



HAL
open science

From Writing Dictionaries to Weaving Lexical Networks

Alain Polguère

► **To cite this version:**

Alain Polguère. From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, 2014, 27 (4), pp.396-418. 10.1093/ijl/ecu017 . hal-01097112

HAL Id: hal-01097112

<https://hal.science/hal-01097112>

Submitted on 18 Dec 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

FROM WRITING DICTIONARIES TO WEAVING LEXICAL NETWORKS

Alain Polguère: *Université de Lorraine & ATILF CNRS, Nancy, France*
(*alain.polguere@univ-lorraine.fr*)

Abstract

Dictionaries are designed as huge texts made up of a collection of much smaller texts, i.e. lexicographic articles. To put it differently, dictionaries are two-dimensional textual models of natural language lexicons. Lexicographers, however, are well aware of the fact that their task is to account for a truly multidimensional entity: a gigantic graph of lexical units connected by various paradigmatic and syntagmatic relations. The most significant advance that computer science will bring to the future of lexicography is therefore not the ability to better store, search and manipulate textual lexicographic data; it will be to allow lexicographers to bypass the text as a formal representation of lexicons and to directly work on lexical networks. Such networks are more suitable to the lexicographic endeavour because they are better formal metaphors of the “natural” structure we are trying to account for. This paper presents lexical systems as graph models of lexicons and introduces the corresponding lexicography of virtual dictionaries. It is based on extensive lexicographic work that is being conducted on the French Lexical Network, a lexical database built according to theoretical and methodological principles borrowed from Explanatory Combinatorial Lexicology.

1. The proper structure of natural language lexicons: all bets are on the table

1.1 Exploring new lexical structures

The idea that dictionaries—huge texts made up of a collection of articles about words—are not the optimum models of natural language lexicons is not new. Lexicographers themselves have experimented with new structures that could better reflect the complex system of lexical relations that is making up our lexical knowledge; for instance, the *Language Activator* for English (Summers 1993) and the *Dictionnaire du Français Usuel* for French (Picoche and Rolland 2002). The advent of electronic publishing gave lexicographers an opportunity to improve the structuring and access to lexicographic information (de Schryver 2003, Lew 2012). But the true revolution in the field of lexical modelling came with the Princeton WordNet project (Miller *et al.* 1990, Fellbaum this volume). There are at least four aspects of this project that, once combined, made it revolutionary.

1. It was based on a theory of how lexical knowledge is structured, this theory imposing strong constraints on how lexical models should be computationally organised as graphs of lexical entities (senses, grouped into hierarchically organised synsets).
2. It was designed as a full-fledged lexicographic project; i.e. the *raison d'être* of the theory was to lead to actual descriptive models of natural language lexicons.
3. It succeeded in producing an extensive description of contemporary American English that reflects the initial theoretical postulates and is fully implemented as a formal

database; this achievement led the path to the construction of WordNets of other languages (Vossen 1998, Fellbaum and Vossen 2012).

4. Last but not least, the Princeton WordNet project is not a commercial enterprise and the resulting model is made available to the community of researchers and developers in Natural Language Processing, linguistics and other language-related fields.

WordNet opened new perspectives in lexicography and demonstrated that actual lexicographic work can be performed building information structures that are radically different from that of dictionaries. In particular, it demonstrated that relational, net-like models of lexicons could indeed be the future of lexicography.

On the negative side, while WordNet aimed at being a cognitively motivated model—Miller *et al.* (1990) promoting an approach identified as *psycholexicology*—, it is being used mainly to support studies and applications that have little to do with cognition. One probable cause of this state of affairs is that WordNet’s ontological structure is not as cognitively relevant as it was expected to be by its designers. Another explanation is to be found in the fact that WordNet became rapidly so popular in the Natural Language Processing community that the focus of the project shifted at an early stage from psycholinguistics to computer applications (C. Fellbaum, personal communication). The WordNet project remains the greatest source of inspiration for anyone interested in finding radically new ways of performing lexicography, except precisely for its ontological synset-based organisation. *Psycholinguistic evidence* put forward by G. Miller (Miller 1990) to support the relevance of WordNet’s global structure is, to say the least, rather thin and in no way prevents researchers from exploring alternative approaches. Furthermore, the present state of knowledge or lack thereof about the actual structure of the mental lexicon (section 2.2.1 below) indicates that no current approach should be taken at face value. The goal of this paper is not to try to disprove the ontological theory behind WordNet; it is to offer an alternative structure for lexical models, called *lexical systems*, and to explain how this alternative approach to lexical structuring can be implemented with a new type of lexicography, where the writing of lexicographic articles is replaced with the weaving of lexical networks.

The general approach to the structuring of lexical knowledge that we present has been implemented in the context of an on-going project that targets the construction of a full-scale lexical resource for the French language called the *French Lexical Network*—hereafter, *fr-LN*. The design of this lexical resource abides to theoretical and descriptive principles put forward by Explanatory Combinatorial Lexicology (Mel’čuk *et al.* 1995, Mel’čuk 2006). It is however radically distinct from previous lexicographic work done within this framework in that it relies on a model of lexical information—lexical system—that is not a dictionary and that is fundamentally non-textual in nature.

The first three-year stage of the long-term fr-LN enterprise will be concluded soon and the initial results are significant enough to serve as a “proof of concept” of the approach. It allows for some generalisations to be drawn—beyond the specific case of the fr-LN—that could be valid for the field of lexicography at large. Several publications are already available on the computer implementation of the fr-LN, its lexicographic editing and its growth process (Lux-Pogodalla and Polguère 2011, Gader *et al.* 2012, Polguère and Sikora 2013). This paper offers a synthesis of the work done, focusing on lexical systems as an alternative to ontological and/or synonymy-based network models of lexical knowledge.

1.2 Ontological vs non-ontological lexical models

Lexical systems fully belong to the family of lexical networks, and can therefore be directly related to WordNets. It should however be highlighted that two radically distinct types of lexical networks can be considered:

1. Ontologically organised lexical networks, such as WordNets, structure lexical information in a mainly hierarchical fashion, in compatibility with Aristotelian concept categorisation principles (Sowa 2010). Though they are not actual ontologies (Hirst 2009, Fellbaum and Vossen 2012: 320), such networks are intended to be usable as linguistic modules in the world-lexicon interface and allow for the connection of the “deeper” realm of concepts with the “more superficial” realm of lexical units.
2. Non-ontological lexical networks, on the other hand, emphasise the relational nature of lexicons, while ignoring or putting on the back-burner the problem of general classification of lexical units and properties inheritance. Lexical systems belong to this second type of lexical networks.

The above classification is necessarily a simplification of the actual picture. A lexical resource such as FrameNet (Fillmore *et al.* 2003, Ruppenhofer *et al.* 2010), for instance, is only mildly a “true” lexical network. Like WordNet, however, it does project a classifying organisation onto the structure of the lexicon by relating lexical senses to semantic frames that are interrelated (inheritance being one of the fundamental relations between frames) and function as a general system of categorisation of lexical units.

We are deeply convinced that both approaches identified above are incompatible: the lexicon is either organised as a hierarchical classifying structure or as a non-hierarchical multidimensional one. Both visions of the lexicon have to be explored, applied and put to the test (in Natural Language Processing, language teaching, studies of language pathologies, etc.) before any conclusion can be drawn on their appropriateness. Up to now, it is mainly the first, ontological vision of the lexicon that has been theorised and systematically applied.

Lexical systems are not only non-ontological lexical networks. They are also distinct from most known net-models in that they are not centred around one specific lexical relation: namely, synonymy. There is no doubt that synonymy is the mother of all lexical relations and that synonymy and, more generally, paraphrase is at the core of natural language semantics. But synonymy should participate in the structuring of lexical networks as only one among many other fundamental lexical relations. At least three arguments can justify this assertion.

Firstly, synonymy is part of speech sensitive and, if taken as structuring principle of lexical models, keeps apart words that have similar meanings but cannot be substituted for each other in sentences. In a synonymy-based lexical network such as WordNet, the noun *theft* is topologically closer to *felony*—their respective synsets are linked by a hypernym relation—than it is to the verb *to steal*, in spite of the fact that this latter has an equivalent semantic content: *her theft* \approx *the fact that she stole something*. Of course, a connection between *to steal* and *theft* can be implemented in a WordNet, but this connection will be structurally marginal and will cut across the ontological organisation of the lexical network.¹

Secondly, even within the same part of speech, words can be semantically related without entertaining a relation of synonymy: *theft* \sim *thief* \sim *loot*. Such semantic relations—that are often but not necessarily backed by a morphological link—play a fundamental role in structuring our lexical knowledge and its use in speech production and understanding.

Finally, there is no reason why only paradigmatic relations should form the global structure of lexical networks. Each lexical unit is no doubt connected to the rest of the lexicon through semantic relations it entertains with other lexical units, but also and in an equally important way through combinatorial (syntagmatic) relations. The structure of a lexical network should display a direct connection between *theft* and the verb *to commit* or the adjective *petty*. For a relational structuring of bilingual lexical information based on syntagmatic relations, see for instance Fontenelle (1997).

Modern linguistics, through the extensive study of paradigmatic and syntagmatic relations found in natural languages, has helped us understand better the logic behind the structuring of natural language lexicons. More specifically, we consider that the universal system of standard Meaning-Text lexical functions (Mel'čuk 1996) is a solid, well-tested hypothesis of the system of lexical relations that weave the web of words. As we will show immediately below, standard lexical functions have been chosen as structuring elements of lexical systems.

2. Lexical systems as non-ontological models of lexicons

This section presents lexical systems from a formal point of view as non-ontological network models of natural language lexicons to be used for both language studies and lexicography. We proceed in two steps: formal properties of lexical systems (2.1) and justification of the approach (2.2).²

2.1 Formal properties of lexical systems

The lexical system of a given language L is a model of L 's lexicon that possesses specific characteristics in terms of informational content and structuring. The notion of lexical system was first elaborated and experimented with using lexicographic data extracted from the SQL version of the DiCo French lexical database (Mel'čuk & Polguère 2013), and a preliminary description of the model has been published in Polguère (2009). The initial proposal was not supported by a full-scale implementation of a lexical system. It is now possible to draw from the concrete results that have been obtained while lexicographically building the fr-LN in order to refine and strengthen the notion of lexical system.

Lexical systems possess four main characteristics that, taken together, distinguish these lexical networks from other existing ones. These characteristics relate to:

1. components of lexical systems as graphs, in the mathematical sense—i.e. types of nodes and edges (2.1.1);
2. non-atomicity of lexical nodes (2.1.2);
3. general non-ontological structure of lexical systems—they belong to the family of so-called *small-world networks* (2.1.3);
4. measure of confidence associated to information stored in lexical systems (2.1.4).

2.1.1. Components of a lexical system

The lexical system of a given language L is formally an oriented graph. Nodes of this graph correspond to lexical entities of L . These entities are mainly lexical units of L ; according to Meaning-Text terminology (Mel'čuk 2012: 21–44), lexical units are of two different types:

- lexemes—e.g. FOOTBALL I [*Paolo has been playing football forever.*], FOOTBALL II [*Paolo kicked the football into the goal.*] ...;

- idioms—e.g. ‘TAKE A RAIN CHECK’, ‘PIECE OF CAKE’ ...³

Marginally, nodes of lexical systems can also be non-lexicalised (i.e. compositional) phraseological expressions such as linguistic clichés—e.g. *Sorry for your loss*, *Post no sign* ...

It is essential that each node of the graph stand for one well-specified meaning of *L*. In particular, it can by no means be a polysemic vocable of *L* (roughly, a dictionary entry), as vocables are groupings of lexical units (vocabulary senses) but not actual units, and lexical links hold between senses only. A lexical system is thus a “disambiguated” graph, unlike vocabulary-based graphs such as Motter *et al.* (2002) for English or Ploux (1997) for French.

Edges of lexical systems are, in their great majority, semantico-syntactic connections known as Meaning-Text lexical functions (Mel’čuk 1996). The system of *standard* lexical functions (Polguère 2007) allows for the rigorous encoding of regular semantic derivations (synonymy, antonymy, conversivity, actant names, actantial adjectives, etc.) and collocational relations (intensifiers, support verbs, etc.). For instance, for any lexical unit LU, **Oper₁**(LU) is the lexical function application that returns all support verbs that take the first actant of LU as subject and LU itself as first complement.

Let us illustrate the **Oper₁** lexical function with data borrowed from the fr-LN, namely, with the French lexical unit HOLD-UP I ‘(robbery) hold-up’, which is the basic non-metaphorical sense of the polysemic Fr. HOLD-UP vocable.⁴

- (1) **Oper₁**(HOLD-UP I) = *commettre* [ART ~], *faire* II.1 [ART ~], *réaliser* III [ART ~]

Figure 1 visualises the lexical graph that underlies (1).

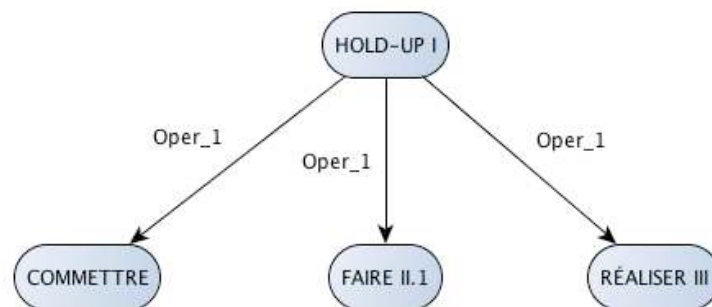


Figure 1. Lexical graph for Fr. **Oper₁**(HOLD-UP I)

Paradigmatic and syntagmatic lexical function links form the backbone of lexical systems, but they are not the only lexical connections present in natural language lexicons. In particular, the three following types of links are added to lexical systems:

1. copolysemy links between lexical units that are senses of the same vocable—Fr. HOLD-UP II [*le hold-up du gouvernement sur les prêts sociaux* lit. ‘the government’s hold-up on social loans’] is connected to HOLD-UP I by a copolysemy link of metaphor;
2. semantically empty links between idioms and lexemes that are formally included in the idiom phrase—Fr. ‘DONNER CORPS’ ‘to give existence’ is connected to lexemes DONNER III.2 ‘give’ and CORPS I.1a ‘body’ in the fr-LN by a link of formal inclusion;
3. semantic links of definitional inclusion—HOLD-UP I contains in its definition the meaning ‘voler ²I’ (‘to steal’) and is thus connected to the corresponding lexeme by a link of semantic embedding.

We are aware of the fact that letting lexical functions be the fundamental structuring principle of lexical systems is a strong theoretical hypothesis. No actual proof can be presented at this point to demonstrate the soundness of our approach, and only arguments in its favour will be provided (see section 2.2 below).

2.1.2. *Lexical nodes as complex entities*

Nodes of a lexical system graph are non-atomic entities: they possess a rich internal structure. If we crack open a lexical node, we access information that is equivalent to the lexicographic article of the corresponding lexical unit. Each lexical unit node is thus associated with a structured set of properties, mainly:

- grammatical characteristics—part of speech, gender (if applicable), inflection, etc.;
- semantic label (Polguère 2011), actant structure and definitional paraphrase;
- syntactic government pattern (Mel’čuk 2013: 307–309);
- paradigmatic and syntagmatic connections—implemented as edges of the graph;
- citations and other lexicographic examples.

These sets of properties correspond to the different zones of a lexicographic article in an *Explanatory Combinatorial Dictionary* (Mel’čuk 2006, Mel’čuk 2013: Chapter 11) or in any lexical database built according to theoretical and descriptive principles of Explanatory Combinatorial Lexicology. Storing such properties “inside” lexical nodes is equivalent to building the microstructure of a lexical system. How this can actually be done is explained and illustrated in section 3.1 below.

To conclude on the topic of their components, lexical systems are rather heterogeneous graphs, with very diverse types of nodes and edges. While this heterogeneity can be a handicap from a formal and computational point of view, we believe that it reflects an intrinsic property of natural language lexicons: they are made up of various types of lexical entities and of a very rich system of connections between them. Our present concern is with lexical systems as theoretical perspective on the structuring of lexical information. We are not dealing with the actual programming of lexical systems as computer databases—on this topic, see Gader *et al.* (2012), where the SQL implementation of the fr-LN is introduced. Implementation is not a trivial issue however, and it is vital for a lexicographic enterprise based on lexical systems to computationally design data structures that will meet the specific needs of lexicographic edition: creation/destruction of lexical nodes, (un-)weaving of lexical links, association of rich information to each lexical nodes, etc., performed by a team of lexicographers working simultaneously on the same set of data.

2.1.3. *Non-ontological small-world graphs*

Lexical systems are non-ontological models of natural language lexicons (see above, section 1.2). This is a direct consequence of the fact that the general organisation of these models is mainly determined by Meaning-Text lexical functions: it is a system of lexico-syntactic relations that is radically distinct from the hierarchical organisation of concepts into classes and subclasses postulated by ontological models. To schematise: the visualisation of an ontological lexicon looks a bit like a flattened Eiffel Tower, while that of a lexical system looks more like a spaghetti dish.

The graph of a lexical system can thus appear to be, at first glance, a “mess” of paradigmatic and syntagmatic relations, where no classifying principle emerges. However, such graph possesses at least four remarkable properties.

- P.1** The graph’s edge density is weak. Lexical graphs, and lexical systems in particular, can be said to contain relatively few edges in regards to their total number of nodes.
- P.2** The average shortest path between nodes is small. In general, there exists at least one relatively short edge path between any two nodes in the graph, in spite of the weak edge density (**P.1**).
- P.3** Lexical nodes form clusters. Though characterised by a global weak density (**P.1**), lexical graphs are locally dense in terms of arcs, as if made up of “lumps” of nodes.
- P.4** The nodes’ degree distribution decreases according to a power law. In less mathematical terms, we find in these weakly dense graphs a small minority of highly connected nodes (cf. crossroads lexical units, section 3.2 below).

Graphs that possess the above properties belong to the family of *small-world graphs*. The above characterisation is of course an over-simplification of the actual mathematical notion of small-world graphs, originally defined in Watts and Strogatz (1998).

Properties of lexical small-worlds have been extensively studied by B. Gaume (Gaume 2004, Gaume *et al.* 2006, Gaume 2008), who has demonstrated that it was possible to use their formal properties to identify in them semantic spaces and, to put it in more general terms, to organise the seemingly disorganised lexical knowledge (see section 2.2.2 below).

Though still limited in terms of coverage, the fr-LN in its current state is already sufficiently rich to display formal characteristics of small-world graphs (Gader *et al.* 2014). In principle, any lexical system, as defined here, should possess this specific formal structuring.

2.1.4. *Relativist model*

Information in a lexical system has to be associated with a degree of confidence. There are two reasons for this:

1. Lexical knowledge is in essence gradable due to several factors such as diachronic changes and potential indecisiveness of speakers relative to their own mastering of the language (which includes lexicographers as well).
2. Lexical systems, as models, have to be compatible with the implementation of logical inferences and the construction of lexical knowledge through a process of approximation—see section 2.2.3 below.

The gradable nature of lexical information is implemented in the fr-LN, where each piece of linguistic information—including the very existence of lexical units, their grammatical characteristics, etc.—is associated with a quantified degree of confidence, set by default to its maximum value when a piece of information is manually entered by lexicographers themselves. Lexicographers can however “relativise” lexicographic information by pushing the cursor down (literally) and thus associate to information a lower degree of confidence. It has been decided that by default (i) the minimal degree of confidence that can be manually set by lexicographers is 60% and (ii) any piece of information that is automatically injected in the lexical model carries a degree of confidence of 50%. It is this latter specific degree of confidence associated to automatically generated data that allows for their identification when they need to be retrieved and processed.

In the course of our experiments with lexical systems, we have computationally generated an *English Lexical Network*—the *en-LN*— by compiling lexicographic information embedded in the Princeton WordNet (Gader *et al.* 2014). Based on the above-stated principle, all linguistic information contained in the generated *en-LN* has been associated with a degree of confidence of 50% and any subsequent modification that is manually performed on this lexical system by means of our lexicographic editor (section 3.1 below) will by default be considered as entirely valid—degree of confidence of 100%.

Clearly, degrees of confidence are used in the *fr-LN* and *en-LN* in a basic way. These lexical models are nevertheless equipped with all necessary means to model potential “fuzziness” of lexical information. This will be particularly relevant in the future, when implementing the semi-automation of lexical system growth—section 2.2.3.

2.2 Justification of the approach

This section offers theoretical and empirical arguments for adopting lexical systems as lexical models. We look at the following three aspects of the question: relevance of lexical systems from a psycholinguistic point of view (2.2.1), application of computational methods based on *Proxemy analysis* (2.2.2) and semi-automation of the growth of lexical systems (2.2.3).

2.2.1. *Lexical systems as models of the mental lexicon*

In what follows, we use the term *mental lexicon* (Aitchison 2012, Wierzbicka 2009) to designate lexical knowledge as stored in the Speaker’s mind. There are two main reasons why it can be important for lexical systems to be plausible representations of the actual informational structure of mental lexicons. Firstly, this would entail that lexical systems are relevant models to be used for lexicological as well as psycholinguistic studies, as they would be closer to “individual lexicons.” Secondly, cognitively relevant lexical models should allow for better computer simulation of access to and processing of lexical information, with an efficiency that would be close to that of the human brain. The quest for a compatibility between formal lexical models and the hypothesised mental lexicon can thus be justified both for research and practical applications.

There is a growing amount of research on the mental lexicon performed on graphs that formally resemble lexical systems in that they are small-world graphs or, at least, are not structured by ontological principles—see Gruenenfelder and Pisoni (2009), Baronchelli *et al.* (2013) and Morais *et al.* (2013). It is therefore not unfounded to postulate, as we do, that lexical systems could be good approximations of the structure of mental lexicons, if by this term we refer exclusively to linguistic lexical knowledge and not to “deeper” knowledge of concepts and the world. (We are in no position to make any assumption on how concepts themselves and extralinguistic knowledge is mentally organised.) Another indication of the potential cognitive relevance of lexical systems is their formal compatibility with proposals that have been recently made for the modelling of vocabulary acquisition as enrichment of lexical networks—e.g. Wolter (2006) and Gaume *et al.* (2008).

Domains of applications that ought to be explored in priority in order to validate the hypothesis of the psycholinguistic relevance of lexical systems are, in our opinion: analysis of lexical errors in spontaneous speech and diagnosis of lexical speech anomalies that result from mental or physiological disorders.

2.2.2. Identification of semantic spaces in lexical systems

By *semantic space*, we mean a rather small set of lexical units in the lexical network of a language that displays some form of semantic cohesion. The notion of semantic space is close to that of *semantic field* (section 3.2 below), but remains distinct as all elements of a semantic space do not necessarily belong to the same semantic field. This is explained by the fact that the meaning of a given lexical unit is directly related to its combinatorics. This is true for so-called *free combinatorics*—that is directly inherited from the definition of the lexical unit—, but also for *restricted combinatorics* (phraseological or, more generally idiosyncratic behaviour in sentences). Thus, even though the phrase Fr. *commettre un hold-up* lit. ‘to commit a hold-up’—cf. section 2.1.1 above—is an **Oper**₁(LU)+LU collocation whose base LU is the lexical unit HOLD-UP I, it is logical to postulate that many nominal lexical units that are semantically close to HOLD-UP I will have a tendency to control the same **Oper**₁ collocate: *commettre un crime, un massacre, un meurtre* ‘murder’, *un vol* ‘theft’, etc.

Lexical units of a same semantic field thus tend to “gang up” in lexical small-world graphs, but they will also tend to exert a force of attraction on collocational lexical units that are very vague or quasi-empty and to suck them into their semantic space. A semantic space is therefore, potentially, an expansion of a semantic field that results from the presence of syntagmatic lexical affinities.

Semantic spaces can be identified in lexical networks by means of *Proxemy analyses*. Proxemy is a measure of similarity defined on small-world graphs (Gaume 2004) that has been mainly applied on synonymy graphs (Gaume *et al.* 2006). It allows for the identification of semantic spaces in a completely non-linguistic manner, i.e. by making use exclusively of the “naked” graph structure, expressed in terms of nodes and edges connecting these nodes.

Proxemy analysis performed with the Prox random walk method (Gaume 2008) is presently used with the fr-LN (for now, without any special tuning for lexical systems) in order to extract lexical clusters related to given lexical units. A graphical visualisation using Tmuse (Chudy *et al.* 2013) allows for 3-D display of Proxemy analysis results, as illustrated in Figure 2. This figure displays the results of a Proxemy analysis performed on the Fr. CIGARETTE vocable, which currently contains the following three senses in the fr-LN:

- CIGARETTE I ‘cigarette’ [*Elle allume sa cigarette.* ‘She lights up her cigarette’]—basic lexical unit of the vocable;
- CIGARETTE II ‘habit of smoking cigarettes’ [*Il faut que tu arrêtes la cigarette.* lit. ‘You have to stop the cigarette’ ‘You have to quit smoking cigarettes’]—metonymic sense;
- CIGARETTE III ‘[piece of clothing] tight, thin and long, looking like a cigarette’ [*Elle porte une jupe cigarette* ‘She wears a cigarette skirt’]—appositive metaphorical sense.

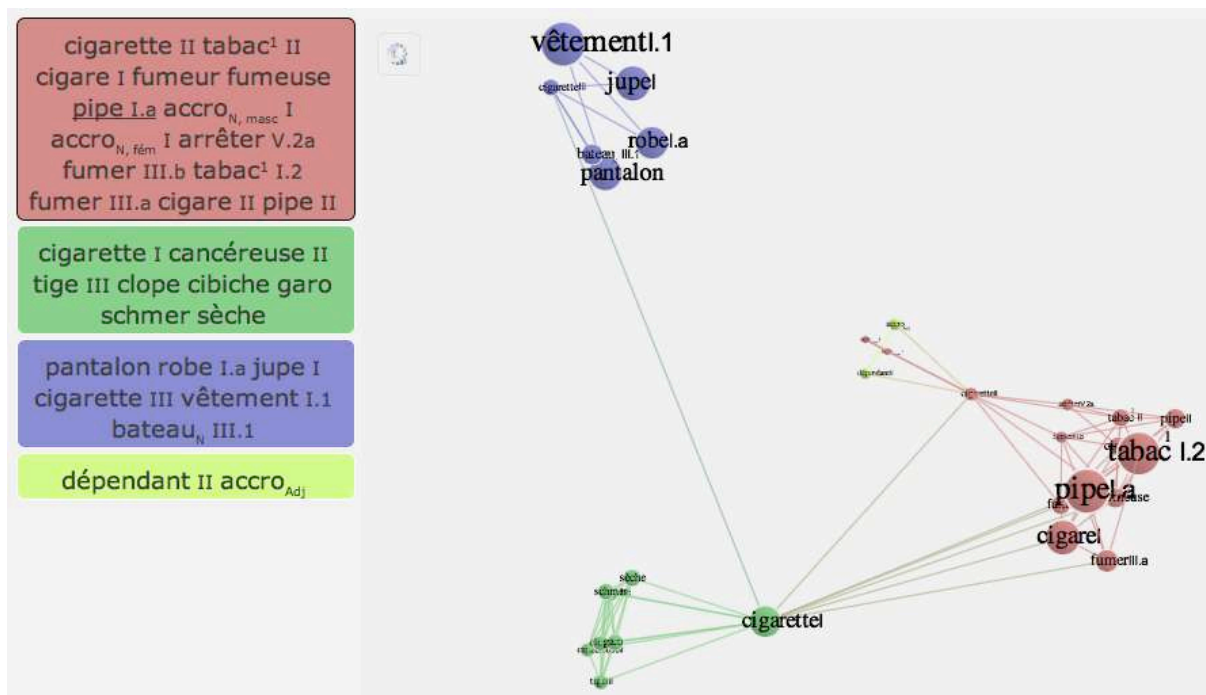


Figure 2. Proxemy clustering and visualisation of Fr. CIGARETTE

Figure 2 clearly shows the presence of three main clusters (though four have been recovered in total). In the graphical display, edge lengths reflect the measure of proxemy: the longer the edge, the weaker the semantic proximity. 3-D visualisation indicates that the topological analysis of the graph has identified a central CIGARETTE I sense, with a distinctively remote cluster corresponding to the use of *cigarette* (CIGARETTE III) as appositive collocational modifier of nouns such as *jupe* ‘skirt’ or *pantalon* ‘trousers’.

Figure 2 is of course far from offering a complete and adequate view of the positioning of Fr. CIGARETTE and all its senses in the global French lexical system. Limitations originate both from the youth of the fr-LN, which is still very incomplete, and from the fact that research remains to be done in order to develop a version of Proxemy analysis that is tailor-made for lexical systems. Work in that direction has just started, with both descriptive and applicative targets, as will now be explained.

2.2.3. Toward semi-automatic growth of lexical systems

There exist at least four main methods for building lexical resources:

1. manual lexicographic construction—cf. WordNet or the fr-LN;
2. lexical data collection by crowdsourcing—cf. Wiktionary or, with a game strategy, JeuxDeMots (Lafourcade 2007);⁵
3. automatic extraction and structuring of linguistic data available in lexical resources—cf. WOLF (Sagot and Fišer 2008) or BabelNet (Navigli and Ponzetto 2012);
4. automatic extraction of linguistic data from text corpora (Hearst 1992).

The second method cannot be trusted if one needs models that offer a fine-grained and high quality description of lexical information that could be used for instance in the context of vocabulary teaching.

The third method presupposes that lexicographers have indeed been at work, manually building lexical resources. Without dictionaries, WordNets and the like, there would not be material that could serve to feed projects of automatic construction of derived lexical databases. Lexical resources, additionally, do not exist once and for all: natural languages endlessly evolve and lexicographic modelling is a perpetually on-going activity.

The fourth method—though particularly interesting if one wants to rapidly gather information on how lexical units do behave in actual speech—generates rough descriptions, suitable for basic Natural Language Processing only.

So it looks like someone has got to do it. And who else but lexicographers? In spite of this, it is perfectly legitimate to envisage the semi-automation of the lexicographic process. It can be performed with drafting tools (Kilgarriff and Rychlý 2010) or through semi-automatic inferential enrichment based on the very structure of lexical information (Sajous *et al.* 2011). The structure of lexical systems is particularly suited for this kind of task and preliminary observations that have been made on Proxemy analysis of the fr-LN graph (cf. section 2.2.2 above) are quite encouraging. For instance, clusterisation can help identify holes in semantic spaces that result from the youth of the fr-LN. It can correspond to lexical units that are missing in the model or, more frequently, to already existing lexical units that are insufficiently connected to the rest of the graph. It is thus worth using algorithms that perform walks through small-world graphs in order to identify in the fr-LN significant recurrent micro-patterns of lexical connections. These patterns can later be exploited in order to automatically generate plausible missing graph elements that will later be manually validated.

Now that lexical systems have been introduced from a formal point of view and that arguments for their construction and use have been discussed, it is high time to explain how the lexicographic process of building lexical systems takes place.

3. Lexicography as weaving of lexical systems

Lexical systems are graph metaphors of natural language lexicons and their construction should be performed as a graph weaving rather than text writing activity. The fr-LN is built using a specially designed graph editor—named *Dicet*—that supports weaving and wading through lexical graphs (Gader *et al.* 2012).⁶ This section focuses on lexicographic processes that are made possible with *Dicet* and that characterise the lexicography of lexical systems. Lexicographic operations will be considered from the viewpoint of both the microstructure (section 3.1) and the macrostructure (3.2) of lexical systems. Notice that metalexigraphy postulates a third level of structural organisation for dictionaries called their *mediostructure*: the system of crossreferences between lexical entries (Gouws & Prinsloo 1998). It is however irrelevant to try to single out such structure for lexical systems as they are nothing but mediostructures. By removing arcs in a lexical system, we do not simply remove crossreferences, we remove the most central element of each lexical unit's description: its system of connections with other lexical units.

3.1 Microstructure level: giving flesh to lexical nodes

The microstructure of a lexical system is the internal structure of nodes of the lexical graph—cf. section 2.1.2 above. We use the term *lexicographic article* to refer to this structure even though it is not a **text**, as will be shown below. Central to this microstructure is the lexical function zone of node articles. It is built by lexicographers by weaving lexical function links between the headword of the article and other lexical units (nodes) of the lexical system.

For instance, the set of **Oper**₁ connections controlled by Fr. HOLD-UP I (Figure 1, section 2.1.1 above) is weaved and edited with the Dicot editor as shown in Figure 3.

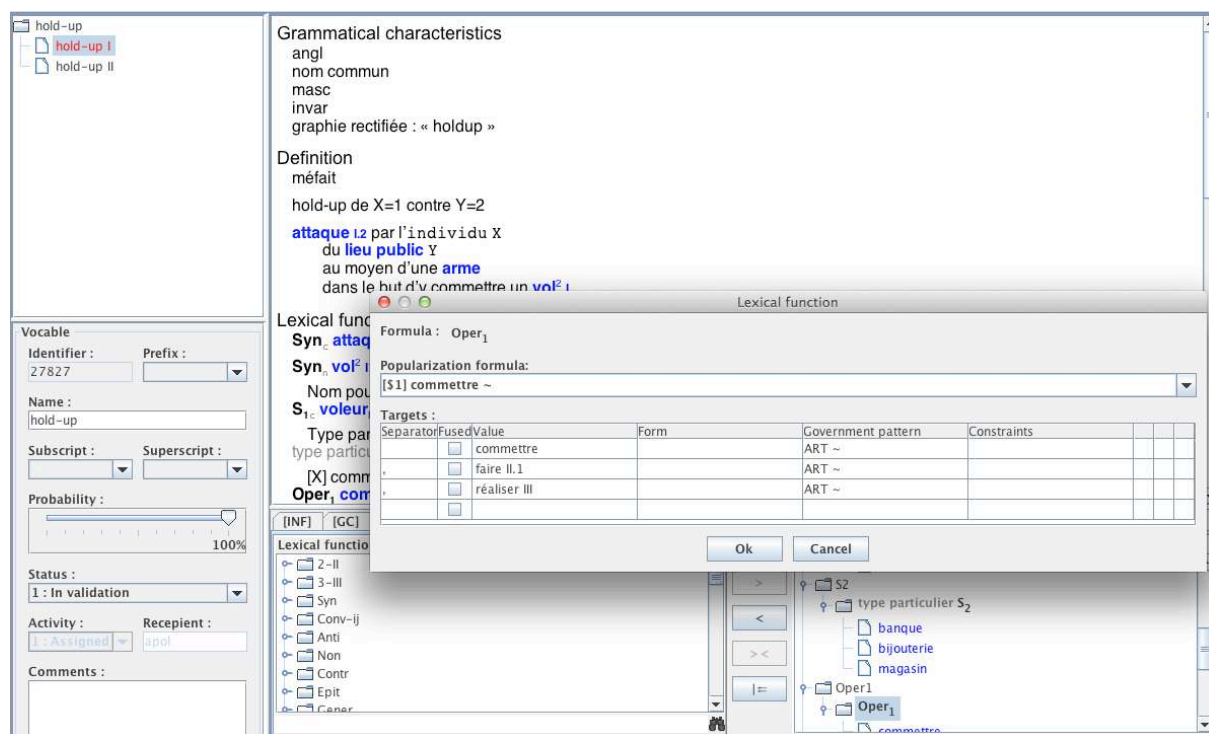


Figure 3. Editing Fr. **Oper**₁(HOLD-UP I) with Dicot

Elements in the **Value** column are not strings of characters but actual uniquely defined lexical nodes of the fr-LN. Lexicographers are establishing links between the headword and these nodes by browsing through the “node-list” of the graph and selecting the appropriate lexical target.

In the process of encoding a lexical function application such as **Oper**₁(HOLD-UP I), lexicographers can actually enter into an embedded lexicographic subprocess in order to perform one the following operations:

- edition of an existing target node—e.g. in order to weave a symmetrical lexical function link from the lexical target back to the headword: weaving **Syn**_n(HOLD-UP I) = BRAQUAGE ‘stick-up’ entails weaving **Syn**_n(BRAQUAGE) = HOLD-UP I.
- “on the fly” creation of an expected lexical target that does not already exist in the database—in which case, minimal lexicographic information is provided: part of speech, at least one (quasi-) synonymy link to an already existing lexical unit and at least one lexicographic example.

Though the lexicographic process is one of graph weaving rather than text writing, lexicographers need to be presented with a textual representation of linguistic information. It is thus essential that the lexicographic editor provides lexicographers with an immediate textual feedback about the current state of the headword modelling in the lexical system. This is performed by the systematic update of an *article-view* of the headword’s article, where information contained in the graph—as **virtual article**—is translated into a dictionary-like

text, with hyperlinks to targeted lexical units. This lexicographic device is illustrated in Figure 4 that shows a sample of the article-view for Fr. HOLD-UP I, as generated by Dicit.⁷

| | |
|------------------|---|
| hold-up I | <p>Caractéristiques grammaticales angl nom commun masc invar graphie rectifiée : « holdup »</p> <p>Définition méfait hold-up de X=1 contre Y=2 attaque I.2 par l'individu X du lieu public Y au moyen d'une arme dans le but d'y commettre un vol² I</p> <p>Fonctions lexicales Syn attaque I.2 Syn vol² I; fam braquage; fam casse Nom pour X S_{1c} voleur_{N, I}, voleuse I; bandit I Type particulier de Y type particulier S₂ banque; bijouterie; magasin [X] commettre ~ Oper₁ commettre [ART ~], faire II.1 [ART ~], réaliser III [ART ~] Ce que X dit quand il commet ~ Enun, Real, Haut les mains !, Les mains en l'air ! AntiBonFact₀ tourner mal</p> <p>Exemples Oui, il a lu le dossier, cet énorme pavé (que j'ai moi-même soigneusement parcouru, dans ma cellule, des jours entiers), où tous les témoignages, les interrogatoires, la reconstitution du hold-up ont été consignés, il y a plus de vingt ans. Frantext CASTEL Hélène, <i>Retour d'exil d'une femme recherchée</i>, 2009, p. 153 Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey. Frantext GOLDMAN Pierre, <i>Souvenirs obscurs d'un juif polonais né en France</i>, 1975, p. 258</p> |
|------------------|---|

Figure 4. Sample of the article-view for Fr. HOLD-UP I in Dicit

We call *lexicography of virtual dictionaries* the above-illustrated approach according to which the dictionary text is not written but is virtually encapsulated in the lexical model and can be automatically generated from it (Atkins 1996, Polguère 2012). It is best suited for building lexical systems because of their fundamentally non-textual nature.

Though lexical systems are essentially graphs of lexical units, it is possible to model “internal” properties of lexical nodes through graph weaving, exactly as for relational properties. Each internal property associated to a lexical node in the fr-LN is implemented as a pointer to an autonomous entity of the base. To indicate that a lexical unit is a common noun, possesses the masculine gender, belongs to familiar language, etc., boils down to weaving links between this lexical unit and metalinguistic entities that are stored in integrated grammatical or, more generally, metalinguistic models (for instance, a model for the system of Meaning-Text lexical functions). The fr-LN goes even further in that even lexicographic examples (citations from corpora or custom-made examples) are autonomous informational entities associated with one or more lexical units, each association identifying the specific segment of the example that corresponds to the exemplified lexical unit. It is possible to use the “lexicographic record” of a given example in order to access all lexical units that are making use of it. This is illustrated in Figure 5 that shows the lexicographic record of the second example that appears in Figure 4 (article-view for Fr. HOLD-UP I).⁸

ID : 470
 Class : Citations de corpus type Frantext
 Source : Frantext
 Status : Pending validation (1)
 Quotation
 Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey.
 Reference
 Frantext GOLDMAN Pierre, *Souvenirs obscurs d'un juif polonais né en France*, 1975, p. 258
 Occurrences

| | |
|------------|--|
| hold-up I | Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey. |
| malfaiteur | Il est par contre exact que j'ignorais (et ignore encore) l'identité des malfaiteurs qui m'ont accompagné lors des hold-up commis contre la société Vog et le payeur des Allocations familiales du passage Ramey. |

Created by debe on january 8, 2013 at 9:16:09 AM
 Updated by mvs on january 23, 2013 at 5:53:59 PM

Figure 5. Record of a lexicographic example in Dicet

This “inverted” navigation in the lexical system—from targeted informational entities back to lexical nodes pointing to them—is available for each piece of information that is mentioned in the microstructure of lexical nodes. For instance, one can open the lexicographic record of a given usage note, such as *vulg* (‘vulgar’), in order to consult the list of all lexical nodes to which this usage note has been associated.

3.2 Macrostructure level: walking and weaving the graph

The fr-LN is being developed around a preselected core vocabulary, called *priming wordlist*; it contains around 3,700 vocables. The birth of the lexical system took place as a process of automatically generating a fully disconnected graph (set of nodes with no arcs): one lexical node for each of the elements of the priming wordlist. Lexicographers then proceeded (i) to give flesh to these nodes and (ii) to manually build the so-called *directly induced wordlist*, i.e. lexical nodes that are close semantic derivatives (exact synonyms, exact verbalisations, nominalisations, etc.) of the elements of the priming wordlists. From this, the growth of the fr-LN proceeded and still proceeds as a *random but systematic process* that is mainly determined by the logical organisation of the system of lexical functions. For more details on the fr-LN growth, see Lux-Pogodalla and Polguère (2011) and Polguère and Sikora (2013).

Lexical system lexicographers follow the basic Explanatory Combinatorial Lexicology principle stating that lexicographic description should not focus on isolated vocables, but on sets of vocables that correspond to given *lexical fields*. In order to explain the notion of lexical field, we need to start with the more basic notion of *semantic field*.

The semantic field of M—for instance, the semantic field of smoking—is the set of **lexical units** (i.e. specific senses) that have in their definition a meaning M in a strategic position.

For instance, the French meaning ‘fumer’ (‘to smoke’) occupies an important place in the definition of Fr. CIGARETTE I (cf. section 2.2.2 above), which thus belongs to the semantic field of smoking. From the notion of **semantic** field, we can now derive that of **lexical** field.

The lexical field of M—cf. the lexical field of smoking—is a **set of potentially polysemic vocables** whose basic lexical unit (i.e. the sense that “controls” their polysemic structure) belongs to the semantic field of M.

For instance, the polysemic vocable Fr. CIGARETTE belongs to the lexical field of smoking because its basic lexical unit—CIGARETTE I—belongs to the semantic field of smoking. Such is not the case for the vocable Fr. HERBE ‘herb’ because only its metaphoric sense HERBE II ‘weed [marijuana]’ belongs to the semantic field of smoking. This latter vocable will thus be left aside by lexicographers working on the lexical field of smoking.

As explained in Polguère (2013), while semantic fields are playgrounds for semanticists, lexical fields are playgrounds for lexicographers because they are more suitable for handling the description of polysemy. Lexical system lexicographers will therefore work primarily with lexical fields and the macrostructure of lexical systems will evolve according to the need to describe lexical units found in each given lexical field under study. This process—typical of “traditional” Explanatory Combinatorial Lexicography—is in competition with another process that takes place at the level of microstructure construction.

While writing a dictionary—and Explanatory Combinatorial Dictionaries are no exception (Mel’čuk *et al.* 1984, 1988, 1992, 1999, Mel’čuk and Zholkovsky 1984)—, one can mention a lexical unit in a headword article and stop there. For instance, one can mention *commettre* ‘to commit’ in the dictionary entry for Fr. HOLD-UP I (cf. *commettre un hold-up*), together with other support verbs, and forget all about the lexical unit COMMETTRE and its lexicographic description. This is impossible in lexical network lexicography, where such information can only be entered if the target of a lexical link does exist in the model. “On the fly” creation of lexical units, described in section 3.1 above, is therefore an essential factor in the evolution of the lexical system macrostructure as lexicographers will have to venture outside their lexical field in order to draft lexical nodes that are required for holding a given microstructure together. Nodes created on the fly will eventually become recurrent meeting points for lexicographers working on different lexical fields. Once the density of a lexical node (its number of incoming and outgoing arcs) becomes high, the corresponding lexical unit acquires the status of *crossroad lexical unit* and the description of the vocable it belongs to becomes a lexicographic priority.

Interestingly, the fact that lexicographers can access many times the same lexical unit as lexical target of the headwords they work on results in the lexicographic drafting of crossroad lexical units. The set of all arcs leading to a crossroad lexical unit can be visualised, thus providing lexicographers with a *negative* of the article for this unit. This negative image will later help to elaborate the actual picture, i.e. the lexicographic article of the crossroad unit.

In conclusion, the growth of the macrostructure of a lexical system can be a rather intricate process. It includes a significant element of randomness conditioned by unpredictable graph walks lexicographers may have to take in the course of working in the microstructure. We think that this relative randomness is not necessarily a handicap as it does not forbid lexicographers from using some form of planning: identification and processing of strategic lexical fields, regular clean-up of vocables that gain importance because they contain crossroad lexical units, etc. It is also worth noting that this random but systematic growth presents some analogy with “natural” vocabulary acquisition by children and with tentative mathematical models of growth of lexical graphs (Dorogovtsev and Mendes 2001).

But we shall stop here, and refrain from claiming cognitive relevance. We do feel that lexical systems have something in common with the actual structure of lexical knowledge and

that weaving lexical graphs is a more “natural” activity than writing dictionaries. Further than that, we don’t know, and probably nobody knows (yet).

Acknowledgements

Work on the fr-LN takes place at the ATILF CNRS laboratory (Nancy, France) in the context of the RELIEF project funded by the Agence de Mobilisation Économique de Lorraine (AMEL) and the European Regional Development Fund (ERDF). RELIEF is a collaboration between ATILF and MVS Solutions (Sainte-Marguerite, France). We are very grateful to Monique Cormier, Candice Delaite, Christiane Fellbaum, Marie-Claude L’Homme and Dorota Sikora for their comments on a first version of this paper.

Notes

¹ In the Princeton WordNet, version 3.1, the path from *theft* to *to steal* is tortuous: sense `theft#1` is related to `stealing#1`, as it belongs to the same synset; this latter sense is itself connected to the verbal sense `steal#1` by a ‘derivationally related form’ link (Fellbaum *et al.* 2009).

² Part of the material presented below is adapted from Polguère (2014), in French.

³ Names of lexical units are written in small capitals, accompanied with bolded lexicographic numbers if senses need to be distinguished. Idioms are signalled by the two ‘...’ enclosing symbols.

⁴ The lexicographic numbering used for French lexical units in this paper is based on the current state of the fr-LN. The fact that no numbering accompanies *commettre* ‘to commit’ in (1) indicates that the polysemy of this verb has not yet been analysed in the fr-LN and that only the sense corresponding to an **Oper**₁ support verb is currently handled in the database (*commettre un hold-up, un péché* ‘a sin’, *une erreur* ‘a mistake’ ...).

⁵ Wiktionary: <http://www.wiktionary.org>; JeuxDeMots: <http://www.jeuxdemots.org/jdm-accueil.php>.

⁶ The Dicet editor has been designed jointly by the fr-LN lexicographic team and N. Gader (MVS Publishing Solutions), who did the actual programming.

⁷ The edition of government patterns (roughly, subcategorisation frames) is not implemented in Dicet for the time being.

⁸ The collection of all examples used in the fr-LN makes up an embedded linguistic corpus (Lux-Pogodalla 2014). At the time of writing, this corpus has reached the non-trivial size of over one million word occurrences. It is consulted by lexicographers by means of the TXM corpus processing software (Heiden 2010). The embedded example corpus is of course not a “text.” It should not be mistaken with the reference corpora used by lexicographers—such as Frantext (Montémont 2008)—from which citations are extracted for insertion into the fr-LN database.

References

A. Dictionaries

Mel’čuk, I. et al. 1984, 1988, 1992, 1999. *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Volumes I-IV.* Montreal: Les Presses de l’Université de Montréal.

Mel’čuk, I. and A. Polguère. 2013. *DiCo lexical database.* On-line “Dicouèbe” version, OLST, Department of Linguistics and Translation, Université de Montréal [<http://olst.ling.umontreal.ca/dicouebe>].

Mel’čuk, I. and A. Zholkovsky. 1984. *Explanatory Combinatorial Dictionary of Modern Russian. Semantico-syntactic Studies of Russian Vocabulary.* Vienna: Wiener Slawistischer Almanach.

Picoche, J. and J.-C. Rolland. 2002. *Dictionnaire du français usuel. 15 000 mots utiles en 442 articles.* Brussels: De Boeck-Duculot.

Summers, D. (ed.). 1993. *Longman Language Activator. The World's First Production Dictionary.* Harlow: Longman Group.

B. Other literature

Aitchison, J. 2012. *Words in the Mind: An Introduction to the Mental Lexicon.* 4th edition, Oxford: Wiley-Blackwell.

Atkins, B. T. S. 1996. 'Bilingual Dictionaries: Past, Present and Future' In M. Gellerstam, J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C. R. Pappmehl (eds), *Euralex '96 Proceedings.* Gothenburg University, Department of Swedish, 515–590.

Baronchelli, A., Ferrer i Cancho, R., Pastor-Satorras, R., Chater, N. and M. H. Christiansen. 2013. 'Networks in Cognitive Science.' *Trends in Cognitive Sciences* 17:7, 348–360.

Chudy, Y., Desalle, Y., Gaillard, B., Gaume, B., Magistry, P. and E. Navarro. 2013. 'Tmuse: Lexical Network Exploration.' *The Companion Volume of the Proceedings of IJCNLP 2013: System Demonstrations,* Asian Federation of NLP, Nagoya, 41–44.

de Schryver, G.-M. 2003. 'Lexicographers' Dreams in the Electronic-Dictionary Age.' *International Journal of Lexicography* 16.2: 143–199.

Dorogovtsev, S. N. and J. F. F. Mendes. 2001. 'Language as an evolving word web.' *Proceedings of the Royal Society of London. Series B, Biological sciences* 268: 2603–2606.

Fellbaum, C., A. Osherson and P. E. Clark. 2009. 'Putting Semantics into WordNet's "Morphosemantic" Links.' In Z. Vetulani and H. Uszkoreit (eds), *Responding to Information Society Challenges: New Advances in Human Language Technologies.* Springer Lecture Notes in Informatics 5603, Dordrecht: Springer, 350–358.

Fellbaum, C. and P. Vossen. 2012. 'Challenges for multilingual wordnet.' *Language Resources and Evaluation* 46: 313–326.

Fillmore, C. J., C. R. Johnson and M. R. L. Petruck. 2003. 'Background to FrameNet.' *International Journal of Lexicography* 16.3: 235–250.

Fontenelle, T. 1997. *Turning a bilingual dictionary into a lexical-semantic database.* Tübingen: Niemeyer.

Gader, N., V. Lux-Pogodalla and A. Polguère. 2012. 'Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor' In *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III).* Mumbai: The COLING 2012 Organizing Committee, 109–125.

Gader, N., Ollinger, S. and A. Polguère. 2014. 'One lexicon, two structures: So what gives?' In H. Orav, C. Fellbaum and P. Vossen (eds), *Proceedings of the Seventh Global Wordnet Conference (GWC2014).* Tartu: Global WordNet Association, 163–171.

Gaume B. 2004. 'Balades aléatoires dans les Petits Mondes Lexicaux.' *I3 Information Interaction Intelligence* 4.2: 39–96.

Gaume B. 2008. 'Mapping the Forms of Meaning in Small Worlds.' *Journal of Intelligent Systems* 23: 848–862.

Gaume, B., K. Duvignau, L. Prévot and Y. Desalle. 2008. 'Toward a cognitive organization for electronic dictionaries, the case for semantic proxemy' In *Proceedings of the workshop on Cognitive Aspects of the Lexicon (COGALEX 2008).* Manchester, 86–93.

Gaume B., Venant F. and B. Victorri. 2006. 'Hierarchy in Lexical Organisation of Natural Languages.' In D. Pumain (ed), *Hierarchy in Natural and Social Sciences.* Methodos series 3, Dordrecht: Springer, 121–142.

- Gouws, R. H. and D. J. Prinsloo. 1998.** ‘Cross-Referencing as a Lexicographic Device.’ *Lexikos* 8: 17–36.
- Gruenenfelder, T. M., Pisoni, D. B. 2009.** ‘The Lexical Restructuring Hypothesis and Graph Theoretic Analyses of Networks Based on Random Lexicons.’ *Journal of Speech, Language, and Hearing Research* 52, 596–609.
- Hearst, M. A. 1992.** ‘Automatic Acquisition of Hyponyms from Large Text Corpora’ In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 1992)*, Nantes, 539–545.
- Heiden, S. 2010.** ‘The TXM Platform: Building Open-Source Textual Analysis Software Compatible with the TEI Encoding Scheme’ In R. Otaguro, K. Ishikawa, H. Umemoto, K. Yoshimoto and Y. Harada (eds), *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*. Sendai, 389–398.
- Hirst, G. 2009.** ‘Ontology and the Lexicon’ In S. Staab and R. Studer (eds), *Handbook on Ontologies*. 2nd edition, International Handbooks on Information Systems Series, Berlin & Heidelberg: Springer, 269–292.
- Kilgarriff, A. and P. Rychlý. 2010.** ‘Semi-automatic Dictionary Drafting’ In G.-M. de Schryver (ed), *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers, 299–312.
- Lafourcade, M. 2007.** ‘Making people play for Lexical Acquisition.’ *Proceedings of SNLP 2007, Seventh Symposium on Natural Language Processing*. Pattaya.
- Lew, R. 2012.** ‘How can we make electronic dictionaries more effective?’ In S. Granger and M. Paquot (eds), *Electronic lexicography*. Oxford: Oxford University Press, 343–361.
- Lux-Pogodalla, V. 2014.** ‘Intégration relationnelle des exemples lexicographiques dans un réseau lexical’ In *Actes de la 21^e Conférence sur le Traitement Automatique du Langage Naturel (TALN 2014)*, Marseille.
- Lux-Pogodalla, V. and A. Polguère. 2011.** ‘Construction of a French Lexical Network: Methodological Issues’ In *Proceedings of the First International Workshop on Lexical Resources, WoLeR 2011. An ESSLLI 2011 Workshop*. Ljubljana, 54–61.
- Mel’čuk, I. 1996.** ‘Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon’ In L. Wanner (ed), *Lexical Functions in Lexicography and Natural Language Processing*. Language Companion Series 31, Amsterdam/Philadelphia: John Benjamins, 37–102.
- Mel’čuk, I. 2006.** ‘Explanatory Combinatorial Dictionary’ In G. Sica (ed), *Open Problems in Linguistics and Lexicography*. Monza: Polimetrica, 225–355.
- Mel’čuk, I. 2012.** *Semantics: From meaning to text*. Volume 1, Studies in Language Companion Series 129, Amsterdam/Philadelphia: John Benjamins.
- Mel’čuk, I. 2013.** *Semantics: From meaning to text*. Volume 2, Studies in Language Companion Series 135, Amsterdam/Philadelphia: John Benjamins.
- Mel’čuk, I., A. Clas and A. Polguère. 1995.** *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve: Duculot.
- Miller, G. A. 1990.** ‘Nouns in WordNet: A Lexical Inheritance System.’ *International Journal of Lexicography* 3.4: 245–264.
- Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross and K. J. Miller. 1990.** ‘Introduction to WordNet: An On-line Lexical Database.’ *International Journal of Lexicography* 3.4: 235–244.
- Montémont, V. 2008.** ‘Discovering Frantext’ In J. Auracher and W. van Peer (eds), *New Beginnings in Literary Studies*. Newcastle: Cambridge Scholars Publishing, 89–107.

- Morais, A. S., H. Olsson and L. J. Schooler. 2013.** ‘Mapping the Structure of Semantic Memory.’ *Cognitive Science* 37: 125–145.
- Motter, A. E., de Moura, A. P. S., Lai, Y.-C. and P. Dasgupta. 2002.** ‘Topology of the conceptual network of language.’ *Physical Review E* 65.
- Navigli, R. and S. P. Ponzetto. 2012.** ‘BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network.’ *Artificial Intelligence* 193: 217–250.
- Ploux, S. 1997.** ‘Modélisation et traitement informatique de la synonymie.’ *Linguisticæ Investigationes* 21.1: 1–27.
- Polguère, A. 2007.** ‘Lexical function standardness’ In L. Wanner (ed), *Selected Lexical and Grammatical Issues in the Meaning-Text Theory. In Honour of Igor Mel’čuk*. Language Companion Series 84, Amsterdam/Philadelphia: John Benjamins, 43–95.
- Polguère, A. 2009.** ‘Lexical systems: graph models of natural language lexicons.’ *Language Resources and Evaluation* 43.1: 41–55.
- Polguère, A. 2011.** ‘Classification sémantique des lexies fondée sur le paraphrasage.’ *Cahiers de lexicologie* 98: 197–211.
- Polguère, A. 2012.** ‘Lexicographie des dictionnaires virtuels’ In Y. Apresjan, I. Boguslavsky, M.-C. L’Homme, L. Iomdin, J. Milićević, A. Polguère and L. Wanner (eds), *Meanings, Texts, and Other Exciting Things. A Festschrift to Commemorate the 80th Anniversary of Professor Igor Alexandrovič Mel’čuk*. Studia Philologica, Moscow: Jazyki slavjanskoj kultury Publishers, 509–523.
- Polguère, A. 2013.** ‘Les petits soucis ne poussent plus dans le champ lexical des sentiments’ In F. Baider and G. Cislaru (eds), *Cartographie des émotions. Propositions linguistiques et sociolinguistiques*. Paris: Presses Sorbonne Nouvelle, 21–41.
- Polguère, A. 2014.** ‘Principes de modélisation systémique des réseaux lexicaux’ In B. Bigi (ed), *Actes de la 21^e Conférence sur le Traitement Automatique du Langage Naturel (TALN 2014)*, Marseille, 79–90.
- Polguère, A. and D. Sikora. 2013.** ‘Modèle lexicographique de croissance du vocabulaire fondé sur un processus aléatoire, mais systématique’ In C. Garcia-Debanc, C. Masseron and C. Ronveaux (eds), *Enseigner le lexique*. Namur: Presses Universitaires de Namur, 35–63.
- Ruppenhofer, J., M. Ellsworth, M. R. L. Petruck, C. R. Johnson and J. Scheffczyk. 2010.** *FrameNet II: Extended Theory and Practice*. Berkeley: International Computer Science Institute.
- Sajous, F., Navarro, E. and B. Gaume. 2011.** ‘Enrichissement de lexiques sémantiques approvisionnés par les foules: le système WISIGOTH appliqué à Wiktionary.’ *Traitement Automatique des Langues (T.A.L.)* 52.1: 11–35.
- Sagot, B. and D. Fišer. 2008.** ‘Building a free French wordnet from multilingual resources.’ *Proceedings of OntoLex 2008*. Marrakech.
- Sowa, J. F. 2010.** ‘The Role of Logic and Ontology in Language and Reasoning’ In R. Poli and J. Seibt (eds), *Theory and Applications of Ontology: Philosophical Perspectives*. Dordrecht et al.: Springer, 231–263.
- Vossen, P. (ed.). 1998.** *EuroWordNet: a multilingual database with lexical semantic networks for European Languages*. Dordrecht: Kluwer.
- Watts, D. J. and S. H. Strogatz. 1998.** ‘Collective dynamics of ‘small-world’ networks.’ *Nature* 393: 440–442.
- Wierzbicka, A. 2009.** ‘The Theory of the Mental Lexicon’ In S. Kempgen, P. Kosta, T. Berger and K. Gutschmidt (eds), *Die slavischen Sprachen/The Slavic Languages: An*

International Handbook of their Structure, their History and their Investigation. Berlin & New York: Mouton de Gruyter, 848–863.

Wolter, B. 2006. ‘Lexical Network Structures and L2 Vocabulary Acquisition: The Role of L1 Lexical/Conceptual Knowledge.’ *Applied Linguistics* 27.4: 741–747.