

Using Lexical Knowledge Patterns for Terminology Work in English and French: Some Important Differences

Elizabeth Marshman

Observatoire de linguistique Sens-Texte (OLST)

Université de Montréal, Montreal, Quebec

Many researchers (e.g., Ahmad and Fulford 1992; Meyer et al. 1999; Barrière 2001; Condamines and Rebeyrolle 2001; Meyer 2001; Condamines 2002; Marshman et al. 2002; Feliu 2004; Bodson 2005) have discussed using knowledge patterns — i.e., recurring linguistic structures that often indicate conceptual relations or attributes — to help terminologists identify specialized texts and contexts that may be useful for the purposes of conceptual analysis, knowledge structuring and terminological description. Many of these applications have focused on the identification of what were called by Meyer (2001) *knowledge-rich contexts* (KRCs), that is, text segments that provide at least one piece of information useful for such tasks.

Computer tools may use knowledge patterns to analyze corpora automatically and extract candidate KRCs that can then be presented to a human user for evaluation. One approach to candidate KRC extraction (e.g., Meyer et al. 1999; Feliu 2004) involves the search for terms or candidate terms of interest occurring with patterns indicating relations that may be useful for describing the concepts they denote. Such tools may not only assist terminologists and terminographers in quickly and easily identifying useful information in corpora — which can then be used in creating knowledge structures, delimiting and defining concepts, linking terms to the concepts they denote, and describing terms and their use — but may also help to provide a coherent, structured overview of the type and density of information about a concept in a corpus. Moreover, tools that can analyze corpora in two or more languages using similar approaches and classifications of candidate KRCs would be invaluable for assisting users in comparing knowledge structures as described in these languages, and in establishing interlinguistic equivalence between terms.

However, little is known about how potential differences in the ways relations and the elements they link are expressed in various languages may affect the development and productivity of such tools. This study looks at some of these factors in order to evaluate possibilities and strategies for developing tools for bilingual work.

The lexical knowledge patterns on which this work focuses are structures that can be represented using a model such as X [*marker*] Y , in which the variables X and Y represent textual manifestations of concepts (i.e., terms or other linguistic units), which are linked by a lexical marker of a conceptual relation — in this case a CAUSE-EFFECT or ASSOCIATION relation — that holds between the two concepts. The CAUSE-EFFECT relations discussed in this work may be expressed by English markers such as *produce, cause, participate in, increase, decrease, prevent* and *limit*, and French markers such as *causer, contribuer à, augmenter, diminuer, empêcher* and *limiter*. The ASSOCIATION relation — defined in this work as an observed co-occurrence of two factors, and which is often a precursor to the identification of CAUSE-EFFECT relations — may be indicated by such markers as *association with* and *link between... and* in English and *associé à* and *lien entre... et* in French.

This presentation will describe an analysis of a set of KRCs extracted from French and English corpora in the field of medicine (specifically heart disease and breast cancer). These KRCs expressed CAUSE-EFFECT and ASSOCIATION relations involving concepts denoted by a set of candidate terms (CTs) in each language. KRCs extracted using these CTs were analyzed manually to identify the lexical relation

markers present and the linguistic units they linked, which were then described and compared in order to evaluate potential differences in the characteristics of the knowledge patterns observed in the two samples. The data obtained in this analysis were then studied to determine whether any discrepancies observed could affect the performance of pattern-based tools for terminological research in a bilingual context.

The KRCs observed in the two corpora revealed some striking differences in the forms of both the markers and the linguistic units they linked. For example, a wider variety of markers was observed in the French sample, suggesting that in order to retrieve a comparable proportion of the available KRCs from a corpus, more markers may be required in this language. Conversely, the representation of the individual markers in the English sample appeared to present some additional complexities as compared to the French, suggesting that challenges in marker representation could contribute to difficulties in identifying KRCs in English. The form of linguistic units linked by markers also showed variation in the two samples, with phenomena such as anaphora and non-nominal forms also more prevalent in the French sample than in the English sample. These variations suggest that the use of previously identified, nominal candidate terms may allow less comprehensive retrieval of relation occurrences in French, and that alternative strategies may be required in order to fully exploit available contexts in this language.

This research has thus permitted the identification and evaluation of a number of issues that may affect pattern-based applications for researching specific terms and the concepts they denote for the purposes of bilingual terminology work. Moreover, the nature of the issues involved may suggest strategies for dealing with these questions in planning and developing KRC extraction tools.

- Ahmad, K. and H. Fulford. 1992. "Knowledge Processing: 4. Semantic Relations and their Use in Elaborating Terminology." (Computing Sciences Report CS-92-07). Guildford: University of Surrey.
- Barrière, C. 2001. "Investigating the Causal Relation in Informative Texts." *Terminology* 7(2): 135–154.
- Bodson, C. 2005. "Termes et relations sémantiques en corpus spécialisés : rapport entre patrons de relations sémantiques (PRS) et types sémantiques (TS)." Doctoral thesis, Université de Montréal.
- Condamines, A. 2002. "Corpus Analysis and Conceptual Relation Patterns." *Terminology* 8(1): 141–162.
- Condamines, A. and J. Rebeyrolle. 2001. "Searching for and identifying conceptual relationships via a corpus-based approach to a Terminological Knowledge Base (CKTB): Method and Results." In D. Bourigault, C. Jacquemin and M.-C. L'Homme, eds. *Recent Advances in Computational Terminology*. 127–148. Amsterdam/Philadelphia: John Benjamins.
- Feliu, J. 2004. "Relacions conceptuals i terminologia: anàlisi i proposta de detecció semiautomàtica." Doctoral thesis, Universitat Pompeu Fabra.
- Marshman, E., T. Morgan and I. Meyer. 2002. "French patterns for expressing concept relations." *Terminology* 8(1): 1–29.
- Meyer, I. 2001. "Extracting Knowledge-Rich Contexts for Terminography: A Conceptual and Methodological Framework." In D. Bourigault, C. Jacquemin and M.-C. L'Homme, eds. *Recent Advances in Computational Terminology*. 279–302. Amsterdam/Philadelphia: John Benjamins.
- Meyer, I., K. Mackintosh, C. Barrière and T. Morgan. 1999. "Conceptual Sampling for Terminographical Corpus Analysis." In *Proceedings of Terminology and Knowledge Engineering TKE '99*. 256–267. Innsbruck, Austria, August 1999.