



Understanding terminology in texts

Rita Temmerman

Centrum voor Vaktaal en Communicatie

Erasmushogeschool Brussel

Summary

In this article we focus on the process of understanding terms in texts. We explain how a method for terminological analysis of specialised texts has been set up in order to develop ontologically-underpinned terminological resources. We will concentrate on the categorisation framework, which is used in terminography for structuring terminological information. This framework is currently implemented in a didactic software tool, called CatTerm. Student translators using CatTerm construct a knowledge model of a given domain while reading a corpus of specialised texts in the target and source languages of the translation exercise.

In this article we deal with the process of understanding, which terminographers and translators need to confront. We introduce our methodology to teach student translators how to acquire domain knowledge prior to translating specialised texts (e.g. technical, scientific, or legal texts). A bilingual corpus of specialised texts can help them acquire at least a basic insight into the domain at hand, familiarise with the terminology of the domain (in both source and target languages) and become aware of linguistic and cultural differences. We consider all this to be part of a domain knowledge acquisition process. From our own experience in teaching specialised translation courses, we have learned that this process is often underestimated by many of the aspiring translators. This is why *Centrum voor Vaktaal en Communicatie* (CVC) of Erasmushogeschool Brussel has developed the CatTerm software tool that guides student translators in constructing a knowledge model of a given domain.

From meaning to understanding

Ogden and Richards (1923) visualised the three aspects of meaning that are traditionally distinguished (language symbols, thoughts and referents) in a triangle, which is commonly known as the semiotic triangle, the triangle of reference or the triangle of meaning. In their analysis thoughts are created in our minds referring to referents (e.g. objects) in reality and language symbols are used to symbolise these thoughts. Meaning is established on the basis of the relation between language symbols and thoughts referring to referents.

The semiotic triangle depicts the relation between language symbols (words), thoughts and reality as a static situation, not as a process i.e. a continual movement back and forth from thoughts to words and reality, from words to thoughts and reality and from reality to words and thoughts.

In sociocognitive terminology (Temmerman, 2000), the shift from meaning to understanding changes the perception of the interaction between language symbols, thoughts and referents.

The focus on understanding places the human subject or ‘conceptualiser’, who is able to understand and to create thoughts based on his understanding, at the centre of attention.

In cognitive linguistics it is believed that thoughts of human subjects about the world cannot be objectified, as much of what we know and understand about the world is embodied, i.e. acquired via sensory perceptions (Chrisley and Ziemke 2002). A thought can never cover the aspects or features of a given referent all at once because of the many perspectives from which the referent can be approached or perceived. Human subjects will never share exactly the same thoughts about given referents because of the different ways these referents can be experienced. Focussing on understanding (instead of meaning) therefore allows us to emphasise the subjective nature of thoughts.

Moreover individual experience is considered to be situated, which means that the domain community’s tradition conditions the embodied experience of each member. Consequently, thoughts are to some degree intersubjective as they emerge in a cultural group (e.g. country, educational institution, company, science lab, etc.) whose members share a more similar approach in the perception of referents and who may negotiate and renegotiate their cultural understanding across time and space.

Understanding is not only diversified. It is also flexible as it derives from a continuously evolving process which relates present perceptions and knowledge via reflection to potential further perceptions and insights. Because of our focus on understanding, we prefer to talk about *units of understanding*. The constant development of units of understanding can be explained as the result of several simultaneously active factors. Examples are the urge for more and better understanding, the interaction between different language users, prototype structure in the understanding of categories which can be seen simultaneously as the result of and as one of the causes of meaning evolution.

Because cognitive models play an important role in the process of understanding terminologists and translators need not only distinguish units of understanding related to a specific subject area in texts, they must also try and get insight in how units of understanding relate to one another. For them gathering the information on how terms relate to units of understanding boils down to specialised discoursed analysis.

From theory to practice

In terminology theory, the perspective is the term. Conscious reflections on the distinction between generalisation and contextualisation in determining the difference between core meaning and various senses of a term are part and parcel of terminological analysis. Methods for the identification, analysis, recording and processing of terms are being developed from a cognitive perspective (e.g. Kerremans et al. 2006). Special attention is given to the role of terms in cognition. Terms are studied from a diachronic perspective in order to get better insight into the role of terminology in a transitional process of understanding (e.g. Dury 2005; Temmerman 2008). Methods of cognitive analysis are being applied in multilingual terminology projects. The focus is shifting to what the discipline of terminology can learn from interdisciplinary studies of cognition: e.g. research into metaphors, the dynamic nature of cognition and the role of language and semantics therein, terminological variation, etc.

The shift from meaning to understanding has important implications for terminography. Terminography incorporates several practical tasks, such as the identification and analysis of terms in specialised texts as well as the organisation and description of these units in terminological resources. Focussing on understanding first of all emphasises the fact that there are different degrees of understanding, depending on the type of user. A technical definition of a given term may be understood by a domain expert but not necessarily by a specialised translator. Consequently, one ‘ideal’ definition of units of understanding – whereby meaning is defined on the basis of unique and sufficient properties – is often neither possible nor desirable (Temmerman 1997). When terminological resources are developed, it is crucial to start from questions such as ‘who are the users and what information do they require?’. Studies of translation processes, for instance, have shown that translators of specialised texts need access to multilingual specialised dictionaries which specify the way terms are structured in a network of intra- and interlingual relations (e.g. Dancette 1994; Temmerman 2003). Intralingual relations specify how terms in a given language and within a given domain are related to one another. Interlingual relations specify how, in a given domain, terms from different languages are semantically related to one another. The dictionary by Dancette and Réthoré (2000) shows how this is done for terminology about retailing.

With respect to the relations between terms in one language and their translations in other languages, it should be noted that differences on the level of semantics do occur and should therefore be made explicit in multilingual terminological resources. These semantic differences may result from differences in the embodied and situated experience of a given referent by members of cultural groups. Many units of understanding are therefore not clear-cut but have prototype structure which should be accounted for in terminological resources, as is e.g. shown in the dictionary of Dancette and Réthoré (2000). The fact that referents may be conceptualised or perceived in different ways, is a possible explanation for both polysemy and synonymy (see e.g. Freixa 2006).

Another aspect that should be accounted for is the dynamic nature of units of understanding. Many (electronic) terminological resources, however, are usually quite static and are only changed or modified once in a while, mainly because they are managed by one person or by only a few people. We believe that terminological resources can better keep up with the changes in disciplines such as biology or medicine if their development and maintenance becomes the responsibility of their entire user community. CVC explored this idea in a project about competency-based occupational profiles, in which a collaborative on-line platform was developed for small medium-sized enterprises. For further details about this project, we refer to Kerremans et al. (2006).

The shift from meaning to understanding clearly has implications for terminography. Obviously, some of the issues addressed in the previous paragraphs are also important for (aspiring) translators. Consider for instance the fact that a term may have different interpretations, depending on the contexts in which it is used, or that several lexicalisations denoting the same referent may be the result of differences in categorisations and should therefore not be considered as fully equivalent.

In order to become aware of these issues, translators first need to acquire insight into a given domain before they start translating specialised texts. In practical specialised translation seminars, aspiring translators are usually asked to read a number of texts in the source and target languages and to extract from these texts terms and translation equivalents which they

add to a bilingual terminological database. Although this is a good exercise for students learning to identify essential units of understanding related to a given subject area, it still needs to be examined whether this level of knowledge suffices to understand the specialised texts and to make the right choices with respect to the translation.

In our view, students have reached a good level of understanding when they can not only identify the main units of understanding in a specific domain but are also able to explain how these units are related to one another.. They also need to be aware that the meaning of terms may vary, depending on the context in which they are used. In other words, it is essential that students go a few steps beyond just copying terms and definitions from texts and pasting them into a term base. In order to understand how units of understanding are related to one another, we teach our students how to analyse linguistic contexts in specialised texts which provide essential information to increase their knowledge of one or several units of understanding. It should be noted that it is the trainer of the translation seminar who decides in a first phase what texts the students should read in preparation of their translation assignment.

Construing domain knowledge via terminological understanding, is a difficult process for which students should get efficient support. The CatTerm software was developed by CVC for guiding students through different knowledge acquisition steps.

The idea of construing a domain knowledge model, which is the main purpose of the didactic software, is based on termontography, a method set up to construct ontologically-underpinned terminological resources. We briefly present the method and reflect on one important methodological step in particular, i.e. the creation of a categorisation framework which defines the scope of each terminological project and allows for the structuring of terms and knowledge-rich contexts (Meyer et al. 1997) extracted from specialised texts.

The categorisation framework in termontography

Termontography combines principles of sociocognitive terminology and principles of text-based, application-oriented ontology development (Kerremans et al. 2003). The method was originally set up to support the work of terminographers in projects about developing multilingual, ontologically-underpinned terminological resources (Kerremans et al. 2003).

The application-oriented view in termontography motivates the need for a knowledge analysis step preceding the compilation of a corpus of specialised texts and the extraction of terms and other information (such as definitions) from these texts (Kerremans et al. 2003). The aim of the knowledge analysis step is to define the user requirements and scope with respect to the development of an ontologically-underpinned terminological resource or 'termontological' resource.

One of the main tasks in termontography is setting up a categorisation framework. This framework is a model of specialised knowledge which – time and financial resources permitting - might be created in collaboration with domain experts. It is further refined on the basis of an analysis and understanding of terms in specialised texts. This entire process is now supported by software tools (see e.g. De Baer et al. 2006a).

The categorisation framework, partly shown in figure 1, is composed of units of understanding which we call ‘categories’ as they are used as a means for classification. The screenshot is taken from an exercise on the topic of beer brewing. Examples of categories in figure 1 are: *abdijbier* (i.e. ‘abbey beer’) and *alcoholische drank* (i.e. ‘alcoholic beverage’). The representation of these categories is not based on formal specifications but relies on lexical items. These lexical items may be single-word terms, complex terms or even longer expressions (cf. Kerremans et al. 2003).

Categories may also be linked to one another on the basis of whole-part relationships. It should be noted that apart from generic-specific and whole-part relationships, the categorisation framework also allows us to specify all sorts of associative relationships. In this sense, its structure is much more refined than thesauri or taxonomies. For a more elaborate discussion on this, we refer to De Baer et al. (2006b and 2008). From the moment we specify an associative relationship, the category is no longer considered as a means for classification (cf. *supra*), but becomes a concept.

Apart from adding units of understanding (i.e. categories or concepts), terms and relationships, the categorisation framework also provides the possibility to add all sorts of extra information, such as definitions or any other type of descriptive information. For instance, in case of a semantic difference between a source language term and its translation equivalent or between denominative variants within the same language. The difference may be explained in a field called ‘transfer comment’. A transfer comment is meant to warn users of the termontological resource (e.g. translators) against mistakes in transferring information from one situational or cultural context to another. In the case of describing a possible semantic difference between a source language term and its translation(s), this descriptive field roughly corresponds to the “relations internationelles” in the dictionary by Dancette and Réthoré (2000).

For a more detailed discussion of the categorisation framework in termontography, we refer to Kerremans (2004) and De Baer et al. (2006b and 2008). In the next section, we will show how the framework is implemented in the CatTerm software tool which was developed for student translators.

Software support for student translators

Translators need to go through a number of knowledge acquisition steps if they want to produce high-quality translations of specialised texts (e.g. technical, scientific and legal texts). Relying on a bilingual corpus of specialised texts they should acquire at least a basic insight into the domain at hand (e.g. Zanettin 1998; Bowker 1999; Wilkinson 2005), become familiar with the terms that are used in the domain (in both source and target languages) and become aware of the semantic differences between terms, synonyms and translation equivalents, which they can specify in the field of transfer comments.

From our own experience in teaching specialised translation seminars, we have learned that these knowledge acquisition steps are often highly underestimated by many of the aspiring translators. Since we believe that during translation training more time should be devoted to guiding student translators in a consistent manner through the knowledge acquisition steps, we developed CatTerm.

As a platform-independent interactive computer programme, CatTerm will guide student translators through four main tasks that we consider necessary for knowledge acquisition: constructing and analysing a bilingual corpus of specialised texts, manually identifying terms in both source and target languages, creating a categorisation framework which is used to classify the terms, describing the terminology in the target language.

The four tasks in the knowledge acquisition process are split up into seven methodological steps that have been incorporated in the wizard interface of the software. As the purpose of this exercise for students is to come up with a categorisation framework that is based on information in the target language (which is by default the student's native language), they first need to read target language texts from a corpus compiled by the translation seminar trainer (step 1). It should be noted that students also learn to look for reliable texts (e.g. on the internet), similar to those in the basic corpus, and learn to consult alternative sources, such as (multilingual) glossaries, thesauri and specialised translation dictionaries that are relevant for the given domain.

By reading the texts in the corpus, they will acquire more insight into the domain and become familiar with the terminology in the target language. They are asked to highlight terms as they read selected texts (step 2). Relying on their understanding of the texts, students start building a categorisation framework (step 3). Afterwards, they assign definitions to terms in the framework (step 4). During this step, they also learn to search and formulate descriptions for the terminology of the domain. Next, students select texts from the corpus in the source language (step 5). The text(s) they will be asked to translate are part of the source language corpus. They first read these text(s) and identify the terms (step 6). The categorisation framework based on the target language corpus analysis is used as a reference model for structuring the extracted units in the source language (step 7). Additional source language texts may be consulted in order to complete the categorisation framework. At the end of the process students can have the computer generate a bilingual ontologically-underpinned dictionary based on their own work. In this respect, CatTerm is also a terminology management software tool. Yet, the emphasis is much more on the user's acquisition of domain knowledge. The bilingual terminological dictionary is a by-product of the knowledge acquisition exercise.

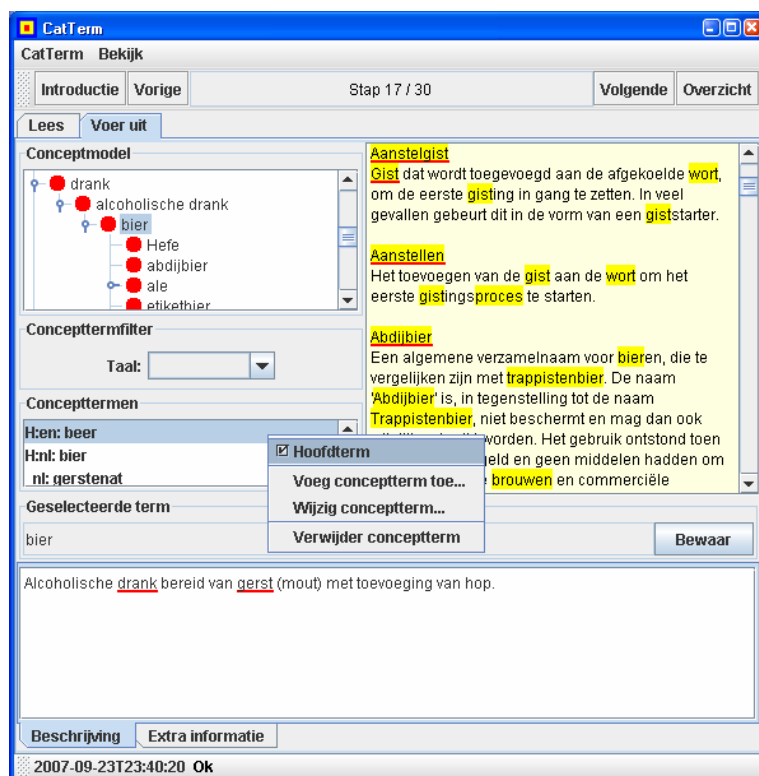


Figure 1: Screenshot of one step in the CatTerm wizard

Figure 1 is a screenshot of the CatTerm tool. The right hand panel in this screenshot shows the selected text in the target language. In this particular example, the text consists of a glossary list of terms that are used to denote units of understanding related to the process of beer brewing. The Dutch language highlighted words in this example such as *gist* (i.e. yeast), *abdijbier* (i.e. abbey beer) or *trappistenbier* (i.e. Trappist beer) are terms that were selected by a student. These terms are structured in the categorisation framework, which is shown in the left-hand panel. The framework can be viewed in either the source or target language.

By selecting a category in the framework, CatTerm shows the list of terms (including synonyms and translation equivalents) that students have assigned to this category. Terms that are preceded by the 'H' label are marked as 'head term'. These terms are used to represent the category in the categorisation framework. Finally, the lower panel in the figure shows that, students can add a definition to a term that has been selected in the term list.

The CatTerm software was tested in the framework of a master's thesis (Van Poelvoorde 2008). The study showed that the software is effective in the sense that it obliges student translators to be critical about the information they find in specialised texts and to reflect thoroughly on their understanding of terms and translation equivalents. At the same time, the study listed some disadvantages concerning the pertinence of the current version and formulated suggestions that may lead to improvements. For instance, our choice to simplify

the categorisation process by allowing student translators only to define generic-specific relationships was considered too restrictive.

Conclusion

By discussing CatTerm, we showed how student translators learn to understand and categorise knowledge by analysing specialised discourse. Starting from a categorisation framework which they set up on the basis of an understanding of a bilingual corpus of specialised texts, they are able to structure and define terms (including synonyms and translation equivalents) and construe domain knowledge.

An additional advantage of our approach is that the trainer of translation seminars has the means to monitor the student's cognitive acquisition process. The use of relevant information sources and texts can be checked. The development of the concept model can be followed from the start. Another advantage is that students can compare their concept models. This allows them to discuss their choices and learn from each other. Most important, however, is that students learn to acquire domain knowledge thanks to CatTerm.

Our work in progress on the impact of the distinction between lexical, situational and cognitive contexts and on the possibilities to apply this distinction in the terminological analysis of a text corpus should bring us to insights into how terminological resources should be adapted when circumstances change, like in the case when textual corpora are updated or when the domain expands.

A test version of CatTerm is freely available from <http://cvc.ehb.be/Software.html>

References

- Bowker, Lynne (1999) Exploring the potential of corpora for raising language awareness in student translators. *Language Awareness*, 8 (3/4). 160-173.
- Chrisley, Ronald /Ziemke, Tom (2002) "Embodiment". In: *Encyclopaedia of cognitive science*. London: Macmillan Publishers: 1102-1108.
- Dancette, Jeanne (1994) Comprehension in the translation process: an analysis of think-aloud protocols. In: Dollerup, Cay /Lindgaard, Annette (eds.) (1994): *Teaching Translation and Interpreting 2. Insights, aims, visions*. Amsterdam /Philadelphia: John Benjamins. 113-120.
- Dancette, Jeanne /Réthoré, Christophe (2000). *Dictionnaire analytique de la distribution. Analytical dictionary of retailing*. Montréal: Les presses de l'Université de Montréal.
- De Baer, Peter /Koen Kerremans /Temmerman, Rita (2006a) Bridging communication gaps between legal experts in multilingual Europe: discussion of a tool for exploring terminological and legal knowledge resources. In: Corino, Elisa Turin et al. *Proceedings of the XII Euralex international congress*. Alessandria: Edizioni dell'Orso. 813-818.
- De Baer, Peter /Kerremans, Koen /Temmerman, Rita (2006b) Facilitating ontology (re)use by means of a categorisation framework. In: Meersman, Robert et al. *On the Move to Meaningful Internet Systems 2006. Proceedings of the AWeSOME workshop*, Montpellier. Berlin: Springer. 126-135.
- De Baer, Peter /Kerremans, Koen /Temmerman, Rita (2008) A Categorisation Framework Editor for Constructing Ontologically underpinned Terminological Resources. In: Daille, Béatrice et al. *Proceedings of the sixth international conference on language resources and evaluation (LREC)*, Marrakech [CD-ROM]. 327-359.
- Dury, Pascaline (2005) Terminology and Specialised Translation, the Relevance of the Diachronic Approach. *LSP and Professional Communication*, 5 (1). 31-42.
- Freixa, Judit (2006) Causes of denominative variation. *Terminology*, 12 (1), 51-77.

- Kerremans, Koen (2004) Categorisation frameworks in termontography. In: Temmerman, Rita /Knops, Eugenia. (eds.). *The Translation of Domain Specific Languages and Multilingual Terminology Management*. [Special issue of *Linguistica Antverpiensia New Series* 3/2004. 263-277.
- Kerremans, Koen /De Baer, Peter /Temmerman, Rita (forthcoming) Towards A Multilingual Termontological Resource of Competency-based Job Descriptions for SMEs. *Proceedings of the International Conference on Terminology*, Antwerp.
- Kerremans, Koen /Temmerman, Rita /Tummers, José (2003) Representing multilingual and culture-specific knowledge in a VAT regulatory ontology: support from the Termontography approach. *Lecture Notes in Computer Science*. 662-674.
- Meyer, Ingrid /Skuce, Douglas /Kavanagh, Judy /Davidson, Laura (1997) Integrating linguistic and conceptual analysis in a WWW-based tool for terminography. In: Lessard, Greg /Levison, Michael (eds.). *Proceedings of the joint international conference of the Association for Computers and the Humanities and the Association for Literary & Linguistic Computing*. Ontario: Queen's University. <http://www.cs.queensu.ca/achallc97/papers/p021.html>
- Ogden, Charles Kay /Richards, Ivor Armstrong (1923) *The meaning of meaning: a study of the influence of language upon thought and of the science of symbolism*. London: Kegan Paul, Trench, Trubner & Co.
- Poelvoorde, Helmut Van (2008) *Opstellen van een categorisatieframework met CatTerm ter voorbereiding van het vertalen van gespecialiseerde (bio)medische teksten over hematologie*. Brussel: Erasmus-hogeschool Brussel (MA thesis).
- Temmerman, Rita (1997) Questioning the Univocity Ideal. The Difference between Socio-Cognitive Terminology and Traditional Terminology. *Hermes. Journal of Linguistics*, 18. 51-90.
- Temmerman, Rita (2000) *Towards new ways of terminology description. The sociocognitive approach*. Amsterdam/Philadelphia: John Benjamins.
- Temmerman, Rita (2003) Innovative methods in specialised lexicography. *Terminology*, 9 (1). 117-135.
- Temmerman, Rita (2008) Sociocultural situatedness of terminology in the life sciences: the history of splicing. In: Roslyn, Frank et al. (eds.). *Body, language and mind. Vol II. Interrelations between biology, linguistics and culture*. Tübingen: Springer. 327-359.
- Wilkinson, Michael (2005) Using a specialized corpus to improve translation quality. *Translation Journal*, 9 (3). <http://www accurapid.com/journal/33corpus.htm>.
- Zanettin, Federico (1998) Bilingual Comparable Corpora and the Training of Translators. *Meta* XLIII (4). 616-630.