Hitcheva, Dobromira

Canadian terminologists on terminological collocations

Dobromira Hitcheva

University of Architecture, Civil Engineering and Geodesy, Sofia, Bulgaria

Canadian Terminologists on Terminological Collocations

Abstract

Collocations in special languages have long been recognized as a major problem in technical translation and as relevant for producing scientific or technical texts. This paper traces the theoretical and methodological arguments some Canadian terminologists developed to analyze the phenomenon in question and to fill the gap of lacking co-occurrents of entries in their terminological collections. Earlier papers tried to distinguish between collocations and specialized lexical combinations (SLCs). In later papers the authors speak of "terminological collocations" and analyze terminological verb collocations using the verb actant structure based on the Explanatory and Combinatorial Lexicology (Mel'čuk et al. 1984, 1988, 1992, 1995, 1999). The latest research in the area is focused on presenting and organizing collocations in specialized dictionaries by grouping lexical functions in larger semantic classes, thus allowing users to access collocations onomasiologically, i.e. from the meaning to the collocate.

Résumé

Les collocations dans les langues spécialisées sont reconnues depuis longtemps comme un problème essentiel dans la traduction technique, pertinent pour la reproduction de textes scientifiques ou techniques. Le présent article retrace les arguments théoriques et méthodologiques que certains terminologues canadiens développent dans le but d'analyser le phénomène en question et pour combler la lacune du manque de collocations dans les entrées de leurs collections terminologiques. Tout au début des études on essaie de faire la distinction entre collocations et combinaisons lexicales spécialisées. Dans les ouvrages parus plus tard les auteurs parlent de "collocations terminologiques" et analysent des collocations terminologiques verbales en employant une structure verbale actantielle, basée sur la Lexicologie explicative et combinatoire (Mel'čuk et al. 1984, 1988, 1992, 1995, 1999). Les recherches récentes dans le domaine sont focalisées sur la présentation et l'organisation de collocations dans des dictionnaires specialisés par groupement de fonctions lexicales en classes sémantiques plus larges, permettant ainsi aux usagers d'accéder aux collocations onomassiologiquement, c.-à.-d. du sens au collocat.

Introduction

Collocations in both language for general purposes (LGP) and language for specific purposes (LSP) play a major role in understanding and producing general-language and special-language texts. Collocations in special languages have long been recognized as a major problem in technical translation. With the latest developments in computational and corpus linguistics, collocations have become an object of intensive research for lexicographic and terminographic

purposes. The followers of the lexico-semantic corpus-based approach to terminology (Observatoire de linguistique Sens-Texte/Observatory of Meaning-Text Linguistics at the Linguistics and Translation Department, University of Montreal) have made a particular contribution to the study of terminological collocations and to developing methods for organizing and presenting them in specialized dictionaries.

The aim of this paper is to briefly outline the development of the methodology for investigating terminological collocations/TCs by representatives of the Canadian School of Terminology. For achieving this aim, three stages in developing models and methods for studying terminological collocations by Canadian terminologists will be considered: first, a treatment of TCs as specialized lexical combinations; then, a model for analyzing terminological verb collocations/TVCs by the actant structure of the verb collocates; and finally, a method for presenting collocations in specialized dictionaries.

Terminological collocations as specialized lexical combinations

L'Homme and Bertrand distinguish between specialized lexical combinations/SLCs and collocations, defining the former as "word groups used in special languages" and the latter as "word groups used in general language". They were interested in specialized combinations comprising 2 lexemes where lexeme 1 prefers the company of lexeme 2 (e. g. *administer medicine* NOT * *give medicine* (cf. Laport and L'Homme 95–101), where *medicine* is the term/keyword, i.e. a unit with specific reference within a specialized subject field, while *administer* is the co-occurrent). The authors claim that SLCs are not prototypical collocations as the co-occurrents can combine with groups of semantically-related terms (e. g. the verb *pilot* can combine with *aircraft, airplane* and *seaplane*). However, they agree that SLCs and collocations share some similarities. Firstly, they are both conventional within a community (Meľčuk et al.): collocations within a given linguistic community, whereas SLCs within a group of specialists (discourse community). Moreover, learners of a language must acquire collocations since they are unpredictable, while learners of a special language must acquire SLCs (e. g. only specialists in the Steel Industry know that *steel* is *alloyed, tempered* or *pre-stressed*). These similarities suggest that that collocations and SLCs behave the same way and could be described using similar descriptive methods.

Studies have shown, though, that SLCs and collocations behave differently. The co-occurrents in SLCs can combine with small or large groups of terms. They are often conditioned by some sort of "definitional knowledge" held by the head/term. On the other hand, the co-occurrents in general language (GL) collocations are dictated by usage. Research results (L'Homme) demonstrate that verbs combine with several terms that share semantic properties (e. g. *install: operating system, package, Word, web surfer*) and that the regrouping of terms within semantic classes/SC (e. g. in "pieces of software" = SC) provides a basis for a description of verbal co-occurrents, i.e. co-occurrents combine with classes of terms. A study on GL collocations (Meľčuk and Wanner) illustrates that there exists a correlation between the meaning of a lexeme and its restricted co-occurrents: lexemes with common collocates share semantic features, though at the same time, the correlation is far from absolute. Therefore, grouping of keywords within

semantic classes is possible but cannot be systematically applied in GL. Furthermore, a typical GL collocation is semi-compositional, i.e. the keyword will combine uniquely with a given co-occurrent whose meaning is altered within that specific combination.

In a corpus-based study, L'Homme and Bertrand use texts related to two fields of knowledge (aeronautics and philosophy) to extract SLCs with the aim to measure the extent to which semantic classes could be observed. They investigate the proportion of lexical collocations (in which a co-occurrent with a given meaning combines with a single term) and conceptual collocations (in which the co-occurrent with a given meaning selects groups of terms). The study follows several steps:

1. Selection of preliminary keywords (e.g. *aircraft*)
2. Extraction of SLCs in which the preliminary keywords are used (e.g. *aircraft takes off, operate an aircraft,* etc.)
3. Selection of co-occurrents and new extraction of SLCs with the selected co-occurrents (e.g. *airplane takes off, rotorcraft takes off,* etc.)
4. Study of the extracted SLCs

The results show that in 86% of the SLCs studied, the co-occurrents could be found in other combinations, i.e. there are conceptual collocations in both fields of knowledge. The remaining 14% are SLCs in which a single term is found for a given co-occurrent, i.e. lexical collocations. Therefore, SLCs cannot be defined as true collocations, i.e. unique combinations.

Analyzing TVCs by the V actant structure

A model for specialized lexicography proposed by L'Homme ("Capturing"), in particular, for a cohesive representation of verbal senses and of verbal derivatives, defends the idea that verbs and their derivatives should be included in specialized dictionaries. Since we accept Sager's view that "it is common to consider all concepts to be represented by a single class, that of nouns…" (Sager 26), we will look at L'Homme's analysis of specialized verbs ("Capturing") from the point of view of the latter being collocators/collocates/co-occurrents in a TVC. The analysis of specialized verbs is a good starting point for discovering the lexical structure of a subject field (L'Homme, "Capturing"), including the identification of TCs. We are interested in the model to the extent of which it can account for the overall semantic structure of a TVC.

Verbs are predicative lexical units/PLUs associated with actants which serve to describe their semantics. These actants are usually represented by terms in noun form. The principles of Explanatory and Combinatorial Lexicology/ECL (Meľčuk, "Lexical") are used to describe the PLUs extracted from large textual corpora. For example, the definition of *install* in a computer science context must include three variables (semantic actants): a person X installs a program Y on a permanent storage device Z (cited after L'Homme, "Capturing" with slight changes). The actants/terms in the vicinity of *install* are semantically related, i.e. they share semantic components: ex. user/programmer, etc. (*animates*) INSTALL driver/routine/program/Word, etc. (*software components*) ON laptop/hard disc, etc. (*storage devices*). Therefore, accounting

for actants in terms of semantic classes seems to be very productive: SCs can be used to specify variables (X, Y and Z). SCs can also be used to capture/identify sets of terms which share semantic components, thus helping to identify TVCs in large text corpora.

Presenting TCs in specialized dictionaries

A novel method is proposed for organizing and presenting collocations in a specialized dictionary of computing and the Internet by Jousse et al. Their aim is to design an access method that would allow users to obtain answers to questions such as:

1. Which verbs express the idea of "using" a dialog box? – Answer: enable/display/open a dialog box.
2. Which verbs express the typical activities carried out by a programmer? – Answer: the programmer writes/debugs/develops programs.

The DiCoInfo, Dictionnaire fondamental de l'informatique et de l'Internet, is an online dictionary containing over 1000 articles in French and about 700 in English (the Spanish version is still under development). Work started in 2009 to provide a more adaptive and user-oriented access to data, which later led to the development of automatic access to the translation of collocations (L'Homme and Leroyer). Collocations sharing identical semantic and syntactic properties were linked up by means of encoding the language functions/LFs (L'Homme, Leroyer and Robichaud). The next step was to design and implement the onomasiological access to collocations (the user knows the meaning of a phraseological unit but is searching for the appropriate collocates).

Lexical relationships/LRs are an important data category in the DiCoInfo. Each entry contains a list of lexical units sharing with the head word a paradigmatic or syntagmatic relationship (synonymy, antonymy, syntactic derivation, collocates). Two systems of explaining LRs are used: 1) LFs (Meľčuk et al., *Dictionnaire*) and 2) a less formal and language-dependent explanation, which is also more transparent for users (Polguère). An example of collocations explained in the database of the dictionary (in the online version LRs are listed in a table) can be seen below:

Key word	Collocations	Lexical function	Explanation
program	quit a ~	FinReal1 (LF code)	The user stops using a p.

The collocations are presented in a section called "Combinations" but are long and difficult to read in some entries; e. g. approximately 100 collocates are listed for *file*.

A need arose to create a Model for grouping and browsing collocations. The authors solved the problem by grouping collocations in transparent classes. First, LFs were grouped into more generic SCs to allow users to access collocations onomasiologically (from the meaning to the collocate). Then, LRs were analyzed and dominant specific classes found (e.g. in Computing

many collocates express the idea of USE and MAKE STH WORK). Finally, all LFs encoding a typical USE were grouped into intermediate classes. The classification of the collocations of *file* illustrates these steps:

Generic class: USE/NOT USE

— to prepare for use/operation (intermediate class): *to install a file, to search for a file* (terminal class)
— to start to use: *to start, to open a file*
— to use/make sth work: *to process, to edit a file*
— to stop using/working: *to close a file*

Next, they organized the classes hierarchically: a root, the generic class, the intermediate classes, and the terminal classes corresponding to the LFs names.

Browsing collocations in the dictionary goes through the following steps:

1. The class hierarchy is loaded as an additional data structure along with the dictionary files, thus providing easy access and a possibility to modify the organization of the classes without having to edit the entries or the programs.
2. Entries are displayed within an outline/tree view section (see above) that holds the collocations according to the hierarchy.
3. This new section is first presented as an ordinary hyperlink.
4. By clicking on this hyperlink, users open the hierarchy and may select different branches/nodes according to the class names presented and their search needs.
5. Ultimately, browsing paths reach the terminal classes and short tables are presented with the usual info about collocations in the respective dictionary files.

Conclusion

The methodology for investigating TCs, proposed by representatives of the Canadian School of Terminology, is original, comprehensive and applicable in studying different aspects of TCs. The claim that TCs, called in earlier papers SLCs, behave differently from typical GL collocations because the keyword in the latter will combine uniquely with a given co-occurrent, is debatable. We have proved (Alexiev and Hitcheva) that among both TCs and GL collocations it is possible to identify low-valence bases/keywords and high-valence bases: e.g. the term 'pile' in construction collocates predominantly with the verb 'drive' whereas the term 'steel' collocates with a large number of verbs such as 'melt', 'cast', 'roll', etc. The same applies to general collocations with a low-valence base (e.g. 'take precaution') and a high valence base (e.g. 'give/offer/provide/reject/etc. advice'). The idea of accounting for actants in terms of semantic classes seems to be very productive for investigating TVCs, especially in identifying TVCs in large text corpora. The model proposed for grouping TCs into more general SCs in specialized dictionaries is a considerable contribution to terminology processing in general and terminographic practice in particular.

Variations sur la Communauté :
l'espace canadien

Works cited

Alexiev, Boyan, and Dobromira Hitcheva. "Terminological Collocations as Knowledge and Translation Items." *Proceedings of the First International Translation Studies and Terminology Conference Turkey.* Kirrikale, Turkey: Vizyon Ofset Matbaa, Kirrikkale University, 2012. Print.

Jousse, Anne-Laure, et al. "Presenting Collocates in a Dictionary of Computing and the Internet According to User Needs." *Proceedings of the Fifth International Conference on Meaning–Text Theory Barcelona.* Barcelona: Universitat Pompeu Fabra, 2011. 134–144. Print.

Laporte, Isabelle, and Marie-Claude L'Homme. "Recensement et consignation des combinaisons lexicales spécialisées: Exemple d'application dans le domaine de la pharmacologie cardiovasculaire." *Terminologies nouvelles* 16 (1997): 95–101. Print.

L'Homme, Marie-Claude. "Capturing the Lexical Structure in Special Subject Fields with Verbs and Verbal Derivatives. A Model for Specialized Lexicography." *International Journal of Lexicography* 16.4 (2003): 403–422. Print.

----------. "Définition du statut du verbe en langue de spécialité et sa description lexicographique." *Cahiers de lexicologie* 73.2 (1998): 125–148. Print.

----------, et al. "Advanced Encoding for Multilingual Access in a Terminological Database – A Matter of Balance." *Terminology and Knowledge Engineering Conference Dublin 2010. Presenting Terminology and Knowledge Engineering Resources Online: Models and Challenges*, 12–13 Aug 2010. Dublin: Fiontar, Dublin City University, 2010. Print.

----------, and Claudine Bertrand. "Specialized Lexical Combinations: Should They Be Described as Collocations or in Terms of Selectional Resrictions?" *Proceedings of the Ninth Euralex International Congress 2000*. Stuttgart, Germany: University of Stuttgart, 2000. 497–506. Print.

----------, and Patrick Leroyer. "Combining the Semantics of Collocations with Situation- driven Search Paths in Specialized Dictionaries." *Terminology* 15.2 (2009): 258–283. Print.

Meľčuk, Igor. "Collocations and Lexical Functions." *Phraseology: Theory, Analysis, and Applications.* Ed. A. P. Cowie. Oxford: Clarendon Press, 1998. 23–53. Print.

----------. "Functions: A Tool for the Description of Lexical Relations in a Lexicon." *Lexical Functions in Lexicography and Natural Language Processing.* Ed. Leo Wanner. Amsterdam: John Benjamins, 1996. 37–102. Print.

----------, et al. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques 1–IV*. Montréal: Les Presses de l'Université de Montréal, 1984–1999. Print.

----------, and Leo Wanner. "Lexical Functions and Lexical Inheritance for Emotional Lexemes in German." *Lexical Functions in Lexicography and Natural Language Processing.* Ed. Leo Wanner. Amsterdam: John Benjamins, 1996. 207–277. Print.

Polguère, Alain. "Collocations et fonctions lexicales: pour un modèle d'apprentissage." *Les Collocations. Analyse et traitement*. Eds. Grossmann Francis, and Agnès Tutin. Amsterdam: De Werelt, 2003. 117–133. Print.

Sager, Juan C. *A Practical Course in Terminology Processing.* Amsterdam/Philadelphia: John Benjamins, 1990. Print.