Semantic approach to Phraseological Patterns in Karstology

Larisa Grčić Simeunović, Paula de Santiago
University of Zadar, King's College London
E-mail: larisa.grcic@gmail.com, pauladesantiago@gmail.com

Abstract

As previous research about collocational patterning (Stubbs 1996, 2001, Flowerdew 2009) has demonstrated, words are restricted in their collocational behaviour. This article aims to illustrate how specialised phraseology reveals restrictional meaning of terminological units as well as patterned syntagmatic relations between terms or terms and other lexical units as a result of a semantic analysis. Two kinds of paradigmatic sets were established: conceptual classes formed by term bases, and semantic sets formed by adjective collocates. Their distributional behaviour is described in order to identify the most recurrent phraseological patterns. Authors point out the importance of combinatorial information for determining specialized meaning and examine the possibility of offering additional information to the terminological definition.
Keywords: collocation, specialized corpus, semantics, selectional preference, terminography

1 Semantics in phraseology

Previous studies have described word combinations in specialized languages, finding out that two lexeme combinations likely depend on semantics apart from usage and that phraseology can illustrate the conventions of specialized communication within a linguistic community (Meyer & Mackintosh 1996; Mel'čuk & Wanner 1996; L'Homme 1998; L'Homme & Bertrand 2000; Martin 1992, 2008; Hanks 2013). According to Hanks (2013: 3) "discovering conventions requires searching for regular patterns of word use shared by speakers and writers". However, a collocation is more than a surface co-occurrence pattern; it also provides a representation of word meaning. Therefore, following Firth's idea (1957: 11), we are able to know the meaning of a word, or more information about the meaning of a word, by the company it keeps. Our interest in collocations goes beyond their conventional nature, we are interested in delving into the semantics of specialized concepts through their most relevant attributes in a given domain.

Following Martin's study (2008: 56), we agree on defining a collocation as a combination of two concepts or conceptual semantic frames that are in dependency relation, one (the collocate) modifying the other (the head). This means that there is a dominant frame and a dependent one clarifying a specific aspect of the dominant frame. Considering the so called dependent frames or collocates essential for the description of specialized concepts, this study finds support in works where collocates have been organized into semantic types (Pustejovsky et al. 1993), also called semantic classes (L'Homme 1998), classes of objects (Gross & Mathieu Colas 2001; Chodkiewicz & Gross 2005) or lexical sets (Hanks 2013) for the purpose of finding definitional criteria of terms.

The concept of semantic classes can be related to the Stubbs' (2001: 449) interpretation of semantic preference defined as "a lexical set of frequently occurring collocates, which share a semantic feature". In this line, Mechura (2010: 1) suggests that terms can be described according to their selectional preferences defined as "the tendencies of words to co-occur with other words that belong to certain semantic types". Introducing semantic classes in specialized language aims not

only to organize collocational profiles of terms but also to point out the inherent aspects of specialized meaning. According to Lauder (2010) information derived from semantic preference "is not restricted to lexical or dictionary meaning but includes pragmatic or encyclopaedic information as well which would not be predictable from a dictionary definition". This paper seeks to find revealing information about the meaning of terms within phraseological combinations in order to support or complement terminological definitions.

2 Methodology

This study is based on a 913.168 tokens monolingual specialized corpus. Two subcorpora were analysed; a scientific one, consisting of PhD theses and research articles, and a didactic one consisting of student manuals and course books. It covers the field of karstology, dealing with karst groundwater systems (hydrogeology) and karst landforms (geomorphology), both above and below the surface.

2.1 Data selection: selecting key words

In order to identify lexical collocations in our corpus, we selected key words within the field of karstology and used them as term bases, thus the starting point for the identification of collocations. A list of 50 most frequent and topic-relevant nouns has been extracted from the English corpus by using the function Wordlist from WordSmith Tools 6.0 (WST 6.0). For this purpose, a general stop list and a specialized lemma list have been used. Finally, the relevance of terms has been confirmed by contrasting them against entries in specialized glossaries. A total of 36 candidate terms matched the entries in these terminological sources, and are considered in this study. The terms are:
karst, cave, flow, limestone, groundwater, rock, aquifer, polje, spring, doline, river, fracture, sediment, stream, discharge, conduit, lake, recharge, dissolution, formation, passage, uvala, calcite, basin, precipitation, sinkhole, joint, porosity, permeability, cavity, infiltration, karstification, layer, fault, corrosion, depression.

2.2 Grouping the selected key words in conceptual classes

As stated in the introduction, we consider important to group the terms in conceptual classes. Scholars such as Pustejovsky et al. (1993: 333) state that by classifying an element into a particular category, many aspects of its semantic structure and hence, its syntactic behaviour, can be generated. Table 1 represents the classification of our key terms in conceptual classes.

Conceptual	Terms
LANDFORMS	cave, polje, doline, uvala, sinkhole, cavity, passage
SEDIMENTS	karst, limestone, rock, sediment, calcite, fracture, fault
HYDROLOGY	flow, aquifer, spring, groundwater, river, basin, stream
PROCESSES	dissolution, precipitation, infiltration, karstification

Table 1. Classification of key terms in conceptual classes.

2.3 Identifying statically significant collocates

All the term bases have been used to extract ADJ + NOUN (term) syntagmatic combinations with a minimum frequency of 3 occurrences appearing in at least 2 different texts. For the selection process, word sketches provided by the tool Sketch Engine were used. This step also allowed us to compare combinatorial potential of terms from the same conceptual class and to point to the different combinatorics between members of several conceptual classes.

2.4 Sorting the list of collocates into semantic classes

Not only terms were organized in semantic classes, but collocates as well. Research conducted by Gross and his team from the LLI group demonstrated the importance of differentiating collocates (i.e. arguments) in classes of objects for the purpose of distinguishing polysemic term bases (i.e. predicates). Gross's lexico-grammar approach described predicate behaviour in the domains of law (Chodkiewicz & Gross 2005) and medicine (Gross & Mathieu Colas 2001) by sorting its arguments within semantic classes of objects.
In this study, 11 knowledge sets generating collocations were detected: location (e.g. subterranean river subsurface karst), shape (e.g. bedded limestone), manner (e.g. steady flow, turbulent flow), age (old cave), size (e.g. enlarged conduit, extensive cave), temperature (e.g. cold water, thermal spring), direction (e.g. horizontal flow, horizontal passage), time (e.g. rapid infiltration, temporary spring), touch (e.g. solid limestone), quality (e.g. pure limestone, impermeable rock) and state (e.g. dry cave, flooded cave).

2.5 Discovering the combinatorial behaviour of phraseological patterns

After selecting two kinds of paradigmatic sets, we analysed their distributional behaviour. This allowed us to describe semantic preferences of terms and establish correlations between conceptual classes of terms and semantic classes of collocates in this particular domain. The semantic values of collocates were interpreted as representing different perspectives of conceptual categories in karst domain.

3 Results

178 collocations have been found. Some semantic classes of terms have shown to be more productive than others. As shown in the chart 1, the heads of 60 (out of 178) collocations denote landforms, 57 denote hydrology, 44 denote sediments and 17 denote processes.

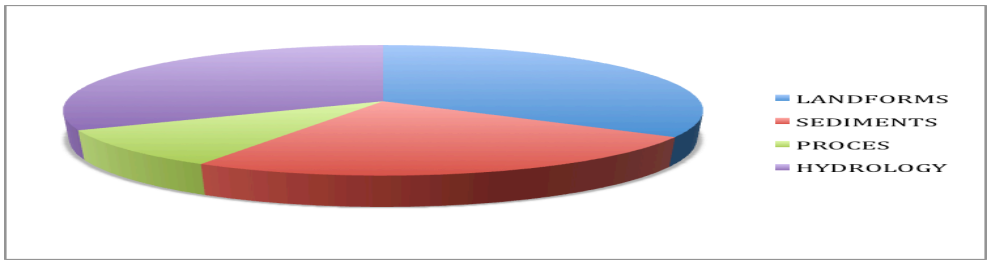

Chart 1. Distribution of term bases in conceptual classes.

The second chart represents the distribution of the 178 collocations according to the semantic classes of collocates modifying nominal terms:

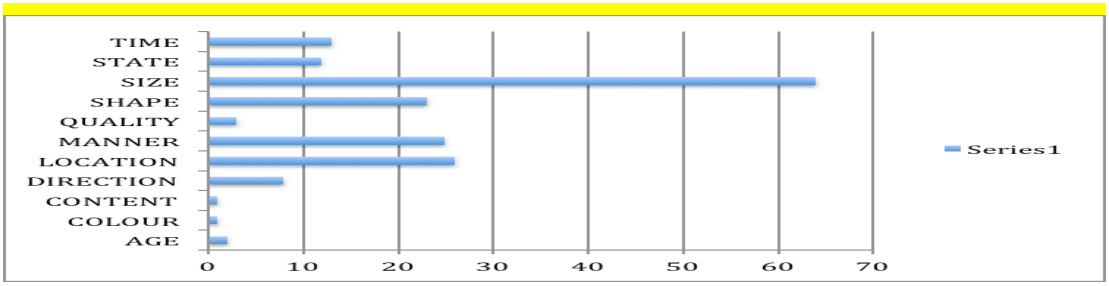

Chart 2. Distribution of collocates in semantic sets.

According to the results, the most frequent collocates express size (e.g. small, major, matrix, wide, single) (64), location (26) (e.g. surface, subterranean, horizontal, vertical, regional), manner (25) (e.g. rapid, direct, intensive, abundant) and shape (23) (e.g. shallow, deep, circular spherical) while semantic sets of time, state, age and colour occur in fewer occasions.

The next step consisted in analysing the combinatorial potential of conceptual classes of term bases and semantic sets of collocates. The results are shown in Table 2.

	Proces	Sediment	Hydrology	Landform	Total
AGE			1	1	2
COLOUR		1			1
CONTENT		1			1
DIRECTION		3	3	2	8
LOCATION	3	5	10	8	26
MANNER	10	6	8	1	25
QUALITY		1		2	3
SHAPE		5	4	14	23
SIZE		14	22	28	64
STATE		8		4	12
TIME	4		9		13
Grand Total	**17**	**44**	**57**	**60**	**178**

Table 2. Combinatorics between conceptual classes and semantic sets.

The results have shown that concepts denoting processes are mostly determined by adjective collocates denoting manner such as *mixture corrosion*, *concentrated infiltration*, etc.; concepts

denoting hydrology combine mostly with collocates expressing size (e.g. single conduit, large basin, etc.) and location (e.g. subsurface conduit, underground conduit, etc.); concepts denoting sediments are determined by collocates designating size (e.g. fault, discrete fracture, enlarged fracture, etc.) and state (e.g. impervious layer, impermeable rock, permeable rock, etc.); and concepts denoting landforms are determined by collocates denoting size (e.g. high passage, large passage, etc.) and shape (e.g. deep cave, horizontal cave, shallow cave, etc.).

Finally, the semantic analysis has led us to describe selectional restrictions of nominal term bases. Even though concepts from the karstology domain are characterised by many different adjectives, we found that some regularities can be established: terms that belong to particular conceptual classes co-occur with adjectives from particular semantic sets. This kind of combinatorial potential is in line with our preliminary hypothesis that semantic classification of collocates can reveal relevant and distinctive characteristics of certain groups of concepts.

The analysis revealed that each conceptual class is determined by a different set of relevant characteristics. For example, *karst landforms* are usually modified by collocates denoting shape, location, state and size. Table 3 represents the overall distribution of semantic sets combined with terms from the conceptual class *landforms*.

	direction	location	manner	quality	shape	size	state	Total
cave		1	1		4	7	2	15
cavity		3			2	2	1	8
depression		1			4	4		9
doline					2	4		6
formation		1		2				3
passage	2	2				3		7
polje					2	2	1	5
sinkhole						3		3
uvala						3		3
Total	**2**	**8**	**1**	**2**	**14**	**28**	**4**	**59**

Table 3. Distribution of semantic sets with conceptual class “landforms”.

The comparison of collocations with terms belonging to another conceptual class revealed that the most frequent co-occurrents with terms denoting “processes” belong to semantic sets of manner (e.g. rapid infiltration, direct infiltration, intensive precipitation, abundant precipitation), time (e.g. early karstification, annual precipitation, hourly precipitation, daily precipitation, short term precipitation) and location (e.g. lateral corrosion, epigene karstification, local precipitation). These differences encourage the semantic classification of multiword units as they reveal stereotyped phraseology combinations. They also confirm a close relation between the key term and its modifiers as each term/concept has its own selectional preferences. The results also confirmed that collocates tend to show selectional preferences as well. While adjectives expressing the attribute of size are combined with three different conceptual classes: landforms (28), hydrology concepts (22) and sediments concepts (14), adjectives denoting shape have shown strong preferences for landform concepts (14 examples). This semantic set appears only 5 times with sediments and 4 times with hydrology concepts. This kind of selectional preferences points to binary attractions described by Almela (2011).

The results of the analysis revealed the relevance of the role of adjective collocates in specialized

communication. The presence of lexical preferences allowed us to propose the use of adjective collocates as key words for searching and predicting of new syntagmatic combinations. This kind of intercollocability was described by Stubbs (2001) who analysed collocations in order to describe mutual prediction of word networks.

4 Terminographic implications

The results of the analysis encouraged us to propose the introduction of combinatorial information inside a term record by grouping collocations in semantic sets and selecting the most relevant characteristics for the description of conceptual meaning. This kind of approach confirms the notion of "concept bound" collocations introduced by Martin (1992). "According to the author, modifying concepts (i.e. cooccurrents) are often conditioned by some sort of "definitional knowledge" held by the head (i.e. terms) and are not strictly dictated by usage" (L'Homme & Bertrand 2000).

On the basis of the information extracted from the corpus, we have developed some term record samples. The aim is to show how the methodology and the data analysis can be applied to glossary making. Each record sample includes several categories: lexical category, definition[4], context from the cited corpus, collocates organized in semantic sets and the conceptual class to which the term belongs. An example of a term record is shown in Table 4:

Cave	
Lexical category	Noun
Definition	Natural underground chamber in a hillside or cliff
Context	Recent diving explorations have found large openings in the submerged walls of the doline and cave entrances at its floor.
Collocations	Shape + cave (e.g. *deep cave, horizontal cave, shallow cave, vertical cave*) Size + cave (e.g. *big cave, extensive cave, large cave, long cave, major cave, single cave, small cave*) Location + cave (e.g. *underground cave*) Age + cave (e.g. *old cave*) State + cave (e.g. *dry cave, unroofed cave*) Manner + cave (e.g. *active cave*)
Knowledge set(s)	Landform

Table 4. Term record sample of the term *cave*.

5 Conclusion

Regularities on the combinations of lexical units in karstology have been found. The interesting fact we discovered was that the terms that belong to particular conceptual classes combine with adjectives from particular semantic sets.

The model of semantic classes allowed us to capture the relevant information about the conceptual structure and to put emphasis on relevant attributes for each concept within a class. Collocations revealed the nature of concepts, by confirming that some concepts are described according to their location while others are distinguished by their size, shape or age. This kind of knowledge modelling is appropriate for representing the multidimensionality of meaning of a concept in different contexts. While terminological definition enumerates only the concept's distinctive

characteristics, phraseology reveals other perspectives of the concept and gives insight into significant aspect of the specialised field. We assume that attributes change according to the domains and therefore our future research will concentrate on comparison of relevant attributes in different domains and different languages.

This kind of analysis allowed us not only to identify normal combinatorial patterns but also to predict other specialized collocations. As multiword units seem infinite, this type of information helped us find clues for discovering what kind of combinatorial patterns are conventional, and which combinations are possible or probable.

References

Almela, M., Cantos, P. & Sánchez, A. (2011) Towards a Dynamic Combinatorial Dictionary: a Proposal for Introducing Interactions between Collocations in an Electronic Dictionary of English Word Combinations. In *Proceedings of the Electronic lexicography Conference, eLEX2011*, *10-12 November 2011.* Institute for Applied Slovene studies, Slovenia, pp. 1-11.

Chodkiewicz, C., Gross, G. (2005) La description de la langue du droit au moyen des classes d'objets. In J.C. Gémar, N. Kasirer (eds.) *Jurilinguistique*. Bruxelles: Bruylant et Montréal, pp. 23-42.

Firth, J.R. (1957) Modes of meaning. In *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.

Flowerdew, J. (2009) Corpora in Language Teaching. In M. H. Long, C. J. Doughty (eds.) *The Handbook of Language Teaching*. Oxford: Wiley-Blackwell, pp. 327-350.

Gross, G., Mathieu-Colas, M. (2001) Description de la langue de la medicine. In *Meta* 46.1, pp. 68-81.

Hanks, P. (ed.) (2013) *Lexical Analysis. Norms and exploitations*. The MIT Press: Cambridge.

Lauder, A. (2010) Collocation, Semantic Preference and Translation: Semantic Preference as a Reference Source for Translation. Accessed at: English.um.edu.my/anuvaada/PAPERS/LAUDER.pdf [04/03/2015].

L'Homme, M.C. (1998) Caractérisation des combinaisons lexicales spécialisées par rapport aux collocations de langue générale. In Th. Fontenelle et al. (eds.) Actes du congrès *Proceedings Euralex ´*98, Liège: Université de Liège, pp. 513-522.

L'Homme, M.C., Bertrand, C. (2000) Specialized Lexical Combinations: Should they be described as Collocations or in Terms of Selectional Restrictions? In *Proceedings of the Ninth Euralex International Congress*. Stuttgart: Stuttgart University, pp. 497-506.

Martin, W. (1992) Remarks on Collocations in Sublanguages. In *Terminologie et traduction*, 2.3, pp. 157-164.

Martin, W. (2008) A unified approach to semantic frames and collocational patterns. In S. Granger, F. Meunier (eds.) *Phraseology. An Interdisciplinary Perspective*. Amsterdam/Philadelphia: John Benjamins, pp. 51-65.

Mechura, M.B. (2010) What WordNet does not know about selectional preferences. In A. Dykstra, T. Schoonheim (eds.) *Proceedings of the 14th Euralex International Congress*, Ljouwert/Leeuwarden: Fryske Akademy, pp. 431-436.

Mel'čuk, I.A., Wanner, L. (1996) Lexical Functions and Lexical Inheritance for Emotion Lexemes in German. In E. Wanner (ed.) *Lexical Functions in Lexicography and Natural Language*

*Processin*g. Amsterdam: Benjamins Academic Publishers, pp. 209-278.
Meyer, I., Mackintosh, K. (1996) Refining the Terminographer's Concept Analysis Methods: How Can Phraseology Help ? In *Terminology* 3.1, pp. 1-26.
Pustejovsky, J., Anick, P. & Bergler, S. (1993) *Lexical Semantic Techniques for Corpus Analysis.* In Computational linguistics 19.2, pp. 331-358.
Stubbs, M. (1996) *Text and Corpus Analysis*. Oxford: Blackwell.
Stubbs, M. (2001) Words and Phrases. Corpus Studies of Lexical Semantics. Oxford: Blackwell.
Thoiron, P., Arnaud, P., Bejoint, H. & Boisson, C.P. (1996) Notion d' "archi-concept" et denomination. In *Meta* 41.4, pp. 512-524.