

Like a Lexicographer Weaving Her Lexical Network*

Alain POLGUÈRE

Université de Lorraine, ATILF, UMR 7118, Nancy, F-54000, FRANCE

Alain.Polguere@univ-lorraine.fr

The spider attitude

The title of our talk—an implicit reference to the English cliché *like a spider weaving her web*—intends to attract one’s attention to the metaphor that can be drawn between the dance of a spider weaving her web and a new *lexicographic gesture* that is gradually emerging from the work on Net-like lexical resources (Fellbaum, 1998; Baker et al., 2003; Gader et al., 2012). Our claim is that the inherent graph structure of natural language lexicons not only determine vocabulary acquisition and use (Wolter, 2006), but also lexicographic activity. In that respect, reflecting on new ways to implement the task of building lexical resources is essential for lexicographers themselves, but also for anyone interested in lexicons as mental structures. After all, lexicographers and language learners are those who have the most direct contact with lexical structures, through closely related activities: describing a natural phenomenon is a form of learning through explicit conceptualization. Lexicographers often experience the fact that by completing the description of a word they achieve a form of understanding and mastering of this word. They do not merely transcribe word knowledge and observations made on word behavior in speech and texts: they “acquire” the word. This makes them feel good and this explains why lexicography is indeed extremely addictive.

Our talk title is also an implicit reference to the English collocation *web of words*, that is so often used to refer to natural language lexicons as messy and too big to be embraced entities—cf. (Murray, 1977), entitled *Caught in the web of words: James A. H. Murray and the Oxford English dictionary*. Of course, webs can be seen as being essentially traps that one gets caught in. This is so to speak the fly or innocent bug perspective. However, lexicographers ought not be caught in the web: they can behave as spiders weaving the web. This is possible if the model they are constructing is indeed a diagrammatic representation—in a semiotic sense (Farias and Queiroz, 2006)—of the natural language lexicon that is being scrutinized. It is when lexicographers run on pages, writing dictionary articles, like flies walking on a glass window, that they have the most chance to get caught in the web of words. This is why lexicographers have long ago introduced systems of cards and records to help them compile data on lexical units. Lexicographic records helped lexicographers free themselves from the two-dimensional prison of the dictionary. Their knowledge about words occupied a “volume,” that of filing cabinets, which is more in line with the three-dimensional nature of the lexicons they had to describe. Later, with the advent of computational lexicography, relational databases replaced filing cabinets as convenient tools. . . and metaphors.

* Extended abstract for CogALex III invited lecture.

Towards a lexicography of virtual dictionaries

New data structures for lexical resources should come together with new ways of building lexical models, and this is the main topic we are dealing with here. In order to propose an alternate perspective on lexicography, one that in our opinion is more cognition-compatible in nature, it is necessary to first eradicate a rather widespread misconception related to the construction of lexical models. According to common perception, lexicography is all about writing dictionaries and, therefore, any activity that targets the construction of other types of lexical models, freed from the two-dimensional (textual) dictionary, is not “true” lexicography. This misconception, very common among laypersons and endorsed by many natural language researchers, originates mainly for the sheer fact that, for centuries, lexicographers had no better medium of encoding than the text and no better physical support for their description than sheets of paper bound together to make dictionaries. However, the dictionary—whether in paper or electronic format—is just one among many possible incarnations of lexical models. What is truly necessary and sufficient for a task to be termed *lexicography* is that:

- it targets the description of lexical units of one or more natural languages in terms of sense, forms and all other relevant linguistic properties;
- it uses a well-defined frame of analysis that allows for a coherent and uniform description of all lexical units;
- it is essentially a hand-made task, but with no limitation to the amount and diversity of tools and external data that can be used to perform this task;
- it “sees big:” the greater the coverage and depth of description for each lexical unit, the more lexicographic the task will be.

This last point is more important than it may appear: when it comes to the lexicon—its description, as well as learning, mastering, etc.—size does matter. To take an extreme case, a person whose only experience in the field is the description of just one or a couple of lexical units can simply not be considered a lexicographer and the task accomplished is all but an exercise in lexicography. By contrast, someone who has achieved the description of tens of thousands of lexical units is no doubt an experienced lexicographer. Somewhere in between, there is the transition from being an apprentice to being an actual lexicographer.

Notice that no mention of the formal nature of lexical models is made in the above characterization of lexicography. In fact, when the construction of a totally new, graph-based model of lexical knowledge was proposed by WordNet initiators (Miller et al., 1990), no claim was made on the advent of a new discipline. On the contrary, lexicography remained the reference, with work performed by individuals called *lexicographers*, who were constructing datasets called *lexicographer files*. And this is entirely justified as, precisely, lexicography is not about writing dictionaries *per se*. This fact has already been pointed at by some dictionary-makers; (Atkins, 1996), for instance, adopts a rather visionary perspective and goes as far as to consider that bilingual lexicography should be aiming at *virtual dictionaries*—cf. S. Atkins’ proposal for “real databases, real links and virtual dictionaries” (section 2.2.1 of her paper).

From writing dictionaries to weaving lexical networks

In our talk, we take the above observations as given, including the fact that lexicography should indeed be targeting virtual dictionaries, generated from non-textual lexical models

(Polguère, 2012). We illustrate how the lexicographic process of building graph-based lexical models can benefit from tools that allow lexicographers to wade through the lexical web, following paradigmatic and syntagmatic paths, while **simultaneously** weaving new links and incrementing the lexical description. Work performed on the *French Lexical Network* (Gader et al., 2012) will serve to demonstrate how the lexicographic process can be made closer to actual navigation through lexical knowledge by the speaker. The main theoretical and descriptive tool that makes such navigation possible is the system of lexical functions proposed by the Meaning-Text linguistic approach (Mel'čuk, 1996). It induces the multidimensional and non-hierarchical graph structure of the FLN that, we believe, is far better suited for designing lexical resources than hyperonymy-based structures.

Computational aspects of the work on the French Lexical Network are dealt with in (Gader et al., 2012). In our presentation, we focus on the actual process of weaving lexical relations.

References

- Atkins, B. T. S. (1996). Bilingual Dictionaries: Past, Present and Future. In Gellerstam, M., Järborg, J., Malmgren, S.-G., Norén, K., Rogström, L., and Pappmehl, C. R., editors, *Euralex'96 Proceedings*, pages 515–590, Gothenburg. Gothenburg University, Department of Swedish.
- Baker, C. F., Fillmore, C. J., and Cronin, B. (2003). The Structure of the FrameNet Database. *International Journal of Lexicography*, 16(3):281–296.
- Farias, P. and Queiroz, J. (2006). Images, diagrams, and metaphors: Hypoicons in the context of Peirce's sixty-six-fold classification of signs. *Semiotica*, 162(1/4):287–307.
- Fellbaum, C., editor (1998). *WordNet: An Electronic Lexical Database*. The MIT Press, Cambridge MA.
- Gader, N., Lux-Pogodalla, V., and Polguère, A. (2012). Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. In *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon. Enhancing the Structure and Look-up Mechanisms of Electronic Dictionaries (CogALex III)*, Mumbai.
- Mel'čuk, I. (1996). Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In Wanner, L., editor, *Lexical Functions in Lexicography and Natural Language Processing*, volume 31 of *Language Companion Series*, pages 37–102. John Benjamins, Amsterdam/Philadelphia.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. J. (1990). Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4):235–244.
- Murray, K. M. E. (1977). *Caught in the web of words: James A. H. Murray and the Oxford English dictionary*. Yale University Press, New Haven.
- Polguère, A. (2012). Lexicographie des dictionnaires virtuels. In Apresjan, Y., Boguslavsky, I., L'Homme, M.-C., Iomdin, L., Miličević, J., Polguère, A., and Wanner, L., editors, *Meanings, Texts, and Other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel'čuk*, *Studia Philologica*, pages 509–523. Jazyki slavjanskoj kultury Publishers, Moscow.
- Wolter, B. (2006). Lexical Network Structures and L2 Vocabulary Acquisition: The Role of L1 Lexical/Conceptual Knowledge. *Applied Linguistics*, 27(4):741–747.