A Constructional Approach to Terminological Phrasemes

Silvia Montero Martínez
Universidad de Granada

Specialized discourse shows regularities in the lexical and syntactic patterning of terminological units. This fact, evidenced by corpus-based analysis, has spurred a number of studies on polilexical terminological units. In spite of the available linguistic data, however, the systematic management of these units in specialized lexicography is still lacking. Apart from a few exceptions, terminological products, especially dictionaries, are inconsistent with their treatment of these units. Such arbitrary approaches are worthless within the context of the newer terminological knowledge bases. In this paper, we describe how the Lexical Grammar Model can offer an in-depth, principled description of such units. Meaning and grammar are seen as interdependent and complementary layers. So, the basic unit of grammar is a form-meaning pairing or construction that can be described as a conventionalized combination of form and meaning. In this vein, the lexical profile of a specialized concept is composed of constructions, which reflect its collocational patterns both at a lexical and a syntactic level. Thus, we use the umbrella term terminological phraseme (Meyer and Mackintosh 1994) to include entrenched, conventional combinations of linguistic units in the form of complex nominals and predicate-argument structures. These units are conceived as constructions codifying conceptual, experiential and syntactic information concerning the lexical concepts of a cognitive frame. Consequently, the frame is the element which constrains the potential relations holding between the lexical concepts, and the construals that the frame allows are only a subset of the construals allowed by the argument-taking heads. The basic qualia structure and the domain-specific relations account for such combinations and for the inheritance phenomenon. In sum, we present a theoretical and methodological approach that accounts for the lexical profiles of concepts in a consistent way, including the description of conceptual relations as well as the terms' combinatorial potential.

1. Introduction

Specialized discourse shows, to a great extent, regularities in the lexical and syntactic patterning of terminological units (López Rodríguez 2000, Heid 2001). This fact, evidenced by corpus-based analysis, has spurred a number of studies on polilexical terminological units. Specifically, it has been shown that compounds (Sager 1997, Picht 1991), taken as nominal combinations of at least two words; lexical or tight collocations (Martin 1992, Heid 2001); and conceptual collocations (Martin 1992, L'Homme 2000) play a key role in specialized texts. All of them are regarded as conventional linguistic units, arising from language-specific communities and reflecting expert meaning configurations.

However, in spite of the available linguistic data, the systematic management of these units in specialized lexicography is still lacking. Apart from a few exceptions (Cohen 1986, Lainé 1993), terminological products, especially dictionaries, are inconsistent in their treatment of these units. Compounds, usually restricted to two-word combinations of N+N and Adj+N in English, are listed as entries or sub-entries, often based on *ad-hoc* criteria. The description given, apart from a list of language equivalents, is generally restricted to the definition and the grammatical category, leaving aside any pragmatic and relational information. A more depressing scenario can be found regarding collocations. It is still rare to find lexical repositories that contain N+V, V+N or Adv+Adj units. When listed, information going beyond the definition and the equivalents is also missing.

Such arbitrary approaches are worthless within the context of the newer terminological knowledge bases (TKBs). In this paper, we describe how the Lexical Grammar Model (Faber

and Mairal 1999; Mairal fc.), an extension of the Functional Lexematic Model (Martín Mingorance 1990), can offer an in-depth, principled description of such units.

2. The PUERTOTERM knowledge base

The PUERTOTERM knowledge base is a dynamic, conceptually-rich resource on coastal engineering. This repository results from the implementation of *frame-based terminography* (Faber et al. 2005, Faber et al. 2006), partially based on Fillmore's Frame Semantics (Fillmore et al. 2003) and a modified version of the FrameNet methodology (Ruppenhofer et al. 2006). Through the analysis of corpus data, our research team created the COASTAL ENGINEERING EVENT (CEE), a frame-based data model (Faber et al. 2005). This relational schema encompasses the objects, processes, and relations typical of coastal management, as seen in Figure 1.

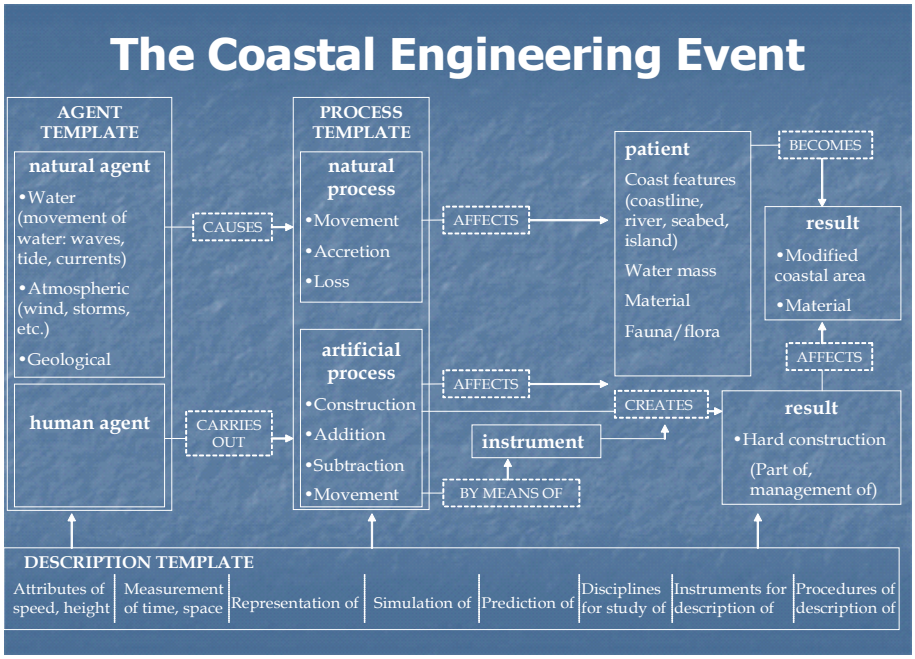
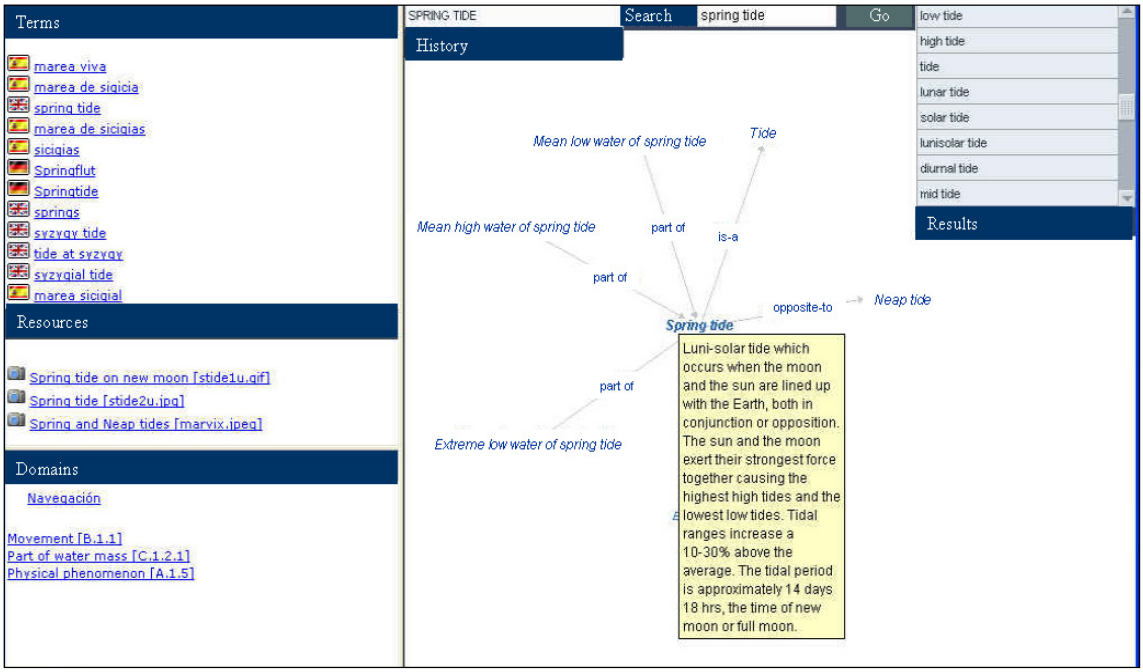


Figure 1: PuertoTerm Coastal Engineering Event

Thus, the CEE has two types of agents that can initiate processes. Such agents can be inanimate (natural forces) or animate (human beings). Natural agents, such as water movement (e.g. waves, tides, and currents) and atmospheric phenomena (e.g. winds and storms) cause natural processes such as littoral drift and erosion. These processes affect other entities or patients (e.g. beaches, sea ports, and seabed) which as a result, may suffer changes (e.g. loss/creation of beaches, modifications in seabed composition, etc.). Human agents can also implement artificial processes (e.g. constructions), which can generate or prevent effects normally caused by natural processes. For instance, a coastal construction can be created with the purpose of controlling erosion. With the help of instruments, man can build structures such as groynes, whose function is to protect a shore area, retard littoral drift, reduce longshore transport, and prevent beach erosion.

Thanks to this conceptual modelling, meaning is contextualized, i.e., the lexical concepts (Evans 2006: 18) and their linguistic designations are placed within specific frames and sub-frames (Barsalou 1999). The term base user has access to information on the conceptual, lexical and pragmatic description of each lexical concept. This encyclopaedic knowledge is displayed in a 3D visual thesaurus, where lexical concepts are placed in terms of the semantic relationships kept with other similar concepts. As an example, Figure 2 shows the entry for the lexical concept SPRING-TIDE.



Source: *http://manila.ugr.es:8080/puertoterm/*

Figure 2. Entry for the concept SPRING TIDE

Each entry includes:

Access to specific frame(s) under the tag *Domain* with hyperlinks to the parent frame(s) of the concept, i.e. PHYSICAL-PHENOMENON, MOVEMENT and PART-OF-WATER-MASS.

Access to frame relations, displayed in a dynamic network of related concepts such as TIDE and NEAP TIDE, related to the concept TIDE through an IS-A and OPPOSITE-OF relation, respectively.

Access to a definition based on a specific schema. The template established for TIDE concepts includes the *qualia* roles (Pustejovsky et al. 2000) and other domain-specific relations such as CAUSED-BY and AFFECTED-BY and the attributes TIDAL-LEVEL, TIDAL-RANGE and TIDAL-PERIOD.

Access to graphic resources (Faber et al. 2007), through the box *Resources* in Figure 2. The images are chosen with a view to highlighting the content of the definition.

Access to the terminological units designating the lexical concept in English, Spanish and German (spring tide, springs, syzygy tide, tide at syzygy, syzygial tide; marea viva, marea de sicigia, marea de sicigias, marea sicigial, sicigias; Springflut, Springtide).

Access to information regarding the lexical concept's *lexical profile*, i.e., "the general selectional patterns in terms of collocations and grammatical tendencies" (Evans 2006: 12). This information is given under each of the units listed in the box *Terms* (Fig. 3).

3. Description of the lexical profiles

The highly structured network found behind the visual thesaurus illustrated in Figure 2 requires a fully adequate account and description of the lexical profiles. This framework can be found within the context of Cognitive Linguistics, specifically within the Lexical Grammar Model (Faber and Mairal 1999, Mairal fc.). As an approach combining Cognitive Semantics and Construction Grammar (CxG), it is consistent with a series of assertions that are conducive to a valid theoretical and methodological framework. Our basic proposals stem from the guiding principles in CxG, the symbolic and the usage-based thesis.

Accordingly, meaning and grammar are seen as interdependent and complementary layers (Evans et al. 2007: 35). Consequently, the basic unit of grammar is a form-meaning pairing or *construction* that can be described as a conventionalized combination of form and meaning; which

includes every word, grammatical "rule" or template (Kuningas and Leino 2006: 302). In this vein, a lexical profile is composed of constructions, reflecting the collocational patterns both at a lexical and a syntactic level. In specialized discourse, these patterns are frequently realized by the traditional compounds and collocations which, according to Meyer and Mackintosh (1996), share a domain-specific semantic load. Thus, we use the umbrella term *terminological phraseme* (Meyer and Mackintosh 1994) to include entrenched, conventional combinations of linguistic units in the form of complex nominals and predicate-argument structures.

3.1. *The approach to complex nominals*

Combining a lexicon and syntax-driven approach to complex nominals, we contemplate I) constructions derived from underlying sentences (traditionally, verbal compounds); II) constructions derived from underlying prepositional phrases (traditionally, non-verbal compounds). For example, units such as *river bank erosion* can be syntactically explained as a process of predicate nominalization, where the head, *erosion*, is derived from the predicate ERODE and takes the non-head as argument$_2$, *river bank*. The compound *river bank* can be seen as a process of predicate deletion, where the head, *bank*, and the non-head, *river,* keep a semantic relation that must be inferred from lexical and contextual information (Borer 2003: 32). Paradoxically, this can be inferred from the definitional template of the non-head, the lexical concept RIVER, which contains a slot for the parts of the object. Under the category of complex nominals, we also account for constructions such as *all-in material*, with a PrepP+N configuration, where the noun exerts the selectional restrictions according to its definitional schema.

Since complex nominals can also be seen as constructions linked by not-overtly expressed grammatical relations (Bisseto and Scalise 2005: 327), we can establish a classification in terms of *relational compounds* and *attributive compounds*. Relational compounds include hyponyms, subsumptive constructions pointing to an IS-A relation (e.g. *spring tide* IS-A TIDE), and coordinate constructions (e.g. *luni-solar tide*). Other examples within coastal engineering include relations such as OPPOSITE-TO (*spring and neap tides*), HAS-FUNCTION (*diversion channel*) and DELIMITED-BY (*intermareal zone*). As attributive compounds, we can consider nominals like *high tide* and *semidiurnal tide*, instantiating modifications in the noun-head. The construction *high tide* results from the modification of the slot for the attribute TIDAL-HEIGHT and *semidiurnal tide* from the modification of the slot for TIDAL-PERIOD.

In *qualia* terms (Pustejovsky and Boguraev 2003), the adjective *high* needs to complement a physical object type. Since *tide* is a natural process, type coercion is required. Thanks to the FORMAL role of TIDE, with values relative to a physical object, *tide* is coerced to object and originates constructions relating to dimensions such as *high tide* and *low tide*. The adjective *semidiurnal* requires an event type. Thus, in *semidiurnal tide*, it complements the TELIC role of the event TIDE, the rising and falling of water, giving way to a frequency related construction.

3.2. *The approach to predicate-argument constructions*

Regarding predicate-argument structures, we study verb-argument constructions and non-verbal, argument-constructions from a lexical and syntactic approach also. We find patterns such as *to drain a river* (V+NP) and *tide rises* (N+V) which respond to the argument selection of the verbal head, the predicates DRAIN and RISE. However, we also contemplate support verb structures (Mel'čuk 1998) where the object, a deverbal noun, adjective or prepositional phrase, uses the verbal predicate to convey aspect and number in constructions like *carry out a drainage*, which instantiates the lexical concept DRAIN; *be beset* and *go adrift*.

Concerning non-verbal argument-like structures, they include argument-selecting heads such as nouns, adjectives and prepositional phrases. They exert a weak regime over their complements (Bierwisch 2005: 2) and project structures like *erosion of the river bank*, *bank of river*, *highly eroded*, and the Spanish *barco a la deriva* [ship adrift]. The first two examples deserve special attention. The pattern headword and preposition in the unit *erosion of the river bank* is taken as a special collocation (Ruppenhofer et al. 2002: 1) where the preposition *of* is a lexical marker of an argument of the head, the arg$_2$ *river bank*. The construction *bank of river* responds to the selection exerted by the second noun in the N *of* N arrangement. Again, as seen with the complex

nominal *river bank*, it is the definitional template of the lexical concept RIVER which determines such combination. Thus, based on the non-selectional properties of the noun head, these constructions are called *collocations with transparent nouns* (Ruppenhofer et al. 2002: 6). The last two examples, *highly eroded* and *barco a la deriva*, respond to the selecting properties of deverbal adjectives and prepositional phrases, respectively. The definitional schemas for the lexical concepts, ERODE and SHIP, are responsible for such combinations and, respectively, contain a slot for the attribute DEGREE-OF-EROSION and the TELIC or HAS-FUNCTION relation. The latter contains the value SAIL and fulfils the type-event complement requirements of the prepositional phrase *a la deriva* [adrift]. Thus, a classification of these constructions in terms of relational and attributive combinations is also appropriate.

3.3. *Terminological phrasemes and phraseme templates*

Under our proposal, all of the above types of constructions are taken as combinations governed by the principle of argument linking, which is neither purely syntactic nor purely lexical (Lieber 2004: 46). In fact, terminological phrasemes are conceived as constructions codifying conceptual, experiential and syntactic information concerning the lexical concepts of a cognitive frame. Consequently, the frame is the element which constrains the potential relations holding between the lexical concepts, and the construals that the frame allows are only a subset of the construals allowed by the argument-taking heads. For example, the argument *tide*, in *tide rises*, constrains the possible construals of the predicate RISE and viceversa.

The identification of lexical profiles also requires determining to what extent a linguistic construction instantiates a lexical concept and, ultimately, the conceptual system of the speaker (Gries 2003: 4). Therefore, we follow the CxG's usage-based thesis (Evans et al. 2006: 36) and carry out validations on the PUERTOTERM corpus, a repository of situated instances of language use. The natural behaviour of speakers is reflected in texts that have a reliable authorship, a modern perspective on coastal engineering, a variety of functions and registers (from the general to the highly specialized texts), and different geographic locations. The analysis of corpus data in the form of clusters allows us not only to draw the inventory of constructions populating a lexical profile, but also to make generalizations regarding such constructions.

Phraseme templates (Montero-Martínez et al. 2002) are used as dynamic metaconstructions that capture analogical relationships between several pairs of constructions (Leino and Ötman 2005: 206-207). They have associated constructional structures and semantics, inherited by all subordinate metaconstructions. The template shows the prototype of a network of related constructions (Gries et al. 2005: 640) which instantiates the central schema in a variety of syntactic arrangements or terminological phasemes. For example, the general template EROSION_AFFECTS_EARTH-CRUST includes the more specific template EROSION_AFFECTS_RIVER-BANK. Instantiations of this schema are the verbal construction *to erode the river bank;* the complex nominals *river bank erosion* and *eroded river bank*; and the non-verbal, argument-like construction *erosion of the river bank*. As explained, all of the different syntactic combinations are governed by weak or strong selections (Bierwisch 2005: 2) and respond to a conceptual template determining the interpretation of arguments (Borer 2003: 32). In regards to the constructional semantics associated with this phraseme template, it is represented by the verbal predicate ERODE and other predicate derivatives (*erosion*, *eroded*) and expresses the relationship between the process and the entity referred to, RIVER BANK.

4. Representation of lexical profiles

The lexical profile of a lexical concept needs an adequate instrument of representation. Within our approach, frames are used to make explicit the potential semantic and syntactic behaviour of the specialized language units instantiating each lexical concept. The frame-based structure of the terminological knowledge base PUERTOTERM allows the inclusion of the information of phraseme templates and their corresponding terminological phrasemes. This underlying metaconstruction is represented thanks to the series of semantic relations holding at a frame and lexical concept level. The basic *qualia* structure and the domain-specific relations account for such combinations and for the inheritance phenomenon.

As seen in Figure 2, each lexical concept entry presents a specific data category for the information on its lexical profile. An example can be found in Figure 3, corresponding to the term entry *spring tide*, an English designation of the lexical concept SPRING TIDE.

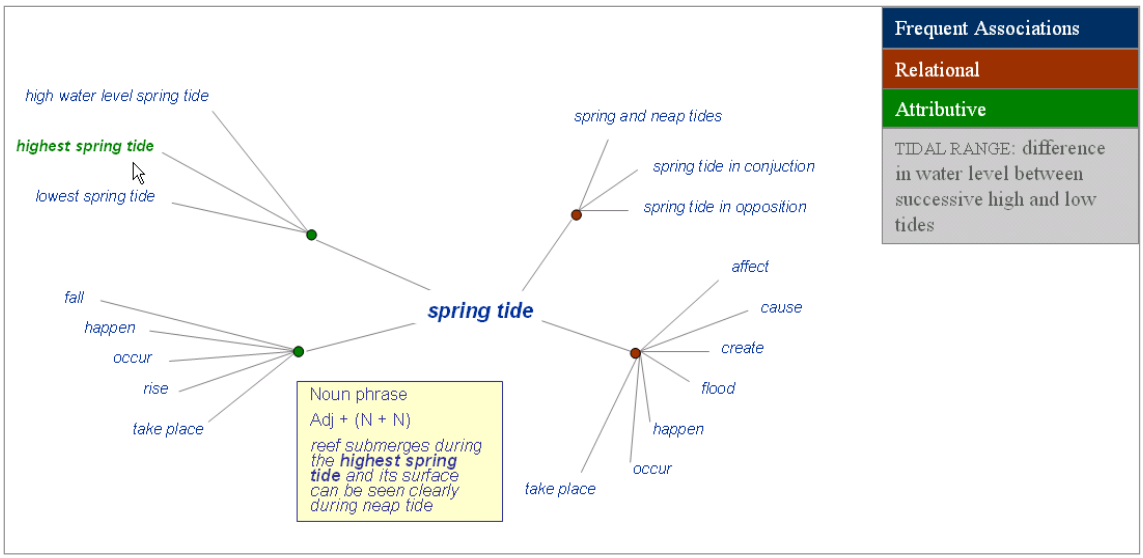


Figure 3. Representation of the lexical profile of *spring tide*

The information on the lexical profile is given through a user-friendly interface, where the elaborated classification and underlying conceptual structure are simplified. Here, terminological phrasemes, projections of the lexical concept, fall under the general category of *Frequent Associations*, both *relational* and *attributive*, according to the information encoded in the corresponding phraseme template.

For example, clicking on the unit *highest spring tide* reveals its attributive nature. The construction instantiates the phraseme template TIDE_TIDAL RANGE, i.e., the classification of tides according to the difference in water level between successive high and low tides. As shown, the terminological phraseme is further specified in terms of the grammatical category, NP; the syntactic structure, Adj+(N+N); and some linguistic contexts of use. Under the category of attributive associations, we also find verbal predicates such as *occur*, *take place* and *happen*. These are frequently found in propositions like "*spring tides occur every 14-15* days", where the prepositional phrase fills the definitional slot for the attribute frequency or TIDAL-PERIOD.

As examples of relational associations we find the verbal predicates *flood*, *affect*, *create* and *cause* which give rise to propositional constructions such as "*the spring tide floods the beach*", an instantiation of the phraseme template TIDE_AFFECTS_COASTLINE. The units *occur*, *take place* and *happen* are also used in relational constructions that refer to the ASTRONOMICAL-BODIES producing this natural phenomenon; for example, "*spring tides occur when the Sun, Moon, and Earth are lined up*". We also find units like *spring and neap tides*, pointing to an OPPOSITE-TO relation; and *spring tide in conjunction* and *spring tide in opposition*, subsumptive constructions (IS-A SPRING TIDE) that respond to the phraseme template TIDE_AFFECTED-BY_ASTRONOMICAL-BODY-POSITION. The prepositional phrases *in opposition* and *in conjunction* complement the slot for the ASTRONOMICAL-BODY-POSITION in the definitional schema for the lexical concept SPRING TIDE. Thus, they complement the physical objects SUN and MOON, the agents of the process.

In sum, we present a theoretical and methodological approach that accounts for the lexical profiles of concepts in a consistent and principled way, including the description of conceptual relations as well as the terms' combinatorial potential (Faber fc.). The results are displayed in a user-friendly THINKMAP representation.

References

Barsalou, L. (1999). "Perceptual symbol systems". *Behavioral and Brain Sciences* 22. 577-609.

Borer, H. (2003). "Exo-skeletal vs. endo-skeletal explanations: Syntactic projections and the lexicon". In Moore, J.; Polinsky, M. (eds.) *The Nature of Explanation in Linguistic Theory*. Stanford: CSLI Publications. 31-67.

Bierwisch, M. (2005). "Nominalization-Interdependence of syntax and lexical information". Paper presented at the workshop *QP structure, Nominalizations, and the role of DP* [on-line]. Center for Language Research and Technology. December 15th-17th, 2005. *http://web.uni-frankfurt.de/fb10/rathert/workshop/oxford/bierwisch.pdf*. [Access date: 20 Oct. 2007].

Bisseto, A.; Scalise, S. (2005). "The classification of compounds". *Lingue e Linguaggio* IV (2). 319-332.

Cohen, B. (1986). Lexique de cooccurrents. Bourse—conjoncture économique. Montreal: Linguatech.

Croft, W.; Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Evans, V. (2006). "Lexical concepts, cognitive models and meaning-construction". *Cognitive Linguistics* 17 (4). 491-534.

Evans, V.; Bergen, B. K.; Zinken, J. (2007). "The cognitive linguistics enterprise: An overview". In Evans, V.; Bergen, B. K.; Zinken, J. (eds.) *The Cognitive Linguistics Reader*. London: Equinox Publishing Company. 1-60.

Faber, P. (Forthcoming). "The cognitive shift in Terminology theory". In Vidal, A.; Franco, J. (eds.) "Una visión autocrítica de los estudios de traducción". *MonTI* 1 (1). Alicante: Universidad de Alicante.

Faber, P. et al. (2006). "Process-oriented terminology management in the domain of Coastal Engineering". *Terminology* 12 (2). 189-213.

Faber, P. et al. (2007). "Linking images and words: the description of specialized concepts". *International Journal of Lexicography* 20. 39-65.

Faber, P.; Mairal Usón, R. (1999). *Constructing a Lexicon of English Verbs*. Berlin, New York: Mouton de Gruyter.

Faber, P.; Márquez, C.; Vega, M. (2005). "Framing terminology: a process-oriented approach". *META* 50 (4). URL: *http://www.erudit.org/livre/meta/2005/index.htm*. [Access date: 15 March 2008].

Fillmore, C. J.; Johnson, C. R.; Petruck, M. R. L. (2003). "Background to Framenet". *International Journal of Lexicography* 16 (3). 235-250.

Gries, S. T.; Hampe, B.; Schönefeld, D. (2005). "Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions". *Cognitive Linguistics* 16 (4). 635-676.

Heid, U. (2001). "Collocations in sublanguage texts: Extraction from corpora". In Wright, S. E. (ed.) *Handbook of Terminology Management*: *Applications-Oriented Terminology Management*. Vol. II. Amsterdam, Philadelphia: John Benjamins. 788-808.

Kuningas, J.; Leino, J. (2006). "Word orders and Construction Grammar". *SKY Journal of Linguistics* 9. 301-309.

L'Homme, M. C. (2000). "Understanding specialized lexical combinations". *Terminology* 6 (1). 89-110.

Lainé, C. (1993). Combinatory Vocabulary of CAD/CAM in Mechanical Engineering. Ottawa.

Langacker, R. (2000). "A dynamic usage-based model". In Barlow, M.; Kemmer, S. (eds.) *Usage-based Models of Language*. Stanford: CSLI Publications. 1-64.

Leino, J.; Ötman, J. (2005). "Construction and variability". In Fried, M.; Boas, H. C. (eds.) *Grammatical Constructions. Back to the roots*. Amsterdam: John Benjamins. 191-209.

Lieber, R. (2004). *Morphology and Lexical Semantics*. Cambridge: Cambridge University Press.

López Rodríguez, C. (2000). *Tipología textual y cohesión en la traducción biomédica inglés-español: Un estudio*. Doctoral Dissertation. Granada: Editorial Universidad de Granada.

Mairal, R. (forthcoming). Lexical templates and syntactic structures: The design of a Lexical Grammar Model. Unpublished manuscript.

Martin, W. (1992). "Remarks on collocations in sublanguages". *Terminologie et Traduction* 2 (3). 157-164.

Martín Mingorance, L. (1990). "Functional Grammar and Lexematics". In Tomaszczyk, J.; Lewandowska, B. (eds.) *Meaning and Lexicography*. Amsterdam: John Benjamins. 227-255.

Mel'čuk, I. A. (1998). "Collocations and Lexical Functions". In Cowie, A. P. (ed.) *Phraseology: Theory, Analysis, and Applications*. Oxford: Clarendon Press. 23-53.

Meyer, I.; Mackintosh, K. (1994). "Phraseme analysis and concept analysis: Exploring a symbiotic relationship in the specialized lexicon". In Martin, W. et al. (eds.) In *Euralex 1994: Proceedings*. Amsterdam: International Congress on Lexicography. 339-348.

Meyer, I.; Mackintosh, K. (1996). "Refining the terminographer's concept-analysis methods: How can phraseology help?" *Terminology* 3 (1). 1-26.

Montero-Martínez, S.; García de Quesada, M.; Fuertes Olivera, P. A. (2002). "Terminological phrasemes in OntoTerm®: A new theoretical and practical approach. *Terminology* 8 (2). 177-206.

Picht, H. (1991). "Fraseología LSP desde el Punto de Vista Terminológico". *Sendebar* 2. 91-105.

Pustejovsky, J.; Boguraev, B. (2003). "Lexical knowledge representation and natural language processing". Class material presented at the course *Morphosyntax-Semantics Interface in Lexicalist Theories.* COLI. Summer Semester 2003. Universitat des Saarlandes.

Pustejovsky, J. et al. (2006). "Towards a generative lexical resource: The brandeis semantic ontology". [on-line]. Language Resources and Evaluation Conference, LREC 2006. URL: *http://www.cs.brandeis.edu/~arum/publications/lrec-bso.pdf*. [Access date: 30 Sept. 2007].

Ruppenhofer, J.; Baker, C. F.; Fillmore, C. J. (2002). "Collocational Information in the FrameNet Database". In Braasch, A.; Povlsen, C. (eds.) In *Proceedings of the Tenth Euralex International Congress*. Copenhagen, Denmark. Vol. I. 359-369.

Ruppenhofer, J. et al. (2006). *FrameNet II: Extended Theory and Practice.* [on-line]. Berkeley: ICSI, University of California. URL: *http://framenet.icsi.berkeley.edu/book/book.html*. [Access date: 30 Jul. 2007].

Sager, J. C. (1997). "Term formation". In Wright, S. E.; Budin, G. (eds.) *Handbook of Terminology Management: Basic Aspects of Terminology Management*. Vol. I. Amsterdam: John Benjamins. 25-41.