: Psycholinguistics. MIT Press, (1999)
8. Dalianis, H.: Aggregation as a subtask of text and sentence planning. In: Florida AI Research Symposium, FLAIRS-96, pp. 1-5. J.H.Stewman, (1996)
9. Fliedl, G.n., Kop, C., and Vo¨hringer, J.r.: Guideline based evaluation and verbalization of OWL class and property labels. Data & Knowledge Engineering. 69, 331-342. (2010)

10. Horridge, M., Drummond, N., Goodwin, J. *et al*.: The Manchester OWL syntax. In: 2006 OWL: Experiences and Directions (OWLED'06). (2006)
11. Liang, S.F., Stevens, R., Scott, D. *et al*.: Automatic Verbalisation of SNOMED Classes Using OntoVerbal. In: 13th Conference on Artificial Intelligence in Medicine, AIME 2011, pp. 338-342. (2011)
12. Liang, S.F., Scott, D., Stevens, R. *et al*.: Unlocking Medical Ontologies for Non-Ontology Experts. In: 2011 Workshop on Biomedical Natural Language Processing, ACL-HLT 2011, pp. 174-181. (2011)
13. Linden, K.V., and Scott, D.: Raising the Interlingual Ceiling in Multilingual Text Generation. Proceedings of the Multilingual Natural Language Generation Workshop. In: International Joint Conference in Artificial Intelligence (IJCAI'95)t, pp. 95-109. (1995)
14. Mann, W.C., and Thompson, S.A.: Rhetorical Structure Theory: toward a functional theory of text organisation. Text. 8, 243-281. (1988)
15. McKeown, K.R.: Discourse Strategies for Generating Natural Language Text. Artificial Intelligence. 27, 1-41. (1985)
16. Reape, M., and Mellish, C.: Just what is aggregation, anyway? In: European Workshop on Natural Language Generation. (1999)
17. Schulz, S., Stenzhorn, H., Boeker, M. *et al*.: Strengths and limitations of formal ontologies in the biomedical domain. Electronic Journal of Communication Information & Innovation in Health. 3, 31-45. (2009)
18. Scott, D., Bouayad-Agha, N., Power, R. *et al*.: PILLS: A Multilingual Authoring System for Patient Information. In: the 2001 Meeting of the American Medical Informatics Association (AMAI'01). (2001)
19. Spackman, K.A., and Campbell, K.E.: Compositional concept representation using SNOMED: Towards further convergence of clinical terminologies. Journal of the American Medical Informatics Association, 740-744. (1998)
20. Sporleder, C., and Lascarides, A.: Using automatically labelled examples to classify rhetorical relations: an assessment. Natural Language Engineering. 14, 369-416. (2008)
21. Stearns, M.Q., Price, C., Spackman, K.A. *et al*.: SNOMED clinical terms: overview of the development process and project status. In: AMIA Fall Symposium (AMIA-2001), pp. 662-666. Henley & Belfus, (2001)
22. Walker, M.A., Joshi, A.K., and Prince, E.F.: Centering Theory in Discourse. Oxford University Press, (1998)