

Helping language professionals relate to terms: Terminological relations and termbases

Elizabeth Marshman, Julie L. Gariépy and Charissa Harms, University of Ottawa

ABSTRACT

Terminological relations constitute critical elements of knowledge in specialised fields and their expression is important for language professionals working in these fields to master. Relations can be expressed using a wide variety of lexical relation markers representing a broad range of relation types and sub-types, as well as additional elements that help to identify the nuances of the relations and the participation of elements in them that must be distinguished for full comprehension. Nevertheless, humans can generally interpret these expressions of relations relatively easily and use them to build their understanding of subject fields. Unfortunately, conventional termbases rarely include examples of these relations, and computer tools are not able to comprehensively and reliably identify them in all cases. We argue that storing examples of terminological relations in (translation-oriented) termbases can benefit language professionals by enhancing both comprehension and expression in specialised fields.

KEYWORDS

Translation, terminology, terminological relations, lexical relation markers, termbases.

1 Introduction and objectives

Terminological relations (i.e. relationships that hold between terminological units or the concepts they denote) are among the key pieces of information analysed by terminologists in the course of their work, and are called upon by writers, translators and subject-field specialists to ensure their own comprehension of specialised domains, to evaluate equivalence between terms in different languages, and to produce clear, precise, high-quality informative texts for readers. Key relations often identified are those between generics and specifics (e.g. *cancer* and *carcinoma*), parts and wholes (e.g. *nucleus* and *cell*), entities and functions (e.g. *mammogram* and *cancer screening*), and causes and effects (e.g. *chemotherapy* and *hair loss*).

Since terminological relations are such key elements in our understanding of concepts in specialised fields, they provide an excellent starting point to help language professionals familiarise themselves with a new domain and its language. Unfortunately, information about terminological relations is largely reduced to a few, subtle elements in traditional term record models. For example, one or two text excerpts containing descriptions of terminological relations may be used as contexts on term records (e.g. Pavel and Nolet 2001), or (as noted e.g. in Meyer *et al.* 1999) such excerpts may be used as raw material for formulating definitions. Nevertheless, much of the information gathered never reaches the final product. Only rare terminological resources¹ explicitly store examples of relations.

In this article, we aim to highlight the information that can be usefully extracted from occurrences of terminological relations in corpora and the ways that making this information easily available in terminology resources could benefit language professionals in specialised fields. We argue that increased attention should be paid to the storage of occurrences of terminological relations in (translation-oriented) termbases. Using observations from a bitext corpus of popularised texts in the medical field, we will highlight the usefulness of information not only about the relations linking specific terms and concepts, but also about the ways these relations are expressed.

We begin by highlighting the context and some of the literature that has discussed the analysis and identification of terminological relations (sections 2.1 and 2.2 respectively) as well as some associated challenges (section 2.3). We then introduce our perspective on terminological relations in translation-oriented termbases (section 2.4). We outline the methodology we used to gather data (section 3) and a sample of results illustrating some of the benefits of storing terminological relations in these termbases, as well as some associated challenges (section 4). Finally, we sum up with some concluding remarks and suggestions for future work (section 5).

2 Context

The design and presentation of terminological resources are evolving rapidly. In the language industry, with intense time pressure and limited resources, terminology management practices must be as efficient as possible. Advances in computational power and the availability of software – and indeed changes in the ways that we view terminology and its goals – have revolutionised the ways terminology is managed. In a few decades, we have progressed from terminology stored on collections of index cards to resources as varied as massive online term banks, large-scale ontologies managing knowledge in specialised fields, and terminology management systems integrated into translation environment tools (TEnTs). Clearly, the structure of term records established decades ago is no longer optimal in all cases. However, what remains to be determined is how – and in how many different ways – terminology resources can be optimised.

Change is being reflected even in resources that we generally expect to be among the most stable: both the Canadian federal government term bank TERMIUM® and the Office québécois de la langue française's *Grand dictionnaire terminologique* are undergoing or have recently undergone transformations behind the scenes to ensure that they can continue to develop and change with the needs of their creators and users. The appearance of formats such as TBX-Basic (Melby 2008, LISA SIG 2009) and TBX Glossary (Wright *et al.* 2010) based on the TBX standard (LISA

2008; Melby 2008) demonstrates that applications and users can be varied enough to justify the development of different standards.

Even as terminology management standards for organisations evolve, individuals are finding their own strategies. Researchers (e.g. O'Brien 1998; Bowker 2011) have noted that users' solutions often differ substantially from traditional terminological models, for example including fewer formal definitions, and most likely relying more on corpus-based data. Essentially, professionals often seek strategies that require less time investment but provide a good return by guiding the correct, precise use of terms. They may store terminology in a variety of formats, from spreadsheets to generic databases to termbases in terminology management systems (e.g. L'Homme 2004).

Terminology storage and consultation options have also evolved. Storage space for electronic data is rarely a significant limiting factor. Electronic formats offer far more freedom in the amount of information that can be stored on a single record. Terminology management systems offer a variety of options for personalising record structures, allowing users to choose the number and types of fields they use and to decide whether these should be single or multiple, optional or mandatory. To compensate for the potential drawbacks of storing more or expanded information on records, software tools offer more choices than ever for displaying data: displaying only completed fields; viewing or hiding specific fields during consultation and/or searching if the user wishes; and in TEnTs, generally displaying only terms and equivalents from termbases during the interactive translation process, making the full records available for manual consultation if more information is required. This means that users may choose to include a wide range of data on records, and may consult the relevant parts of this data at any given moment with relative ease.

Therefore, it is appropriate to consider adjustments and additions to terminology management practices and to examine their potential contribution to the work of language professionals. The focus of this article is the identification and storage of information about terminological relations, and the balance that we believe is possible between a reasonable investment of time in the identification and storage of this information and the potential return in better understanding and expression in specialised fields.

2.1 Terminological relations and their analysis

The understanding of concepts and the terms that denote them is dependent in large part on the understanding of relationships that link concepts to others (and terms to other terms) and that ultimately structure specialised fields. The identification, analysis and expression of terminological relations are central in learning, writing and translating in specialised domains.

From a traditional, conceptual perspective in terminology, researchers (e.g. Sager 1990, Nuopponen 2005, 2010, 2011) have identified a considerable range of potentially relevant relations. It is generally agreed (e.g. Sager 1990, Meyer 2001) that the two most commonly studied terminological relations are specific to generic (e.g. carcinoma is a type of cancer) and part to whole (e.g. the nucleus is part of a cell). These hierarchical relations are used to generate the types of concept systems traditionally used in terminology projects (the biological taxonomy being the best-known example). The generic-specific relation also constitutes the starting point for the traditional Aristotelian definition of *genus* plus *differentia* (e.g. carcinoma in situ is a carcinoma that is confined to the epithelial tissues in which it originated), making it a natural indicator of defining information in texts (e.g. Pearson 1999, Rebeyrolle 2000).

However, researchers are increasingly considering a number of equally relevant non-hierarchical relations. For example, it is hard to imagine grasping the intricacies of the biomedical field without considering cause-effect relations (e.g. the causes of diseases or the effects of their treatments), understanding the field of epidemiology without studying association (i.e. the significant co-occurrence of variables, cf. Hennekens and Buring 1987: 30)² (e.g. the link between physical exercise and incidence of breast cancer), or comprehending the field of computer science without considering entity-function relations (e.g. that a monitor is used to display data, and a printer used to print documents).

The creation of concept systems is still often considered to be a necessary part of thematic terminology work, and the terminological relations that hold between elements of this system are some of the most important elements in the *crochet terminologique* (Dubuc 2002) that helps to establish equivalence between terms. However, all too often the extensive analysis of terminological relations required for this work (e.g. choosing terms and concepts to be included in terminological resources, evaluating equivalence between terms, and describing concepts) is minimised in term records, dictionaries or glossaries, or must be unearthed from definitions, contexts, or observations by attentive users.

A number of researchers have highlighted the considerable gap between terminology practice and products and the need for terminological resources that make information accessible to both human and machine users. For many years, the relation-rich terminological resource envisioned was referred to as a terminological knowledge base (TKB) (e.g. Meyer *et al.* 1992, Condamines and Amsili 1993, Otman 1994, Meyer 2001, Condamines and Rebeyrolle 2000, 2001). Today, interest is often focused on detailed and machine-readable knowledge representation in ontologies (e.g. Gillam *et al.* 2005, Malaisé *et al.* 2005, Roche (ed.) 2010). These perspectives share an emphasis on the fundamental nature of terminological relationships for understanding and representing

specialised fields, and the importance of storing them in an accessible and usable way.

2.2 Discovering relations

In today's digitised and technologised world, it is almost unthinkable to do terminology work without electronic corpora and corpus analysis tools such as monolingual and bilingual concordancers (e.g. Bowker and Pearson 2002, L'Homme 2004). Corpora serve as the basis for the discovery of terms, their attestation, the identification of information about their meanings, and the study of the conditions of their use. Moreover, corpora are not beneficial for terminologists alone. Trainee and professional translators and writers can make use of corpora to research vocabulary and terminology (Meyer and Mackintosh 1996a, Pearson 1998, Bowker and Pearson 2002), and in fact it has even been noted that the analysis of corpora may be preferred in some cases to the use of more conventional resources such as term records (Bowker 2011). They are also rich sources of information for concept analysis (e.g. Meyer and Mackintosh 1994, 1996b), and specifically information about terminological relations.

Moreover, as translators make increasing use not only of comparable but also of translated and aligned documentation (e.g. bitexts and translation memories) (Bowker 2011), they often have access to parallel relation occurrences and the useful information they include in two or more languages.

A number of strategies can be employed for the identification and extraction of terminological relations from corpora. They can largely be divided into two categories, the first relying mainly on statistical approaches to corpus analysis (e.g. co-occurrence and distribution) and the second on the recurrence of specific linguistic and paralinguistic items (e.g. L'Homme and Marshman 2006). The most commonly used of the linguistic approaches focuses on the identification of what Meyer (2001) referred to as *lexical knowledge patterns*. These are recurrent patterns in which a lexical unit or series of lexical units expresses the relation between two terms or other items (e.g. the marker *is a type of* identifying the presence of a generic-specific relation in statements such as *carcinoma is a type of cancer*, or *leads to* indicating a cause-effect relation in statements such as *chemotherapy leads to hair loss*).

2.2.1 Using lexical relation markers

Human readers tend to interpret fairly easily the relation expressed in lexical knowledge patterns. However, computers may also be programmed to use patterns to analyse corpora. Hearst (1992) is most often credited with the early use of lexical markers of relations for automatically identifying relations in general language, but was swiftly followed by many

others in the terminology field (e.g. Ahmad and Fulford 1992, Jouis 1993, 1995, Bowden *et al.* 1996, Meyer *et al.* 1999, Morin 1999, Séguéla 1999, Feliu 2004, Gillam *et al.* 2005, Malaisé *et al.* 2005, Halskov 2007, Halskov and Barrière 2008). These projects revealed both the potential usefulness of lexical relation markers for finding occurrences of terminological relations and some of the challenges of this task. While tools for this purpose are currently not widely available commercially, we can explore their potential to understand how they compare with other options.

Approaches using lexical markers to locate occurrences of terminological relations generally depend on pattern-matching: searching for lexical markers represented by character strings or regular expressions, often in proximity to a term being researched. Once occurrences of these markers are located, they can be analysed to identify the specific terminological units or other items they link, and the specific relationship between them. Such analyses can be done manually, or assisted by computer tools. The relations discovered can then be represented in various ways to make them readily accessible to users.

2.3 Challenges of discovering and analysing terminological relations

While it may at first seem fairly straightforward, automated identification of useful terminological relations in texts using lexical relation markers involves a number of significant challenges. These challenges result for instance from the nature of terminological relations and their role in domains and the nature of lexical knowledge patterns.

2.3.1 Terminological relations

While many relations and their importance are easily recognised, their analysis is often complex. Many studies have analysed the definition, nature and representation in texts of specific relations including part-whole (e.g. Winston *et al.* 1987, Iris *et al.* 1988, Borillo 1996, Jackiewicz 1996, Otman 1996, Condamines 2000), cause-effect (e.g. Nuopponen 1994, Garcia 1996, 1997, Nazarenko 2000, Barrière 2002, Cabré *et al.* 1996, 2001, Feliu 2004, Marshman 2006), instrumentality (Sambre and Wermuth 2010) and association (Feliu 2004, Marshman 2006, Marshman and Vandaele 2010).

Formal relation classifications must first begin by defining the limits of the relation (to use the example of cause-effect relations, when does association of two variables cross the line into cause and effect? do cause-effect relations include causing something to happen, but also causing it not to happen, i.e. preventing something? what about changing how it happens, i.e. modifying something?). In addition, it is essential to consider a variety of relation sub-types (is there a single cause-effect pair, or does the effect lead in turn to another effect in a chain of cause? Is the

cause sufficient in itself to lead to an effect, or does it contribute to the effect along with other factors?). When the various perspectives are combined, fully representing all of the complexities and nuances of the relationships that can be relevant from different perspectives is clearly a momentous task.

Another type of challenge lies in the relevance of specific relations being greater or lesser depending on the field of work and the classes of concepts involved in relations; in fact, some relations may be particularly relevant only in a restricted set of fields (e.g. Séguéla 1999). This means that approaches in each new field may require considerable adjustment of the relations to be taken into account.

2.3.2 Nature of lexical knowledge patterns

Using lexical knowledge patterns to identify information automatically or semi-automatically also involves several challenges, largely because these patterns are segments of authentic texts, composed of lexical units. There is thus substantial potential for variation, not only of the lexical markers, but also the items they link within a given context and the structure in which they are found.

Perhaps most challenging, it is extremely difficult (if not impossible) to predict all of the possible lexical markers of a given relation in a given language. Research (e.g. Ahmad and Fulford 1992, Morin 1999, Séguéla 1999, Barrière 2001, Marshman *et al.* 2002, Feliu 2004, Malaisé *et al.* 2005, Marshman 2006) has identified a wide range of markers for various relations. Not all markers, however, are universally relevant: some are used primarily in specific domains, while others combine most frequently or even exclusively with certain classes of concepts. One example is the marker *chez* in French in the domain of natural sciences, used to indicate part-whole relations (Condamines 2000) (e.g. in Condamines' example, *chez les primates, le mandibule...* 'in primates, the jaw...').³ This marker is not a prototypical part-whole relation marker in general (cf. *est une partie de* 'is a part of' or *est composé de* 'is composed of'), but in the natural sciences was fairly commonly observed to refer to parts of living creatures' anatomy. Similarly, *is a species of* could identify specific types of living things (e.g. *the Spanish shawl nudibranch is a species of nudibranch*), but this would be difficult to imagine in another field and with another class of concept (e.g. **the lithium ion battery is a species of battery*). Lexical relation markers can be said to participate in collocations in specialised language, and to present both their relevance and their challenges (e.g. Clas 1994, L'Homme 1997, Heid 2001).

Text genre (Lee 2001, Condamines 2002, 2008, Jacques and Aussenac-Gilles 2006) may also influence the choice of markers: those used in scientific journals, for instance, may not be those chosen in popularised texts. For example, while the verb *inhibit* may be used to express a sub-

type of causal relation in specialised articles in the medical field, *reduce* or *prevent* might be more frequent in popularised texts in the same field (as they are likely to be more immediately understood by the intended audience). It can thus be challenging to guarantee the 'portability' of markers from one field to another and from one corpus to another. Studies analysing occurrences of markers found in one domain and text genre in others (e.g. Marshman and L'Homme 2008, Marshman *et al.* 2008a, 2008b, 2009) have noted that while some individual markers show consistent occurrences from corpus to corpus, some found in one corpus may be absent in others, or may be far more or less frequent. Some (e.g. Séguéla 1999) have postulated the existence of a fairly consistent, 'portable' core set of markers, which may then be complemented by more corpus-specific markers. As more and more analyses of various corpora are carried out, a 'core' set for key relations may begin to emerge. However, the wide range of communicative situations, genres and domains may considerably limit a standard marker set's usefulness both for locating relations and for expressing them.

If a standard set of markers is difficult to discover in one language, the task is even more complex in bilingual or multilingual work. Even with a set of known markers, it is extremely difficult to identify a corresponding set of markers in another language without independent analysis: many markers have multiple possible equivalents, each of which may have its own particular level of frequency, limitations and associations (Marshman and Van Bolderen 2008).

Another challenge of lexical knowledge patterns is their natural ambiguity as units of natural language. Ambiguity (e.g. Meyer *et al.* 1999, Séguéla 1999, Meyer 2001, Condamines 2002, Marshman 2006, Marshman and L'Homme 2006) may be observed in markers that can in some cases indicate a relevant terminological relation and in some cases another, non-pertinent sense (e.g. in the case of the marker *lead to*, which can indicate a causal relationship in structures such as *the mutation leads to uncontrolled cell growth*, but a completely different sense in *lymph vessels lead to lymph nodes*). In other cases, a marker can indicate more than one type of potentially relevant relation (e.g. the marker *includes*, which can indicate a generic-specific relation as in *ductal carcinomas include ductal carcinoma in situ and invasive ductal carcinoma* or part-whole relations as in *the treatment protocol includes chemotherapy and radiation*). Thus the use of lexical markers to identify specific types of relationships may produce 'noise' (i.e. non-pertinent results) and may require human intervention to identify relevant results. In some cases, even humans may have some difficulty in identifying the relation linking two items. This ambiguity is a concern for identifying and expressing relations in texts.

Moreover, as occurrences of natural language, lexical knowledge patterns do not follow invariable structures: they can change form and order (e.g.

this mutation causes uncontrolled growth; uncontrolled growth is caused by this mutation), can be interrupted by elements such as modals, intensifiers, attenuators, and modifiers (e.g. *this mutation can cause uncontrolled growth; this mutation invariably causes uncontrolled growth; this mutation sometimes causes uncontrolled growth; this mutation causes rapid, uncontrolled growth*). Expressions of uncertainty (studied e.g. in Marshman 2006, 2008), including modal verbs (e.g. *can, may*), hedges (e.g. *sometimes, potentially*) and even negation (e.g. *not, never*) obviously affect the ultimate usefulness of occurrences. While they still very often provide useful information, their content must be carefully evaluated to determine how the information should be interpreted. Another frequently observed phenomenon is the combination of multiple participants in relations (studied e.g. in Marshman 2006, 2007). In many occurrences of terminological relations, multiple participants may be indicated on one side of a relation (e.g. *the treatment protocol includes radiation and chemotherapy; chemotherapy can cause side effects such as fatigue, nausea and hair loss; inflammation may result from either infection or trauma*). These participants can be linked by conjunction (e.g. *X and Y*), disjunction (e.g. *X but not Y*) or even more complex relationships (e.g. generic-specific in *Xs such as Y, Z and W*). The need to determine whether the relationship in question holds between one or more than one pair of the participants adds a layer of complexity to interpreting the relation present.

Clearly, relations can be extremely useful for understanding the conceptual structures of specialised fields. However the tasks of identifying, then classifying and interpreting them according to the fine-grained analysis that may be required can challenge the human user. Identifying the participants in the relation and the certainty with which the relation is present, and expressing the relation with equal precision, can also be challenging. These tasks are even more difficult for computer applications. It is thus no wonder that mass-market commercial tools have not yet integrated functions to automatically identify and classify relations. However, humans can often interpret key information about relations expressed by relation markers with relative ease and precision.

2.4 A different perspective

We might conclude, then, that translators' best option for uncovering relations would be the simplest: to set aside the idea of identifying relation occurrences automatically and go directly to the corpus when information about terms is required, in a process that is becoming more and more commonplace (as noted e.g. by Bowker 2011). However, it is important to note that this kind of approach also has drawbacks. First, it could well lead to duplication of effort, with the translator repeating corpus searches multiple times to refresh his or her memory of specific information, or to look for new kinds of information involving a term or concept. Moreover, the almost inevitable investment of time in filtering

out noise from the occurrences identified would need to be repeated with each search.

One alternative, discussed in Marshman and Van Bolderen (2009), is that translators and other language professionals could reduce inefficiencies by storing contexts containing expressions of terminological relations in their termbases as they encounter them, and ideally annotating them with a minimal amount of information. This would give direct access to the original description of the occurrence, but also facilitate the analysis of key information for future use.

Below, we use examples from our bitext corpus to illustrate the various types of information useful for language professionals that can be obtained through corpus analysis and managed in termbases. First, however, we describe how we identified this information.

3 Methodology

For this project we built a bitext corpus of English and French Web documents for laypersons (e.g. patients) in the field of breast cancer. The corpus consisted of 16 pairs of Web documents from 6 Canadian organisations that provide information about the nature, diagnosis, prevention and treatment of breast cancer (e.g. the Canadian Breast Cancer Foundation, the Canadian Cancer Society, Health Canada). The corpus contained approximately 123,000 English and 143,000 French tokens.

The English and French texts were aligned using the LogiTerm aligner (Terminotix 2010) and candidate terms were extracted from the collection of English texts using the term extractor TermoStat Web (Drouin 2011) and the measure of specificity (Drouin 2003).

Following the extraction, approximately 150 of the most highly ranked candidate terms we considered relevant in the field of breast cancer were chosen for inclusion in a termbase in Microsoft Access.

We used the LogiTerm bilingual concordancer to search for occurrences of English terms, identified the French equivalent(s) present in the text, and then complemented this research by searching for occurrences of the equivalents to identify synonyms of the original term candidate identified. In addition, concordances were analysed to identify occurrences of five key terminological relations that involved the candidate terms: generic-specific, part-whole, cause-effect, association and entity-function. Occurrences were manually identified, extracted and added to the termbase in a relations table linked by the English term to the main term records. We then identified the relation type, the other item participating in the relation and the base form of the lexical marker of the relation.⁴

Figure 1 shows a model record for an occurrence of an association relation.

Relation:		Source:	
Association		Smith and Jones 2010	
English term		French term	
breast cancer		cancer du sein	
English relation marker		French relation marker	
link between... and		lien entre... et	
Related English term(s)		Related French term(s)	
melatonin		mélatonine	
Example of English term		Example of French term	
Some studies suggest a link between melatonin and breast cancer.		Selon certaines études il y aurait un lien entre le cancer du sein et la mélatonine.	
Notes:			

Figure 1. Analysed association relation occurrence

4 Results and discussion

The analysis produced a set of 920 annotated relation occurrences: 289 generic-specific, 101 part-whole, 338 cause-effect, 114 association and 78 entity-function. Based on these results, we discuss below the range of potentially useful markers expressing key terminological relations in association with domain terms in texts for laypersons, as well as the challenges of translating these markers, in order to highlight how such information can be useful for language professionals.

4.1 Possible applications

Since terminological relations are such key elements in our understanding of concepts in specialised fields, their collection from texts can provide an excellent starting point to help translators familiarise themselves with a new domain. Translators may consult stored relation occurrences for individual terms in order to get a quick overview of the pertinence of a term or the concept it denotes within a field, information that a standard terminological definition or a limited number of contexts could not fully provide. In a relational database structure such as ours, users can also consult the set of occurrences of a particular type of relation in order to view the markers commonly used to express it and how they are used (e.g. the types of terms, expressions of uncertainty, or modifiers with which they tend to combine).⁵ Finally, in a bilingual database, users can compare occurrences of relations and the markers used to indicate them in two or more languages to consider potential equivalents of the markers and the structures in which they typically appear. Examples of these uses are discussed below.

4.1.1 Understanding concepts through relations

Simply consulting a number of 'unprocessed' occurrences of terminological relations in texts can help users to better understand the place and significance of a given concept in a field. Figure 2 below shows terminological relations extracted from the English texts in our corpus that provide information about the concept expressed by the term *hormonal therapy* (shown in the centre of the figure in red) and drawn from a range of 33 relation occurrences involving this term. In this figure, generic-specific relations are represented by shades of blue, the generic in aqua and the specifics in darker blue. Association relations (in this case, expressed as risks) appear in green, cause-effect relations (largely involving intended effects, although side effects are also present) in purple, and function relations describing the purposes for which hormonal therapies are used in yellow. Labeling the arcs are the markers that identify the relation in each context, accompanied where appropriate by expressions of uncertainty or hedging (e.g. *can*, *is likely to*, *is not likely to*) that may affect their interpretation.

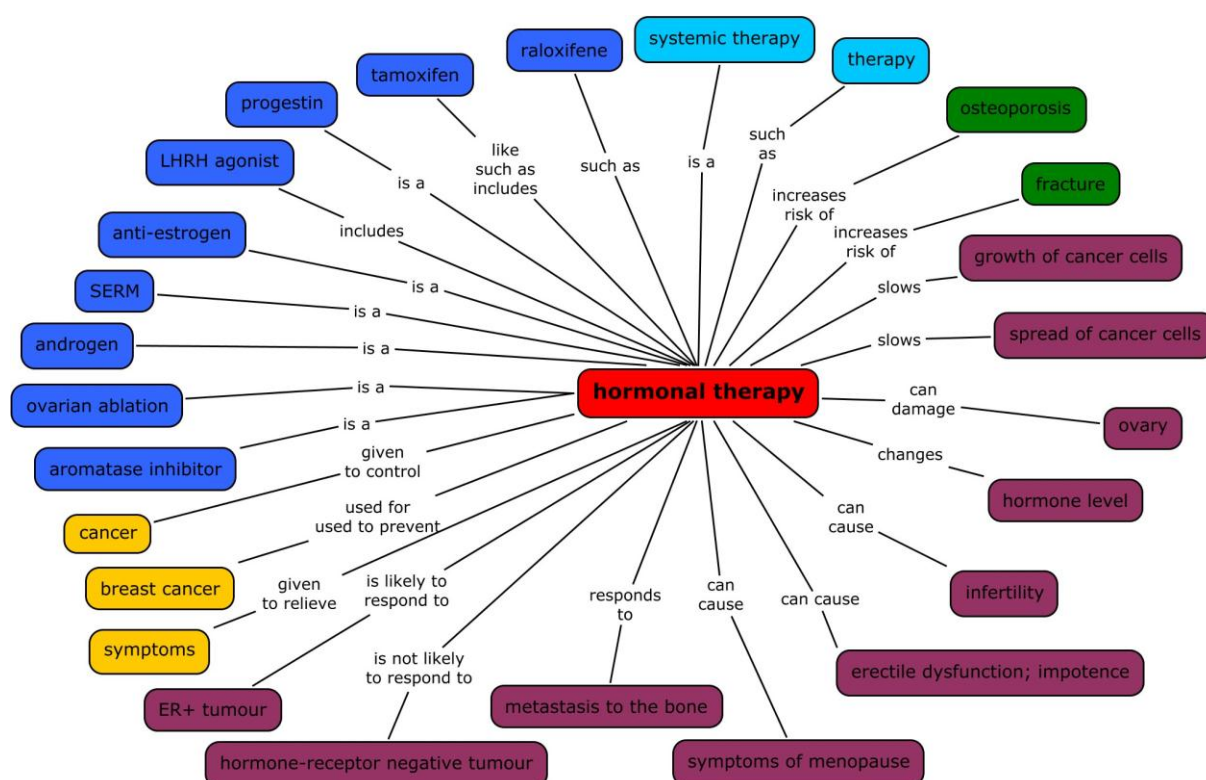


Figure 2. Relations extracted from the corpus for *hormonal therapy*

A language professional who accesses this information in a corpus and stores it for future use can easily review and identify not only minimal defining information (e.g. that hormonal therapy is a systemic treatment that uses means such as medications to slow the growth and spread of cancer by blocking the action of hormones) but also other key information for understanding the full significance of the concept in the field (e.g. the cases in which the treatment is most likely to be useful, the side effects it

may have). By condensing these relevant excerpts into a list of relations that can be sorted and grouped if desired, language professionals can simplify and accelerate future searches focusing on this term, and can also retain important information about nuances between the relations observed that might otherwise be lost or overlooked.

4.1.2 Choosing and varying markers

Figure 2 above shows some of the relation markers that can be used to identify key relationships, and reflects the variety of markers that may be used to express even a single relation involving a specific term in a given type of text (e.g. for generic-specific relations, *is a*, *include*, *such as*, and *like*).

As the relation occurrences were gathered, it became evident that certain markers were very frequently used in the occurrences identified, and that these were not necessarily the clearest or most precise options (see Table 1 below). For example, the generic-specific marker *is a*, as in *a carcinoma is a cancer*, is so multi-purpose that it may present ambiguity for the reader. Nevertheless, it was observed as the sole marker of generic-specific relations in 117 (40%) of the 289 identified relation occurrences of this type. Similarly, the marker *cause* was found in 42 (almost 12.5%) of the 338 occurrences of cause relations. In both cases, a number of other markers could be used, adding variety and in some cases precision to the expression of the relation. Certainly, it is possible that given the nature of the corpus texts used, which targeted laypersons, it was considered advisable to use very simple markers. However, if the existence of equally simple but much less ambiguous markers (e.g. *such as*, *including*, *type of*) were called to the attention of the language professionals who produce these kinds of texts, they might be encouraged to write in a more varied and/or more precise way.

Relation	Top markers identified	Examples of terms observed with markers
Association	risk of	aromatase inhibitor; coronary heart disease; hormone replacement therapy; lymphedema; mastectomy; recurrence
	associated with	coronary heart disease; hormone replacement therapy; mutation; radiation; risk factor
	after	breast cancer surgery; lymphedema; radiation therapy
	chance of	lymphedema; radiation therapy; recurrence
	is linked to	breast cancer risk; heart disease
Cause-effect	cause	alcohol; biological therapy; cancer

		treatment; chemotherapy drug; disease; lump; lymphedema; mutation; radiation therapy; side effect
	reduce	aromatase inhibitor; cancer treatment; mastectomy; radiation therapy; tamoxifen
	increase	hormonal therapy; radiation; risk factor; side effect; tamoxifen
	respond to	hormonal therapy; tamoxifen; trastuzumab
	affect	biological therapy; breast cancer surgery; breast tissue; diagnosis; hormonal therapy; radiation therapy; surgery; treatment option; tissue
Entity-function	is used to	hormone replacement therapy; surgery; chemotherapy drug; cancer cell; mammography; radiation therapy; tamoxifen
	do to	biopsy; diagnosis; lump; mammography
	given to	breast tumour; cancer cell; hormonal therapy; side effect
	goal of... is to	cancer cell; radiation therapy; surgery
Generic-specific	is a	abnormality; aromatase inhibitor; biopsy; breast-conserving surgery; breast reconstruction; chemotherapy drug; clinical breast examination; disease; hormonal therapy; inflammatory breast cancer; lobule; lump; lumpectomy; lymphedema; mammography; mastectomy; physical examination
	such as	aromatase inhibitor; biopsy; bone scan; breast-conserving surgery; chemotherapy drug; chest wall; heart disease; hormonal therapy; lump; lumpectomy; side effect; ultrasound
	include	chest wall; family history; hormonal therapy; lymphedema; mastectomy; physical examination; progesterone; treatment option; radiation therapy
	like	aromatase inhibitor; cancer treatment; chemotherapy drug; hormonal therapy; inflammatory breast cancer; lymph node; surgery; tamoxifen
	type of	biopsy; in situ breast tumour; invasive breast cancer; mastectomy; radiation therapy
Part-whole	in	axillary lymph node; blood vessel; cell;

		chest wall; duct
	of	abnormality; chest wall; duct; lobule; tamoxifen
	contain	cancer cell; cell; dioxin; lump; nutrient; progesterone
	from	blood vessel; cell; healthcare team; lump; radiation; radiation therapy; tissue
	found in	cancer cell; cell

Table 1. Top markers and examples of terms for relations analysed

The inclusion of examples of terminological relations in terminology resources (especially if these were minimally annotated) would provide users with access to lists of potentially appropriate markers that have been combined with the terms they are researching (or similar terms) as well as a means of comparing and contrasting markers. A list of possibly useful markers accompanied by examples illustrating their use could be a valuable asset, particularly for translators who are as yet unfamiliar with a domain and have not fully assimilated its language.

The potential benefits of increased text quality and precision offered by easy access to a list of candidate markers can be illustrated by examples involving the expression of association relations. The distinction between association and causation is a critical one (particularly in the health field), but laypersons (including translators who are unfamiliar with fields in which association is important) may not be sensitive to the distinction and how it is expressed. They might well benefit from being reminded of the various possible means of expressing relationships to help them to find the most appropriate one. (This will be discussed below in the context of translation.) Another example involves the rendering of the marker *affect* by *affecter* in French, a verb that is considered by some (e.g. de Villers 2003: 43) to be an anglicism in this sense. Access to alternative markers might help language professionals to avoid this and similar issues.

4.1.2.1 Translating markers

Whether for identifying relations automatically in corpora or expressing them in texts, establishing equivalence between markers or sets of markers is challenging (e.g. Marshman and Van Bolderen 2008). In the bitext corpus analysed in this project, none of the frequently observed markers shown in Table 1 had only a single observed equivalent. Numbers ranged from 2 (e.g. *réagir à* and *répondre à* for the cause-effect marker *respond to*, observed in 10 occurrences) to a wide range (e.g. as illustrated in Figure 3 and Figure 4 below).

The presence of a range of potentially useful markers for expressing the various types of relationships is evident when a network of markers is analysed. Our networks begin with the most frequent, prototypical marker

for a relation (i.e. *is a* for generic-specific relations and *cause* for cause-effect relations; shown in green in Figure 3 and Figure 4) and then the identification of the French equivalents in the relation occurrences analysed (shown in blue), followed by identification of other English equivalents of the French markers (shown in purple), and so on. The product of these analyses is shown below, the arcs labeled with the number of times the pair of markers was observed in the analysed relation occurrences. The analysis of the generic-specific markers (see Figure 3) identifies a series of 26 potential French markers to express the relation (e.g. *comme*, *consister en*, *est un*, *est un exemple de*, *est une forme de*, *par exemple*, *parmi*, *tel que*, *y compris*) and 10 potential synonyms or replacements for the marker in English (e.g. *include*, *is an example of*, *is a type of*, *such as*).



Figure 3. Network of markers starting with "is a"

The network of cause-effect relation markers (see Figure 4) is even more complex, with 26 possible French markers (e.g. *donner lieu à, provoquer, en raison de, engendrer, entraîner, mener à*) and 19 other English markers (e.g. *result in, lead to, play a part in, produce, due to, because of*).

Once again, a list of potential markers can facilitate and increase the quality of translation work by allowing users to compare alternatives and choose a marker that is precise, appropriate and suited to a given context.

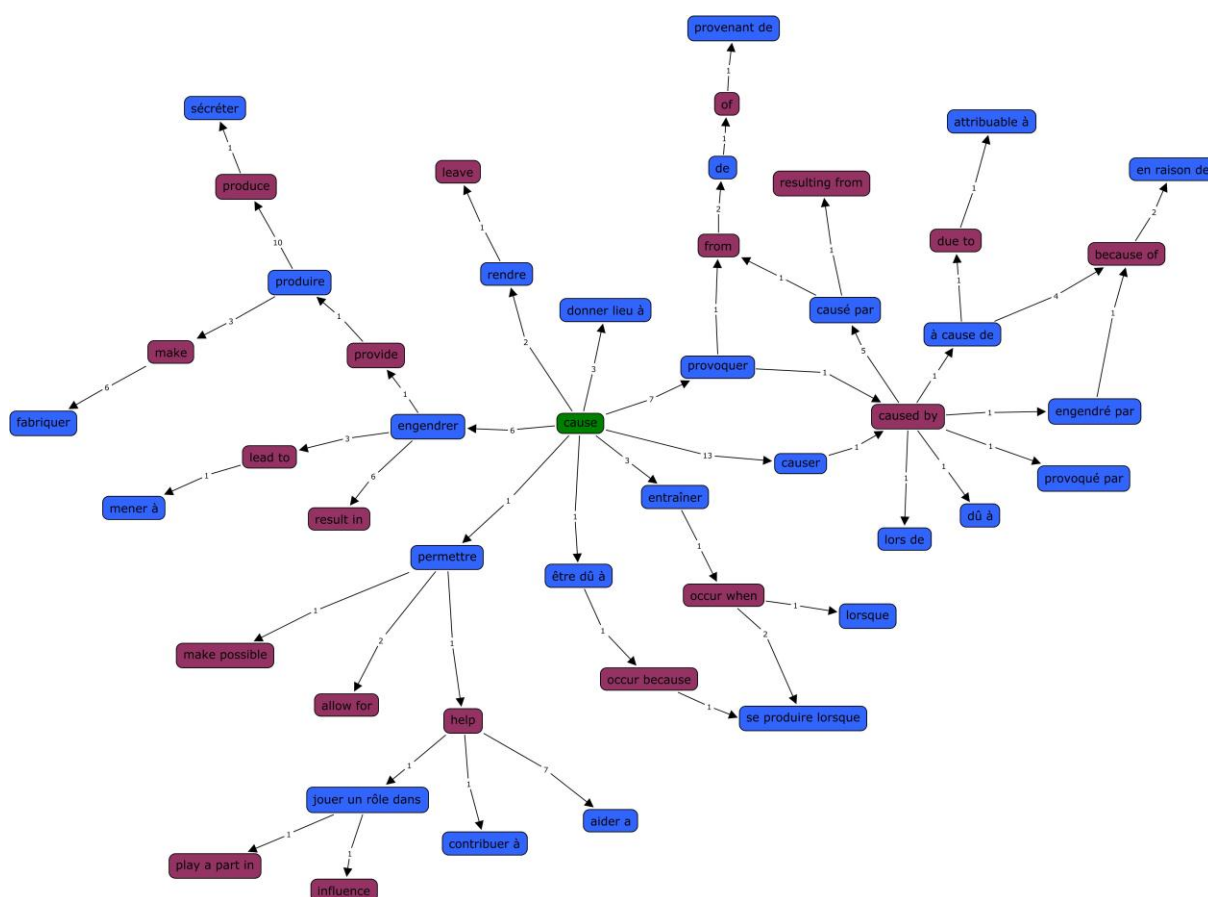


Figure 4. Network of markers starting with "cause"

As noted above, lack of familiarity with the fine distinctions between relations and the markers that express them may result in slippage in use (e.g. translation) of markers which can have a serious impact on the meaning of a text. Although these phenomena were rare in the corpus, a number of occurrences were identified in which English markers of association (e.g. *associated with*, *linked to*, *related to*) corresponded in the aligned document to markers of cause-effect relations (e.g. *engendrer* 'bring about', *causer* 'cause', *causé par* 'caused by', *entraîner* 'lead to'). Certainly, the presence of an association does not rule out the possibility of a cause-effect relation (and may even suggest it), but the French markers do convey a much stronger probability or even certainty of the existence of such a relationship than do the English. The consequences of such a slip if a cause-effect relation has not in fact been established could be significant for both the translator and the client, and avoiding such a problem would be to the advantage of both. Such problems could be avoided for example by providing translators with guidance in the form of examples.

4.2 Limitations and challenges

Although we feel there are considerable potential benefits to storing and consulting occurrences of terminological relations in termbases, it is important to recognise potential challenges. As noted above, any

approach to terminology management must be as efficient as possible. Time required to store and manage additional information must be offset by gains in time and/or in quality of the ultimate product. We believe that the benefits of including terminological relations in many cases will outweigh the modestly increased workload, and that (as is the case with translation memories) the gradual accumulation of information will ultimately form a useful resource. However, as noted above, each situation is different and the return on investment may vary depending on user needs and situation of use.

Making the storage of terminological relations as efficient as possible could require the development of a tool to accelerate and facilitate storage and annotation of occurrences, and a termbase structure that is adequate for storing the information and providing quick and multifaceted access depending on what the translator requires in any given search. Increasing flexibility in commercial tools is promising: further developments in searching and display options could make today's commercial tools even better adapted for handling this kind of information.

Increasingly, as the growing interest in exchange formats for translation memories and termbases as well as data-sharing initiatives such as the TM Marketplace and TAUS Data demonstrate, translators and clients are exchanging data of various kinds. The benefits of an individual's investment in storing terminological relations could then be multiplied by sharing this data.

Facilitating the sharing of information between users and exchange between termbases is also a relevant issue. Standards such as the TBX family in their default forms do not currently account for all of the types of relations and data (e.g. relation markers) explored here. At the present time, the sharing of relation information would require that users develop extensions of the core frameworks and agree on their use in order to exchange data.

5 Conclusions and future work

We believe that with this study we have highlighted key benefits of storing relation occurrences in translation-oriented terminology databases. In the process, we have highlighted the relevance of lexical relation markers for both identifying specific, useful information about terminological relations in texts and for expressing these relations clearly and precisely in writing and translation in specialised fields. Human language professionals can often easily interpret the relevance of relations based on these occurrences, a task that has proven extremely complex in even semi-automated approaches to relation extraction.

The variability and associations observed in the use of markers nevertheless demonstrates the relevance of making lists of markers

available for human users, to assist them in choosing precise and appropriate relation markers for use in specific texts and contexts and with specific terms, as well as in the translation of markers as required. The possibility of storing relation occurrences encountered in the course of corpus-based terminological research in a term base structure appears to be a promising avenue for future investigation.

Among the tasks in future work is the exploration of strategies for identifying the occurrences of terminological relations that are most relevant for users, and for storing the occurrences identified in terminology resources to make both the relations and their markers easily accessible and usable for the language professionals who may benefit from them.

It would also be beneficial to continue studying the usefulness of various types of terminological information and user reactions to its presentation by analysing users' reactions to the inclusion of annotated terminological relation occurrences in termbases.

Acknowledgements

The authors wish to thank: Trish Van Bolderen for her valuable contributions to previous phases of this project; the Canadian Breast Cancer Foundation, the Canadian Cancer Society, the Canadian Medical Association Journal, Health Canada, and the Hereditary Breast and Ovarian Cancer Foundation for their kind permission to analyse texts gathered from their web sites; and the University of Ottawa Faculty of Arts, Office of the Vice Rector and School of Translation and Interpretation, as well as the Social Sciences and Humanities Research Council of Canada for financial support for the project. They also wish to thank the anonymous reviewers of the article for their helpful suggestions, and Kara Warburton and Alan K. Melby for various helpful discussions about TBX standards.

Bibliography

- **Ahmad, Khurshid and Heather Fulford** (1992). "Knowledge processing: 4. Semantic relations and their use in elaborating terminology." *Computing Sciences Report CS-92-07*. Guildford: University of Surrey.
- **Barrière, Caroline** (2001). "Investigating the causal relation in informative texts." *Terminology* 7(2), 135–154.
- — (2002). "Hierarchical refinement and representation of the causal relation." *Terminology* 8(1), 91–111.
- **Borillo, Andrée** (1996). "Diversités des sources : La relation partie-tout et la structure [N1 à N2] en français." *Faits de langues* 7, 111–120.

- **Bowden, Paul Richard, Peter Halstead and Tony G. Rose** (1996). "Extracting conceptual knowledge from text using explicit relation markers." Nigel Shadbolt, Kieron O'Hara and Guus Schreiber (eds) (1996). *Advances in Knowledge Acquisition, Proceedings of the 9th European Knowledge Acquisition Workshop, EKAW'96*. New York/Berlin: Springer, 147–162.
- **Bowker, Lynne** (2011). "Off the record and on the fly: Examining the impact of corpora on terminographic practice in the context of translation." Alet Kruger, Kim Wallmach and Jeremy Munday (eds) (2011). *Corpus-based Translation Studies: Research and Applications*. London/New York: Continuum, 211–236.
- **Bowker, Lynne and Jennifer Pearson** (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. New York: Routledge.
- **Cabré, Maria Teresa, Jordi Morel and Carlos Tebé** (1996). "Las relaciones conceptuales de tipo causal: un caso práctico." *Actas del V Simposio Iberoamericano de terminologie RITerm* (Mexico City, 3–8 November 1996). <http://www.unilat.org/dtil/MEXICO/cabremt.html> (consulted 06.08.2004).
- — (2001). "Propuesta metodológica sobre cómo detectar las relaciones conceptuales en los textos a través de una experimentación sobre la relación causa-efecto." Maria Teresa Cabré and Judit Feliu (eds) (2001). *La terminología científico-técnica: Reconocimiento, análisis y extracción de información formal y semántica*. Barcelona: Institut universitari de lingüística aplicada, Universitat Pompeu Fabra, 165–170.
- **Clas, André** (1994). "Collocations et langues de spécialité." *Meta: journal des traducteurs* 39(4), 576–580.
- **Condamines, Anne** (2000). "Chez dans un corpus de sciences naturelles : un marqueur de relation meronymique?" *Cahiers de lexicologie* 77, 165–187.
- — (2002). "Corpus analysis and conceptual relation patterns." *Terminology* 8(1), 141–162.
- — (2008). "Taking genre into account when analysing conceptual relation patterns." *Corpora* 3(2), 115–140.
- **Condamines, Anne and Pascal Amsili** (1993). "Terminology between language and knowledge: an example of terminological knowledge base." Klaus-Dirk Schmitz (ed.) (1993). *Proceedings of Terminology and Knowledge Engineering, TKE'93*. Frankfurt: INDEKS-Verlag, 316–323.
- **Condamines, Anne and Josette Rebeyrolle** (2000). "Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthode." Jean Charlet Manuel Zacklad, Gilles Kassel and Didier Bourigault (eds) (2000). *Ingénierie des connaissances, évolutions récentes et nouveaux défis*. Paris: Eyrolles, 127–147.
- — (2001). "Searching for and identifying conceptual relationships via a corpus-based approach to a Terminological Knowledge Base (CKTB): Method and Results." Didier Bourigault, Christian Jacquemin and Marie-Claude L'Homme (eds) (2001). *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: John Benjamins, 127–148.
- **Dancette, Jeanne, Christophe Réthoré and Léon F. Wegnez** (1997). *Dictionnaire analytique de la distribution*. Montreal: Presses de l'Université de Montréal. <http://olst.ling.umontreal.ca/dad/> (consulted 07.10.2011).

- **de Villers, Marie-Eve** (2003). *Multidictionnaire de la langue française*. 4e édition. Montreal: Québec-Amérique.
- **Drouin, Patrick** (2011). *TermoStat* Web. http://olst.ling.umontreal.ca/~drouinp/termostat_web/index.php?lang=en_CA (consulted 24.09.2011).
- **Drouin, Patrick** (2003). "Term extraction using non-technical corpora as a point of leverage." *Terminology* 9(1), 99–115.
- **Dubuc, Robert** (2002). *Manuel pratique de terminologie*, 3e édition. Brossard: Linguatéc éditeur.
- **Feliu, Judit** (2004). *Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica*. PhD thesis. Universitat Pompeu Fabra.
- **Garcia, Danela** (1996). "COATIS, un outil d'aide à l'acquisition des connaissances causales exprimées dans les textes." *Actes du Colloque Linguistique et Informatique de Montréal, CLIM'96*. (Université de Montreal, 8–10 June 1996), 97–103.
- — (1997). "Structuration du lexique de la causalité et réalisation d'un outil d'aide au repérage de l'action dans les textes." Équipe de Recherche en Syntaxe et Sémantique (1997) *Actes des deuxièmes rencontres — Terminologie et Intelligence Artificielle, TIA '97* (Toulouse, France, 3–4 April 1997), 7–26.
- **Gillam, Lee, Mariam Tariq and Khurshid Ahmad** (2005). "Terminology and the construction of ontology." *Terminology* 11(1), 55–81.
- **Halskov, Jakob** (2007). *The semi-automatic expansion of existing terminological ontologies using knowledge patterns on the WWW – An implementation and evaluation*. PhD thesis. Copenhagen Business School.
- **Halskov, Jakob and Caroline Barrière** (2008). "Web-based extraction of semantic relation instances for terminology work." *Terminology* 14(1), 20–44.
- **Hearst, Marti** (1992). "Automatic acquisition of hyponyms from large text corpora." Christian Boitet (ed.) (1992). *Proceedings of COLING-92* (Nantes, France, 23–28 August 1992), 539–545.
- **Heid, Ulrich** (2001). "Collocations in Sublanguage Text: Extraction from Corpora." Sue Ellen Wright and Gerhard Budin (eds) (2001). *Handbook of Terminology Management*. Vol. 2. Amsterdam/Philadelphia: John Benjamins, 788–808.
- **Hennekens, Charles H. and Julie E. Buring** (1987). *Epidemiology in Medicine*. Sherry L. Mayrent (ed.). Boston/Toronto: Little, Brown and Co.
- **Iris, Madelyn A., Bonnie E. Litowitz and Martha W. Evens** (1988). "Problems of the part-whole relation." Martha W. Evens (ed.) (1988). *Relational Models of the Lexicon*. Cambridge, M.A.: Cambridge University Press, 261–288.
- **Jackiewicz, Agata** (1996). "L'expression lexicale de la relation d'ingrédience (partie-tout)." *Faits de langues* 7, 53–62.
- **Jacques, Marie-Paule and Nathalie Aussenac-Gilles** (2006). "Variabilité des performances des outils de TAL et genre textuel." *Traitement automatique des langues* 47(1), 11–32.

- **Jouis, Christophe** (1993). *Contribution à la conceptualisation et à la modélisation des connaissances à partir d'une analyse linguistique de textes. Réalisation d'un prototype : Le système Seek*. PhD thesis. École des hautes études en sciences sociales de Paris.
- — (1995). "SEEK: Un logiciel d'acquisition des connaissances utilisant un savoir linguistique sans employer de connaissances sur le monde externe." *Actes des Journées d'Acquisition de Connaissances du PRC-GDR-IA du CNRS*. (Grenoble, April 1995), 159–172.
- **Lee, David** (2001). "Genres, registers, text types, domains, and styles: Clarifying the concepts and navigating a path through the BNC jungle." *Language Learning and Technology* 5(3), 37–72.
- **L'Homme, Marie-Claude** (1997). "Méthode d'accès informatisé aux combinaisons lexicales en langue technique." *Meta: journal des traducteurs* 42(1), 15–23.
- — (2004). *La terminologie : principes et techniques*. Montreal: Presses de l'Université de Montréal.
- — (2011a). *Dictionnaire fondamental d'informatique et d'Internet (DiCoInfo)*. <http://olst.ling.umontreal.ca/cgi-bin/dicoinfo/search.cgi> (consulted 07.10.2011).
- — (2011b). *Dictionnaire fondamental de l'environnement (DiCoEnviro)*. <http://olst.ling.umontreal.ca/cgi-bin/dicoenviro/search-enviro.cgi?ui=en> (consulted 07.10.2011).
- **L'Homme, Marie-Claude and Elizabeth Marshman** (2006). "Extracting terminological relationships from specialized corpora." Lynne Bowker (ed.) (2006). *Lexicography, Terminology, Translation: Text-Based Studies in Honour of Ingrid Meyer*. Ottawa: University of Ottawa Press, 67–80.
- **Localization Industry Standards Association (LISA)** (2008). "Systems to manage terminology, knowledge, and content - TermBase eXchange (TBX)." http://www.ttt.org/oscarStandards/tbx/tbx_oscar.pdf (consulted 07.10.2011).
- **Localization Industry Standards Association (LISA), Terminology Special Interest Group (SIG)** (2009). "TBX-Basic." <http://www.ttt.org/oscarStandards/tbx/tbx-basic.html> (consulted 07.10.2011).
- **Malaisé, Véronique, Pierre Zweigenbaum and Bruno Bachimont** (2005). "Mining defining contexts to help structuring differential ontologies." *Terminology* 11(1), 21–53.
- **Marshman, Elizabeth** (2006). *Lexical Knowledge Patterns for Semi-automatic Extraction of Cause-effect and Association Relations from Medical Texts: A Comparative Study of English and French*. PhD thesis, Université de Montréal. http://www.ling.umontreal.ca/lhomme/docs/marshman_thesis.zip (consulted 03.10.2011).
- — (2007). "Towards strategies for processing relationships between multiple relation participants in knowledge patterns: An analysis in English and French." *Terminology* 13(1), 1–34.
- — (2008). "Expressions of uncertainty in candidate knowledge-rich contexts: A comparison in English and French specialized texts." *Terminology* 14(1), 124–151.

- **Marshman, Elizabeth and Marie-Claude L’Homme** (2006). “Disambiguating lexical markers of cause and effect using actantial structures and actant classes.” Heribert Picht (ed.) (2006). *Modern Approaches to Terminological Theories and Applications. Proceedings of the 15th European Symposium on Language for Special Purposes, LSP 2005*. New York: Peter Lang, 261-285.
- — (2008). “Portabilité des marqueurs de la relation causale : étude sur deux corpus spécialisés.” François Maniez et al. (eds) (2008). *Corpus et dictionnaires de langues de spécialité : Actes des Journées du CRTT*. Grenoble: Presses universitaires de Grenoble, 87-110.
- **Marshman, Elizabeth and Patricia Van Bolderen** (2008). “Interlinguistic variation and lexical knowledge patterns: Comparing data in English and French.” Bodil Nistrup Madsen and Hanne Erdman Thomsen (eds) (2008). *Managing Ontologies and Lexical Resources. Proceedings of the 8th International Conference on Terminology and Knowledge Engineering, TKE 2008*. (Copenhagen Business School, 19-20 August 2008), 263-278.
- — (2009). “Towards an integrated analysis of aligned texts: The CREATerminal approach.” Marie-Claude L’Homme and Amparo Alcina (eds) (2009). *Proceedings of Terminology and Lexical Semantics 2009*. (Montreal, June 2009), CD-ROM.
- **Marshman, Elizabeth and Sylvie Vandaele** (2010). “Metaphorical conceptualization of associations in medical texts: An analysis in English and French.” Walther von Hahn and Cristina Vertan (eds) (2010). *Fachsprachen in der weltweiten Kommunikation / Specialized Language in Global Communication* (Akten des XVI. Europäischen Fachsprachensymposiums, Hamburg 2007 / Proceedings of the XVIth European Symposium on Language for Special Purposes (LSP), Hamburg (Germany), August 2007. Frankfurt am Main: Peter Lang, 335-344.
- **Marshman, Elizabeth, Tricia Morgan and Ingrid Meyer** (2002). “French patterns for expressing concept relations.” *Terminology* 8(1), 1-29.
- **Marshman, Elizabeth, Marie-Claude L’Homme and Victoria Surtees** (2008a). “Portability of cause-effect relation markers across specialized domains and text genres: A comparative evaluation.” *Corpora* 3(2), 141-172.
- — (2008b). “Verbal markers of cause-effect relations across corpora.” Bodil Nistrup Madsen and Hanne Erdman Thomsen (eds) (2008). *Managing Ontologies and Lexical Resources. Proceedings of the 8th International Conference on Terminology and Knowledge Engineering, TKE 2008*. (Copenhagen Business School, 19-20 August 2008), 159-173.
- — (2009). “Marqueurs de la relation cause-effet: stabilité et variation dans des corpus de nature différente.” *Proceedings of the 8th International Conference on Terminology and Artificial Intelligence* (Toulouse, France, 18-20 November 2009). <http://www.irit.fr/TIA09/thekey/articles/lhomme-marshman-surtees.pdf> (consulted 18.06.2012).
- **Melby, Alan K.** (2008). “Translation-oriented terminology made simple.” *Tradumática* 6. <http://www.ttt.org/tbx/AKMtradumaArticle-publishedVersion.pdf> (consulted 07.10.2011).
- **Meyer, Ingrid** (2001). “Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework.” Didier Bourigault, Christian Jacquemin

and Marie-Claude L'Homme (eds) (2001). *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: John Benjamins, 279–302.

- **Meyer, Ingrid and Kristen Mackintosh** (1994). "Phraseme analysis and concept analysis: Exploring a symbiotic relationship in the specialized lexicon." Willy Martin *et al.* (eds) (1994). *Proceedings of Euralex '94*. Amsterdam: Vrije Universiteit, 339–348.
- — (1996a). "The corpus from a terminographer's viewpoint." *International Journal of Corpus Linguistics* 1(2), 257–285.
- — (1996b). "Refining the translator's concept analysis methods: How can phraseology help." *Terminology* 3(1), 1–26.
- **Meyer, Ingrid, Lynne Bowker and Karen Eck** (1992). "COGNITERM: An Experiment in Building a Terminological Knowledge Base." Hannu Tommola *et al.* (eds) (1992). *Proceedings of the Fifth Euralex International Congress* (Tampere, Finland, 4–9 August 1992), 159–172.
- **Meyer, Ingrid et al.** (1999). "Conceptual sampling for terminographical corpus analysis." Peter Sandrini (ed.) (1999). *Proceedings of Terminology and Knowledge Engineering TKE '99*. (Innsbruck, Austria, 23–27 August 1999), 256–267.
- **Morin, Emmanuel** (1999). "Acquisition de patrons lexico-syntaxiques caractéristiques d'une relation sémantique." *Traitement automatique des langues (TAL)* 40(1), 143–166.
- **Nazarenko, Adeline** (2000). *La cause et son expression en français*. Paris: Ophrys.
- **Nuopponen, Anita** (1994). "Causal relations in terminological knowledge representation." *Terminology Science and Research* 5(1), 36–44.
- — (2005). "Concept relations: An update of a concept relation classification." Bodil Nistrup Madsen and Hanne Erdman Thomsen (eds) (2005). *Terminology and Content Development: Proceedings of the 7th International Conference on Terminology and Knowledge Engineering, TKE'05*. (Copenhagen, 17–18 August 2005), 127–138.
- — (2010). "Methods of concept analysis – towards systematic concept analysis." *LSP Journal* 1(2). <http://rauli.cbs.dk/index.php/lspcog/article/view/3092/3275> (consulted 04.02.2012).
- — (2011). "Methods of concept analysis – tools for systematic concept analysis." *LSP Journal* 2(1). <http://rauli.cbs.dk/index.php/lspcog/article/view/3302/3500> (consulted 04.02.2012).
- **O'Brien, Sharon** (1998). "Practical Experience of Computer-Aided Translation Tools in the Software Localization Industry." Lynne Bowker *et al.* (eds) (1998). *Unity in Diversity? Current Trends in Translation Studies*, Manchester: St. Jerome Publishing, 115–122.
- **Otman, Gabriel** (1994). "Pourquoi parler de connaissances terminologiques et de bases de connaissances terminologiques." *La banque des mots* NS6, 5–27.
- — (1996). "Expression lexicale de la relation partie-tout: Le traitement automatique de la relation partie-tout en terminologie." *Faits de langues* 7, 43–52.
- **Pavel, Silvia and Diane Nolet** (2001). *Handbook of Terminology*. Ottawa: Public Works and Government Services Canada.

<http://www.btb.gc.ca/publications/documents/termino-eng.pdf> (consulted 01.10.2011).

- **Pearson, Jennifer** (1998). *Terms in Context*. Amsterdam/Philadelphia: John Benjamins.
- — (1999). "Comment accéder aux éléments définitoires dans les textes spécialisés?" *Terminologies nouvelles* 19, 21–28.
- **Rebeyrolle, Josette** (2000). *Forme et fonction de la définition en discours*. PhD thesis, Université de Toulouse II.
- **Roche, Christophe** (ed.) (2010). *Proceedings of Terminology and Ontology: Theories and Applications*. (Annecy, France, 3-4 June 2010). <http://www.porphyre.org/toth/proceedings> (consulted 07.10.2011).
- **Sager, Juan Carlos** (1990). *A Practical Guide to Terminology Processing*. Amsterdam/Philadelphia: John Benjamins.
- **Sambre, Paul and Cornelia Wermuth** (2010). "Instrumentality in cognitive concept modelling." Marcel Thelen and Frieda Steurs (eds) (2010). *Terminology in Everyday Life*. Amsterdam/Philadelphia: John Benjamins, 233-254.
- **Séguéla, Patrick** (1999). "Adaptation semi-automatique d'une base de marqueurs de relations sémantiques sur des corpus spécialisés." *Terminologies nouvelles* 19(1), 52–60.
- **Terminotix** (2010). LogiTerm 5. <http://www.terminotix.com> (consulted 01.10.2011).
- **Winston, Morton, Roger Chaffin and Douglas J. Herrmann** (1987). "A taxonomy of part-whole relations." *Cognitive Science* 11(4), 417–444.
- **Wright, Sue Ellen et al.** (2010). "TBX Glossary: A Crosswalk between Termbase and Lexbase Formats." Jennifer DeCamp (ed.) (2010). *Proceedings of the workshop 'Developing, Updating, and Coordinating Terminologies, Dictionaries, and Lexicons for Terminological Consistency' at AMTA 2010* (Denver, 31 October – 4 November 2010). http://amta2010.amtaweb.org/AMTA/papers/TBX-Glossary_2010-10-29.pdf (consulted 07.10.2011).

Websites

- "TAUS Data." www.tausdata.org (consulted 04.07.2012).
- "TM Marketplace." <http://www.tmmarketplace.com> (consulted 25.06.2012).
- "Visual DiCoInfo." <http://olst.ling.umontreal.ca/dicoinfo/visuel.php> (consulted 25.06.2012).

Biographies

Elizabeth Marshman has been an Assistant Professor at the University of Ottawa School of Translation and Interpretation (UO-STI) and a regular member of the Observatoire de linguistique Sens-Texte since 2007. Her research interests include computer-assisted terminology, language

technologies and the teaching of language technologies in translator education programs. She can be reached at elizabeth.marshman@uottawa.ca.



Julie L. Gariépy is currently a student at the UO-STI, conducting her M.A. research in Translation Studies with a focus on collaborative terminology and wikiterminology. She can be reached at jgari085@uottawa.ca.



Charissa Harms is currently a student at the UO-STI, conducting her M.A. research in Translation Studies with a focus on media representations of political narrative. She can be reached at charm100@uottawa.ca.



Notes

¹ As some exceptions we can mention the *Dictionnaire analytique de la distribution* (Dancette *et al.* 1997), the *Dictionnaire fondamental d'informatique et d'Internet* (*DiCoInfo*) (L'Homme (ed.) 2011a) and related projects including the *DiCoEnviro* (L'Homme (ed.) 2011b) and the Visual DiCoInfo.

² Observations of association are often precursors to concluding the existence of cause-effect relations. However, they are not sufficient to draw conclusions of a causal relationship: considerable and consistent evidence of association and a plausible mechanism for causation are required. For this reason, it is important to distinguish the two types of relations. More discussion of these relations from the perspective of corpus-based terminology can be found in Marshman (2006).

³ All translations in single quotation marks are our own.

⁴ Occurrences of relations that were incomplete in one or both of the languages or that in our estimation could not be reliably classified were set aside for the purposes of this study. As occasionally sentences containing occurrences of relations are repeated within or between documents and/or may have been identified using more than one candidate term, duplicate occurrences were removed for the purposes of this analysis. The final collection contained relation occurrences for 92 English terms.

⁵ This could also be achieved in some other tools such as terminology management systems, generic database management systems or office software, provided that this information has been stored in fields that can be processed using the available search, sorting and/or filtering options.