

Fishing out regular polysemy with the *Gener* (generic) relation

Alain Polguère

Université de Lorraine, CNRS, ATILF

alain.polguere@univ-lorraine.fr

Abstract

Lexical Systems are formal network models of natural language lexicons. They are characterized by their graph structure that is mainly woven by paradigmatic and syntagmatic lexical relations known as lexical functions. Being organized as “social networks of words,” Lexical Systems are radically distinct from taxonomic models which organize lexical units within hierarchized classes. In spite of their apparent messiness, these structures are perfectly suited for the storage of lexical regularities. This paper is devoted to the *Gener* (generic) lexical function and its significance for the identification and modeling of regular polysemy in network models of the Lexical System family. We show how the systematic lexicographic encoding of *Gener* relations allows for the study of polysemy patterns and the uniform modeling of this phenomenon in lexicographic models. The approach is illustrated based on the lexicographic construction of a Lexical System – the French Lexical Network (fr-LN) –, using the generic term Fr. *fleur* ‘flower’.

Keywords: regular polysemy, Lexical System, lexical function, generic, lexicographic definition

Résumé

Les Systèmes Lexicaux sont des modèles formels réticulaires des lexiques. Ils se caractérisent par leur structure de graphe tissée en premier lieu par les relations lexicales paradigmatiques et syntagmatiques appelées *fonctions lexicales*. Du fait de leur organisation en « réseaux sociaux de mots », les Systèmes Lexicaux se distinguent radicalement des modèles taxonomiques, qui organisent les unités lexicales en classes hiérarchisées. En dépit de leur désordre apparent, ces structures sont parfaitement adaptées au stockage de régularités lexicales. Le présent texte est consacré à la fonction lexicale *Gener* (générique) et à sa pertinence pour l’identification et la modélisation de la polysémie régulière dans les modèles réticulaires de la famille des Systèmes Lexicaux. Nous montrons comment l’encodage lexicographique systématique des relations *Gener* permet l’étude de patrons de polysémie et la modélisation uniforme de ce phénomène dans les modèles lexicographiques. L’approche est illustrée à partir de la construction lexicographique d’un Système Lexical – le Réseau Lexical du Français (RL-fr) –, en s’appuyant sur le terme générique *fleur*.

Mots-clés : polysémie régulière, Système Lexical, fonction lexicale, générique, définition lexicographique

1. Working with network models of natural language lexicons

This paper addresses the topic of regular polysemy from a lexicographic perspective, more precisely from the perspective of lexical network lexicography.¹ The present section lays the ground for the discussion, with a general characterization of natural language lexical networks (Section 1.1), some considerations about lexical functions that are at the core of our methodology (Section 1.2) and the relational perspective on polysemy adopted here (Section 1.3). Later, we proceed with a detailed presentation of the *Gener* lexical function (Section 2) whose interest in the exploration and modeling of regular polysemy we intend to demonstrate (Section 3).

1.1. Shape and size of the lexicon in lexical network terms

We are using as theoretical and descriptive framework *Explanatory Combinatorial Lexicology and Lexicography* (Mel'čuk, 2013: ch. 11) augmented by the lexicography of *Lexical Systems*: an approach to the construction of lexical models where lexical networks rather than dictionaries or other “textual” representations of lexical knowledge are targeted (Polguère, 2014). Lexical Systems are network models of natural language lexicons whose graph structure belongs to the family of *small-world networks* (Watts & Strogatz, 1998). Such lexical models can be formally characterized in layperson terms as “social networks” of lexical units: huge graphs of dispersed nodes – lexical units – connected by arcs – lexical relations – whose number is small in comparison with the number of nodes. Expressed in mathematical terms, such graphs have a low *density*, i.e., a low average *node degree* – a node's degree being the number of connecting arcs for this node. Within this globally dispersed structure, many clusters are formed of mutually interconnected nodes – “parties” of lexical units. Additionally, the existence of highly connected nodes – nodes with exceptionally high density corresponding to lexical units with many “friends” – ensures that within this widely dispersed structure, there is always a relatively short path (sequence of arcs) that connects any two nodes. Hence the “small world” metaphor. Let us now estimate the size of the lexical network of natural languages; the reasoning at this stage is based on the English language.

As for lexical nodes, standard general-public dictionaries (*American Heritage*, *Oxford Learners' dictionary*, etc.) can help estimate the size of the basic English lexicon. To make things simple, let us consider that the wordlist of such dictionaries may contain around 60,000 entries; with a conservative average polysemy of three senses per entry, we obtain an estimated size of 180,000 lexical units that constitute a significant and representative part of the English lexicon. A minimal

¹ In preparing the final version of this text, we have greatly benefited from comments made by Lucie Barque, Pauline Hass, Veronika Lux-Pogodalla, Igor Mel'čuk, Polina Mikhel and two anonymous reviewers of *Lexique*. We did our best to take their feedback into consideration.

lexical graph of the English language should therefore contain at least 180,000 nodes – a node is a specific word sense, not a polysemous word. This is a rough number that gives us a minimal order of magnitude. Additionally, entries in most commercial dictionaries are lexemic (i.e., single word) entities and the description of idioms is embedded in lexemes' articles. Idioms are nonetheless full-fledged lexical units and they number many thousands in all natural languages. Consequently, we should revise our largely underestimated calculation up to a still conservative 250,000 lexical units. Though this number may seem huge from the perspective of manual lexicographic modeling, it is relatively small in comparison to the node cardinality of graphs representing other types of natural entities: e.g., a graph of the world's population or a graph of all celestial objects in the galaxy (this latter having literally and metaphorically astronomical proportions).

As for the number of lexical relations, doing an estimate is trickier than for lexical nodes, for two reasons. Firstly, lexical units can entertain very diverse relations: formal, semantic, combinatorial, etc. Secondly, in line with the characteristics of small-world networks (see first paragraph of the present section), some lexical units can entertain a large number of direct relations – e.g., the noun *FEELING* and the verb *DO* – while others do not occupy such strategic structural position of “crossroad” lexical units (see, Section 2.2). Let us then take a purely formal strategy to solve the problem of estimating the number of lexical relations in lexicons. If we take as axiom the fact that lexical networks are small-world networks, it is possible to use the mathematical properties of such graphs to approximate the number of relations in lexicons once we have an estimate for the total number of nodes: let us call this latter N . The average degree (number of connecting arcs) of a node in a small-world network is estimated (see e.g., Gaume *et al.*, 1988, p. 88) to have a maximal value of $\ln(N)$.² We just made an estimate of the value of N by reasoning on the English language: 250,000 lexical nodes. We can now estimate the number of arcs with a simple formula: $\ln(250,000) \times 250,000$. If we round up the returned value,³ we can conclude that the estimated size of a complete Lexical System model of English, or of any natural language, is thus: 250,000 nodes and 3,110,000 arcs. It is big, but not **that** big.

The conclusion one can draw from the above reasoning is that natural language lexical networks are vast and rich structures but their manual (lexicographic) construction remains a feasible task provided lexicographers develop special computational tools for weaving and processing them (Gader *et al.*, 2012).

Basic terms and writing conventions. By *lexical unit*, we mean a unit of the lexicon and, consequently, a unit of lexicographic description associated with a well-specified meaning;

² Natural log of N .

³ A more precise value is 3,107,304.

formally, lexical units can be either *lexemes* or *idioms*. Names of lexical units are written in small capitals, eventually accompanied by a lexicographic number in case different senses of a polysemous *vocable* have to be distinguished – e.g., MOON_(N) I.1 ‘satellite of the earth’, MOON_(N) I.2 ‘satellite of a planet’, MOON_(N) II ‘what is seen of the moon_(N) I.1 in the sky’, etc. are lexemes of the polysemous vocable MOON_(N). Names of idioms are written between ‘...’ to highlight the fact that these lexical units are semantically non-compositional expressions that function as a whole – e.g., ‘ASK FOR THE MOON’ ‘have exaggerated expectations’. Names of *lexical functions* (see 1.2) are written in a monospaced font – e.g., Gener, Magn, etc. Finally, names of copolysemy relations (see, Section 1.3) are written in a **bold monospaced font** – e.g., **Extension**, **Metonymy**, **Metaphor**, etc.

1.2 Remarks on the universal system of simple standard lexical functions

The lexical function *Gener* is at the core of the present work. Before we deal specifically with it (Section 2), it is important to briefly recapitulate the purpose and characteristics of the system of simple standard lexical functions. We presuppose in what follows a minimal familiarity with lexical functions on the part of the reader, one that can be achieved by studying, for instance, Wikipedia’s “lexical function” entry⁴ or any succinct presentation such as Mel’čuk (2007). In other words, we do not offer a new “lexical function in a nutshell” presentation but we try to increase the reader’s familiarity with lexical functions by proposing an operational definition of the notion suited for the present paper.

At the time of writing, Mel’čuk and Polguère (2021) is the latest detailed presentation of the full system of standard lexical functions.⁵ We base the content of the present section on the explanations and presentation strategies adopted in this text.

Let us consider an elementary situation of language production where a Speaker has lexicalized a meaning with a given lexical unit *L* and needs to express another meaning ‘*σ*’ in the sentence in combination with the meaning of *L*. Two main cases can be envisaged: (i) either ‘*σ*’ is a “normal” meaning and its lexicalization can be performed by the Speaker independently from *L* itself – e.g., HORSE + ‘beautiful’ → *beautiful/gorgeous/magnificent/... horse*; (ii) or ‘*σ*’ belongs to a small set of meanings that are flagged by natural languages as having to be lexicalized contingent upon a preexisting lexicalization it is to be combined with – e.g., HORSE + **‘to make use [of]’** → ***to mount/to ride a horse***.

⁴ https://en.wikipedia.org/wiki/Lexical_function.

⁵ A much expanded presentation of standard lexical functions, in English, is currently being written in the form of a book (Mel’čuk & Polguère, in preparation). We allow ourselves to borrow a few ideas from this upcoming publication in the discussion of *Gener* (Section 2).

The idea that there exists a small set of meanings that are universally (i.e., in all natural languages) expressed contingent upon preexisting lexicalizations is at the heart of the system of standard lexical functions as stated in the definition of the notion offered by Mel'čuk and Polguère (2021, p. 81):⁶

[...] any LF [= Lexical Function] f possesses a meaning of the second type – noted ' σ^f ' – whose expression is contingent upon the preliminary choice of a lexical unit L . The LF f takes L as argument – which is noted $f(L)$ – and returns as value the set of all expressions of ' σ^f ' that are suitable relative to L .

Meanings of **standard** lexical functions – hence standard lexical functions themselves – have the three following properties.

PROPERTY 1. They are vague (i.e., poor) and therefore compatible with a very large, heterogeneous set of lexical units; e.g., the meaning of the $Real_1$ lexical function – 'to make use [of L]/to realize [L]/...' or, more generally, 'to do what is expected to be done [relative to L]' – is compatible with lexemes that have semantically nothing in common such as HORSE and DESIRE_(N):

$Real_1(horse) = to\ mount, to\ ride$ [ART ~]
 $Real_1(desire_{(N)}) = to\ fulfill, to\ satisfy$ [ART ~]

PROPERTY 2. They are expressed by a large set of distinct lexical values, though semantic classes of lexical units may be associated with "joker" values for specific lexical functions – e.g., the $Real_1$ to play for names of musical instruments in English (*to play the guitar, the piano, ...*).

PROPERTY 3 (semiotic consequence of Properties 1 and 2). Because they are basic items of linguistic expression, such meanings have a special semiotic status: (i) they are universally expressed in all natural languages; (ii) these latter tend to offer morphological means of expression for such meanings.

Although lexical functions were originally discovered while reflecting on the problem of translating collocations (see, Mel'čuk, 2015, p. 269), it soon became obvious that the notion applies not only to syntagmatic (i.e., combinatorial) lexical relations but equally to paradigmatic (i.e., semantic) relations. Synonymy (Syn), antonymy (Anti) or typical names for actants ($S_{1/2}/...$), to cite only a few examples, correspond to *semantic derivations* (Mel'čuk, 2007, Section 2) and can also be accounted for by lexical functions. In prototypical (or true) cases of $L_1 \Rightarrow L_2$ semantic derivation, L_2 is a semantic derivative of L_1 if (i) it is defined in terms of L_1 and (ii) the semantic

⁶ Our translation.

difference between L_2 and L_1 corresponds to a meaning that possesses all three above-mentioned properties of standard lexical function meanings.

The system of standard lexical functions can thus be divided into two subsets:

- *paradigmatic lexical functions* that correspond to semantic derivations;
- *syntagmatic lexical functions* that correspond to base-collocate relations.

For lack of space – and also because there already exists an extensive literature on this topic –, we will not go further into the discussion of the notion of lexical function. Our goal in this section was to provide the reader with essential information that is relevant for the rest of the discussion.

1.3 Relational model of polysemy

Now that we have dealt with the topic of lexical functions, let us examine the second notion that is at the heart of the present study: *polysemy* and its relational modeling in Lexical Systems. This is treated in detail in Polguère (2018) and we will simply recall here essential elements of our relational approach to polysemy.

To start with, it is worth mentioning that the traditional, most common view of polysemy is based on a decoding (rather than encoding) perspective,⁷ where the polysemy “of a word” is explained in terms of *ambiguity* of a lexical form; e.g., Gries (2015, p. 472):

The probably most widely accepted definition of polysemy is as the form of ambiguity where 2 + related senses are associated with the same word; consider the meanings of *glass* in *I emptied the glass* (‘container’) and *I drank a glass* (‘contents of the container’).

We shall neither accept nor adopt such definition because the fact for a given lexical signifier to be ambiguous is nothing but a practical consequence of a vocable’s polysemy; ambiguity is by no means a definitional notion as regards to polysemy.

First of all, let us establish the fundamental distinction between the polysemy of a vocable, that is the **property of a vocable** to be a set of more than one lexical unit – e.g., the polysemy of the $MOON_{(N)}$ vocable previously mentioned (final paragraph of Section 1.1) –, and polysemy as a **relation between two word senses** of a polysemous vocable such that one “is accounted for” (in the most general sense) relative to the other.

⁷ For a criticism of the decoding approach in linguistics (qualified as “upside-down approach”), see Polguère (2015, Section 3.2).

We call *copolysemy* the relation holding between two lexical units L_1 and L_2 of the same polysemous vocable, and we represent it with the following formula: $L_1 \mapsto L_2$. Lexical units involved in such relation are mutual *copolysemes*.

This being stated, it becomes obvious that what really matters in polysemy, as an immanent lexical phenomenon, is the existence of copolysemy relations throughout the structure of natural language lexicons. Each polysemous vocable is in fact a polysemy structure made up of lexical units connected by copolysemy relations. Figure 1 visualizes the polysemy structure of the $MOON_{(N)}$ vocable.

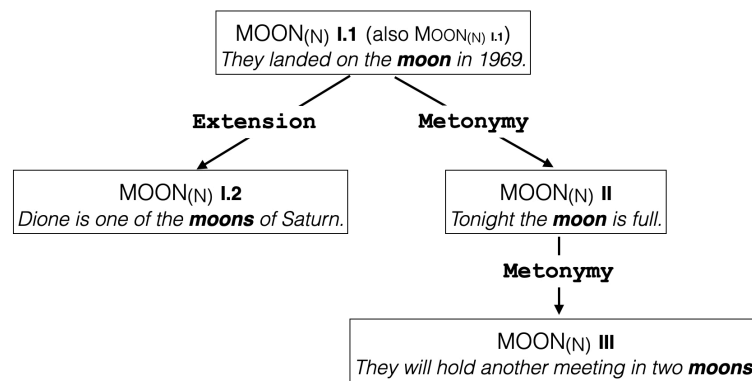


Figure 1. Polysemy structure of $MOON_{(N)}$ (simplified).

All lexemes in Figure 1 are copolysemes: they are connected either directly or indirectly by copolysemy relations. The lexeme that is at the top of such hierarchical structure is called the *basic lexical unit* of the vocable. This figure shows that, in our approach to modeling polysemy structures, each copolysemy relation $L_1 \mapsto L_2$ is systematically labeled in order to identify the type of copolysemy relation that connects L_2 to L_1 .

Formally, the structure given in Figure 1 is a (tiny) subnetwork of the Lexical System of English, where copolysemy relations cohabit with lexical function relations, among others. They play an important role in the topology of each Lexical System as they participate in its small-world structure by connecting areas of the graph that are otherwise topologically rather distant; for instance, in the polysemy of $MOON_{(N)}$, the areas corresponding to the semantic fields of celestial bodies (to which $MOON_{(N)}$ I.1 belongs) vs. time periods ($MOON_{(N)}$ III). This is even more so when copolysemes are connected by the relation of **Metaphor** – e.g., LOTUS I ‘flower’ \mapsto LOTUS II ‘yoga position’.

Copolysemy graphs such as Figure 1 make clear why we refer to our approach to polysemy as being *relational*. Though it may seem obvious that polysemy is based on relations, it is worth noting that the standard approach to its modeling is *enumerative* and not relational: dictionaries – including *Explanatory Combinatorial Dictionaries* such as Mel’čuk *et al.* (1984, 1988, 1992, 1999) – and other types of lexical resources simply list senses under entries for polysemous vocables without explicitly accounting for their relational structure. Two types of indications are used to vaguely guide dictionary users toward polysemy structures that remain largely implicit:

- a more or less elaborate system of sense identification signs – Roman/Arabic numbers and/or various symbols (•, ◇, etc.) – that weakly reflects an implicit hierarchical organization of the vocable without systematically identifying copolysemy relations;
- the use of indications such as *by metaph.*, *by meton.*, *by ext.*, etc., that are often ambiguous as regards to the exact source of the copolysemy relation and are used very sporadically throughout the dictionary.

In spite of the recourse to such techniques, the standard lexicographic approach to modeling polysemy remains largely enumerative and non-relational.

Note that we adhere to a strictly synchronic approach. We do not rely on the historical development of vocables to identify copolysemy relations between their senses. For each identified $L_1 \mapsto L_2$ relation, the diagnosis is based on a synchronic analysis of the meanings of both L_1 and L_2 , by drafting their respective *lexicographic definition*, and on a measurement of their semantic relation – for an extensive exploitation of lexicographic definitions in the study of (regular) polysemy, see Barque (2008). Of course, if no such relation between definitions is found, we are faced with a case of *homonymy* rather than copolysemy.⁸

The model of copolysemy relations presented in Polguère (2018) was built inductively based on the French lexicon through the lexicographic construction of a Lexical System for that language: the *French Lexical Network*, hereafter *fr-LN*. At the time of writing, twelve types of copolysemy relations have been identified, with additional subtypes. The fr-LN contains 29,463 lexical nodes grouped into 18,720 vocables and 9,569 copolysemy links have been woven.⁹ We can safely say that we already have at our disposal enough linguistic data on copolysemy relations to undertake large-scale studies on polysemy, and regular polysemy in particular.

At this point in our exposé, we have achieved the presentation of the core notions of network lexicography that underlie our proposal for exploring regular polysemy. We can now proceed with the examination of the *Gener* lexical function (Section 2) and its potential for studying and modeling regular polysemy (Section 3).

⁸ There are special cases of copolysemy relations where a connection between lexical definitions of related senses cannot be found. We cannot enter into such detail here; for more information, see the two relations of **Grammaticalization** and **Coderivation** in Polguère (2018, Section. 3).

⁹ By comparison, the fr-LN contains 63,815 lexical function links. Current statistics on Lexical Systems developed at ATILF are made available by Sandrine Ollinger on her website – URL: <https://perso.atilf.fr/sollinge/ressources/systemes-lexicaux/>.

2. The Gener lexical function

The *Gener* lexical function belongs to the family of paradigmatic lexical functions. A lexical unit that is a *Gener*(L) is a *generic term* for L. As stated in Mel'čuk and Polguère (in preparation), “it stands for the name of one of the *semantic classes* to which L's denotation belongs”; for instance:

Gener(*daffodil*) = *flower*_(N)
Gener(*crawl*_(V)) = *move*_(V)
Gener(*visual*) = *perceivable*

Just like *Syn* (synonym), *Gener* applies to lexical units of all parts of speech (Verbs, Nouns, Adjectives, Adverbs and Clausatives) and returns values of the same part of speech, though it is mostly relevant to nouns because large clusters of lexical units forming denotational classes develop mainly among nouns. Unlike *Syn*, the mother of all paradigmatic lexical functions, *Gener* possesses a borderline status for at least two reasons that we shall detail using properties of lexical functions introduced in Section 1.2. Firstly, a prototypical *Gener*(L) is a hyperonym of L¹⁰ and is therefore included in the meaning of L : ‘tulip’ = ‘flower_{*Gener*(tulip)} [...]’, ‘to crawl’ = ‘to move_{*Gener*(to crawl)} [...]’, etc. It is therefore clearly not a semantic derivative of L. Secondly, *Gener*, unlike other standard lexical functions, is not associated to a given semantic content ‘σ^ε’ (see Section 1.2 above). Whatever content is associated to *Gener* in *Gener*(L) depends entirely on the meaning of L.

Clearly, *Gener* occupies a special place in the system of standard lexical functions within which it appears isolated.¹¹ The aim of this section is to characterize the “true nature” of the *Gener* lexical relation (Section 2.1) and to explain its structuring role in Lexical Systems (Section 2.2).

2.1. Linguistic rather than “conceptual” foundation

Lexical units that function as values for *Gener* are lexical classifiers. They are directly related to “concepts” that Wierzbicka (1985, ch. 3) identifies as *categories of categories*: a significant number of lexical units denote specific instances of these concepts. As Wierzbicka (1985, p. 189) puts it:

We can safely assume that *tree* is conceptualized in English as a category of categories because there are many other words in English which clearly designate ‘a kind of tree’, for example *oak*, *maple* or *fir*. Similarly, we can safely assume that *bird* is conceptualized in English as a ‘category

¹⁰ We will see in Section 2.1 that it is not necessarily the case.

¹¹ The *Redun* (redundant modifier) syntagmatic lexical function (Mel'čuk & Polguère, 2021, Section 3.2, 121) shares classifying properties with *Gener* that are studied in Polguère (to appear). It is quite possible that the combined use of *Gener* and *Redun* can be a substitute for so-called semantic labels used for the semantic classification of lexical units within Lexical Systems (Polguère, 2011).

of categories' because there are many words in English which clearly designate 'a kind of bird', for example *swallow*, *magpie* or *parrot*. The same holds for the concept *fish* (see *cod*, *trout* or *herring*), *animal* (see *dog*, *cat* or *squirrel*) and *flower* (see *rose*, *tulip* or *carnation*).

Clearly, one would not expect a lexeme such as NAPKIN to be a value for *Gener* as this would require the English language to possess multiple names for kinds of napkins. But *Gener* is strictly a **lexical** function; it operates independently of anthropological perspectives on natural languages. Consequently, there is nothing that theoretically forbids a lexeme such as NAPKIN to be a *Gener*. It could be if we consider the terminology of the industry of napkin making, provided that such industrial sector exists and that many terms have been coined by “napkin makers” to denote kinds of napkins. The reason that makes a lexical unit of a given language be a *Gener* is not to be found in the extralinguistic realm of concepts but in the structure of the lexicon of that language itself. Of course, it goes without saying that general ontological classes must exist and that the existence of *Gener* relations in lexicons is not fully autonomous from human ontologization of the World. It is nonetheless essential to make it clear that *Gener* is only remotely connected to ontological and/or taxonomic approaches to the structure of natural language lexicons. Psycholinguistics, in particular, has a soft spot for classificatory approaches to the lexicon and the identification of taxa as structuring principle for modeling lexicons. The classificatory anthropological notion of “unique beginner” (Berlin *et al.*, 1975), for instance, played an essential part in the initial organization of the English nominal vocabulary in WordNet (Miller *et al.*, 1990).¹² Within WordNet’s taxonomy of nouns, Miller (1990, p. 251) defines a unique beginner as being “[...] a primitive semantic component of all words in its hierarchically structured semantic field” and postulates 25 such unique beginners expressed as *synsets* = sets of (quasi-)synonyms (see Miller, 1990, Table 1, p. 252): {*act*, *action*, *activity*}, {*animal*, *fauna*}, {*artifact*}, {*attribute*, *property*}, {*body*, *corpus*}, {*cognition*, *knowledge*}, {*communication*}, {*event*, *happening*}, {*feeling*, *emotion*}, {*food*}, etc. This establishes the roadmap for organizing English nouns in WordNet.

Taxon, *unique beginner*, *generic term* are mainstream notions in current approaches to the formal structuring of lexical models. But this goes against the non-taxonomic Lexical System approach (see, Section 1.1) or, even, against traditional, dictionary-based Explanatory Combinatorial Lexicology. For this reason and because of its atypical properties as lexical function, *Gener* is a black sheep

¹² Recall that WordNet is in fact the assemblage of four substructures – for Nouns, Verbs, Adjectives and Adverbs – based on psycholinguistic assumptions on the structure of natural language lexicons: “[...] nouns are organized in lexical memory as topical hierarchies, adjectives are organized as N-dimensional hyperspaces, and verbs are organized by a variety of entailment relations” (Miller *et al.*, 1990, p. 237). It is not the proper place to discuss such assumptions; for an alternate view of the “geometry” of the lexicon within the Lexical System approach, see Polguère (2014).

that has to be kept at arm's length for fear of letting the taxonomic perspective sneak into our lexical models through the back door. On the other hand, generic groupings allow for the modeling of regularities in lexical networks, as will soon be demonstrated, and the extensive weaving of *Gener* relations in Lexical Systems has proven to be an important task that has unfortunately been overlooked in our methodology until recently.

In order to have a true lexicographic approach to the weaving of *Gener* relations, we should distance ourselves from psychological or anthropological prejudices about the taxonomic structure of natural language lexicons and characterize this lexical function in “purely” linguistic terms. For this, we need to find within natural languages themselves criteria for the identification of *Gener* relations. Following Mel'čuk and Polguère (2021, Section 2.1, pp. 101-102), we propose below a two-fold characterization of *Gener* based on (i) general semantic properties and (ii) a combinatorial criterion for identifying a given $\text{Gener}(L_1) = L_2$ relation.

GENERAL SEMANTIC PROPERTIES. $\text{Gener}(L)$ is a lexical unit that denotes a class of “things” all possible referents of L belong to and, therefore, that can be used as a generic term for L .¹³ By extension, $\text{Gener}(L)$ designates a semantic lexical class L belongs to. For a set of lexical units to qualify as a **class**, it has to possess a minimal cardinality: two elements is insufficient, fifty is largely enough. The proposed definition needs to be paired with combinatorial criteria that can be used to test if, indeed, a candidate generic term qualifies as true *Gener* – i.e., is considered by the language itself as being classifying.

COMBINATORIAL CRITERION. There exist specific constructions that function as some sort of linguistic reactants allowing for the identification of a *Gener* relation. The compatibility of the candidate generic term for L with these constructions has to be tested in order to validate its $\text{Gener}(L)$ status. The most significant construction is (in its version for the English language):¹⁴

- for nouns, “ L , L' , L'' , ... *and other* $\text{Gener}(L)$ ” – e.g., *daffodils, roses, hyacinths and other flowers* $\Rightarrow \text{FLOWER}_{(N)}$ is a *Gener* of DAFFODIL.
- for verbs, adjectives and adverbs, “ L , L' , L'' , ... *and* $\text{Gener}(L)$ *in other ways*” – e.g., *to crawl, to walk, to run and to move in other ways* $\Rightarrow \text{MOVE}_{(V)}$ is a *Gener* of CRAWL_(V).

No extralinguistic consideration comes into play here. We simply follow what the English language tells us: the above combinatorial property of a $\text{Gener}(L)$ demonstrates its ability to function as some sort of “semantic pronoun” for L , as illustrated below.

¹³ A more convoluted though maybe more precise way to phrase this is to say that $\text{Gener}(L)$ is a lexical unit whose denotation identifies a class of denotations to which the denotation of L belongs.

¹⁴ See Mel'čuk and Polguère (2021, Section 2.1, pp. 101-102) for a more detailed presentation.

- (1) a. **Daffodils** are dangerous if eaten: **they** can poison badly.
 b. **Daffodils** are dangerous if eaten: **these flowers** can poison badly.

Let us conclude with an important remark about *Gener* and *lexicographic definitions* in preparation for the discussion in Section 3. We consider here *analytical definitions* only – see, for instance, Mel'čuk and Polguère (2018). An analytical definition for a lexical unit L is an equivalence between two elements:

1. the *definiendum* (= what is defined) that is a linguistic expression featuring L together with the actant slots it controls – e.g., *X steals Y from Z*, *admiration of X for Y*, *finger of X*, etc.;
2. the *definiens* (= what is defining), or definition proper, that is a structured paraphrase of the definiendum made up of meanings that are simpler than (\equiv included in) the meaning of L.

It is essential to stipulate that the definiens is structured: it has to abide by some logical principles that organize its content. In particular, at the highest level of organization, it is structured into:

- a *central component* (sometimes called *generic component* or *closest kind*) that corresponds to the core of L's meaning;
- a series of *peripheral components* that present specific semantic content in L's meaning as regards to (i) the central component and/or (ii) the meaning of other lexical units whose definitions possess the same or a similar central component as L's definition.

Clearly, *Gener*(L) is a good candidate as central component of L's definition, but this is by no means systematic, for at least two reasons. Firstly, L may not possess any *Gener*. For instance, there is no *Gener*(SKY_(N)) and the central component of the definition of SKY_(N) needs to be formulated as a nominal phrase expressing a configuration of semantemes and not by a single lexicalized meaning. Secondly, L can very well possess more than one *Gener*(L) that are not even quasi-synonyms; for instance:¹⁵

$$\text{Gener}(\textit{flute}_{(N)}\mathbf{3}) = \textit{glass}_{(N)}\mathbf{2a}; \textit{tableware}$$

The semicolon in bold that separates *glass*_(N)**2a** and *tableware* in the above formula indicates a significant semantic distance between these two lexemes.¹⁶ Their validity as *Gener*(*flute*_(N)**3**) is illustrated below:

¹⁵ The lexicographic numbering in what follows is borrowed from the online edition of the *American Heritage Dictionary* – URL: <https://www.ahdictionary.com>.

¹⁶ Coincidentally, *tableware* is a value of *Gener* for *glass*_(N)**2a**.

Lexique, 31 (2022), 27-54.

e-ISSN 2804-7397.

- (2) a. **flutes**, coupes, highballs and other **glasses**
- b. **flutes**, carafes, plates and other **tableware**

However, only the lexeme $\text{GLASS}_{(N)}2a$ qualifies as central component of the definition of $\text{FLUTE}_{(N)}3$ – of which it is a hyperonym, by the way – because only $\text{GLASS}_{(N)}2a$ denotes a class of actual objects: to use Wierzbicka (1984)’s formulation, a flute *is a kind of* glass. The lexeme TABLEWARE , on the other hand, denotes a very heterogenous class of objects that are brought together with respect to shared characteristics of function/usage; it would be extremely odd to say that a flute *is a kind of* tableware – for this reason, this second $\text{Gener}(\text{flute}_{(N)}3)$ cannot be at the same time a hyperonym. There are other cases where lexical units can have multiple Gener (e.g., in cases of semantic ambivalence analyzed in Milićević & Polguère, 2010) but exploring further this question would lead us astray while our reader is probably getting more and more impatient to hear about regular polysemy. We are almost there.

2.2. Role of Gener in the topology of the lexicon as lexical network

Lexical units whose corresponding node in a Lexical System is the target of Gener relations belong to the topological category of *crossroad lexical units* (Polguère, 2014, p. 18): they are significant meeting points in the lexical network that can be loci for the identification of lexical regularities. From a graph-theory perspective, the existence of such nodes is predicted by the small-world nature of Lexical Systems (see, Section 1.1): while the average degree (i.e., number of connecting arcs) of nodes in small-world networks is low, a minority of nodes set themselves apart by their (very) high degree. Returning to the social network metaphors, they are the nodes of “famous” lexical units, those with many “friends” in the lexicon. Gener lexical units participate actively in the small-world structure of natural languages lexicons (at least, as we model them) because, through their semantic content, they establish classes of lexical units.

We shall call Gener *clusters* in Lexical Systems those groupings of lexical units that are united by one (or more) identical Gener . The display of these Lexical System clusters with the *Spiderlex* graphical browser,¹⁷ at least for prolific generic terms, results in characteristic “dandelion” structures, as illustrated in Figure 2 with the generic French noun FLEUR 1.2 ‘flower’ of the French Lexical Network (fr-LN).

¹⁷ See the Spiderlex webpage: <https://lexical-systems.atilf.fr/en/spiderlex-en/>.

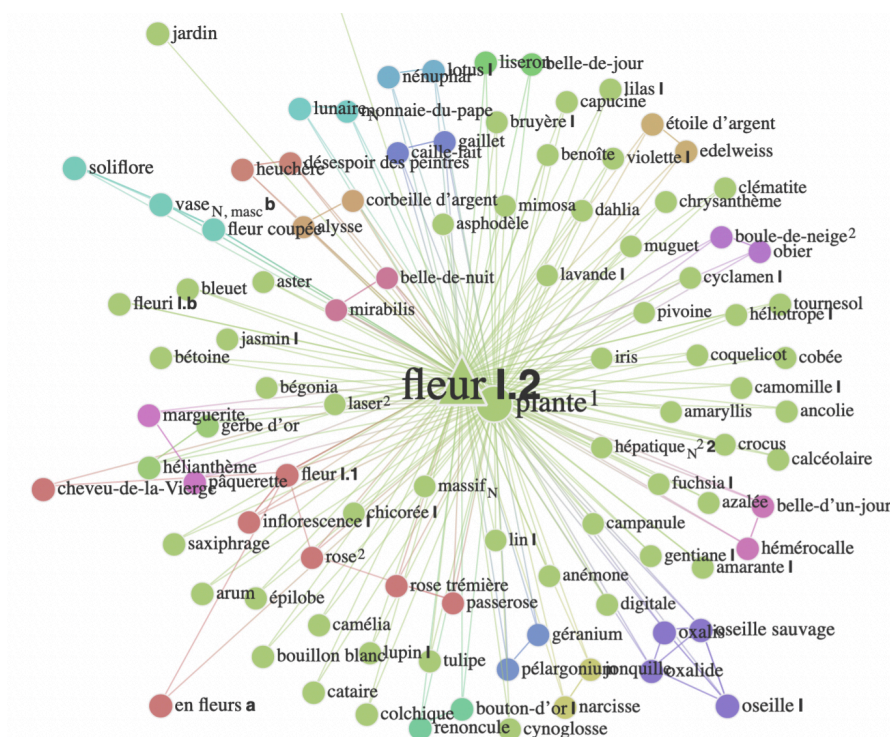


Figure 2. Prototypical dandelion-shaped Gener cluster around FLEUR I.2 ‘flower’ in the fr-LN.

The polysemy of the French vocable FLEUR and the numbering of its senses are discussed in Section 3.1 below. Note that Figure 2 shows more than just the Gener cluster controlled by FLEUR I.2: other lexical units than flower names appear – e.g., PLANTE¹ ‘plant_(N)’, JARDIN ‘garden_(N)’, BOSQUET ‘copse’, etc. This is due to the design of the underlying algorithm used by Spiderlex to conduct the topological exploration of the lexical graph – see, B. Gaume’s principle of *proxemy* (Gaume, 2004; Gaume *et al.*, 2008) – and to the actual content of the fr-LN at the time the figure was generated.¹⁸

¹⁸ Here is a technical explanation of why Figure 2 contains a semantically heterogeneous set of lexical units (not only flower names). A browsing query to Spiderlex concerns lexical entities (graph nodes), and not specific lexical relations (arcs). Such query returns the *proxemy graph* of the queried lexical entity. This graph is entirely determined by the topological structure of the French or English Lexical Network through so-called *random walks* originating for the queried item. Spiderlex’s behavior can be configured in order to take into consideration the semantic weight of the arcs: e.g., a synonymy relation (Syn) between two nodes implies a strong semantic relation between the corresponding lexical units while an intensification (Magn) relation does not have the same implication (no significant semantic intersection between the base and the collocate in the Magn collocation *a fat check*). This add-on to basic proxemy computation (the one described in B. Gaume’s papers mentioned above) allows for the extraction of either *lexical spaces* (semantic weight is ignored) or *semantic spaces* (only arcs with semantic weight are followed during random walks), thus introducing some semantics in the way queries are executed. However, a semantic space, such as in Figure 2, remains semantically heterogeneous in terms of nodes and arcs. Additional navigation tools are available in Spiderlex in order to selectively display lexical nodes clusters or lexical relations extracted from lexical/semantic spaces (Ollinger & Polguère, 2020).

Crossroad lexical units, as mentioned earlier in the present section, are access points to lexical regularities and can be used to identify general lexical regularities that structure each natural language lexicon. Another family of crossroad lexical units can help illustrate this point: “joker” collocates for given semantic classes, such as Fr. ÉPROUVER ‘feel_(v)’ as support verb of the $Oper_1$ type for nouns of feelings – *éprouver de l’amour, du désir, de l’antipathie* ... ‘to feel love, desire, antipathy ...’. A lexeme such as ÉPROUVER is the target of a large number of $Oper_1$ lexical function links in the French lexicon and as such it can be used to:

- classify lexical units and establish correlations between various lexical function connections – e.g., there is a logical correlation in French between the class of lexical units L for which $Oper_1(L) = éprouver$ [ART ~] and those for which $Gener(L) = sentiment$ ‘feeling’;
- systematically explore the lexical and grammatical combinatorics of collocational lexical units, dealing with collocations the other way round, from the lexicographic description of the collocate to that of the base.

To our knowledge, Mel’čuk and Wanner (1994) is the first exploration of regularities among collocations performed within Explanatory Combinatorial Lexicology: its target was German names of feelings. This work later inspired lexicological and lexicographic research on Spanish (Barrios Rodríguez, 2008, 2009; Barrios Rodríguez & Boguslavsky, 2019). For Lexical Systems, punctual studies have been conducted on French (Delaite, 2012) and Russian (Mikhel, 2018).

This shows the important role crossroad lexical units can play in the structure of natural language and in lexicography as well. We can now proceed with an account of an experiment we conducted on the use of $Gener$ relations for the exploration of regular polysemy.

3. Using $Gener$ clusters to study and model regular polysemy

In order to experiment with the extraction of $Gener$ -induced lexical generalizations, we applied the technique of upstream weaving of $Gener$ relations – from a given generic term to lexical units it is a $Gener$ of – using the French generic term *fleur* ‘flower’. There were two reasons for doing so. Firstly, we wanted to go off the beaten tracks and work on other semantic domains than, for instance, the overused feelings/emotions. Secondly, we knew that, with flower names, we were dealing with (i) a sizable $Gener$ cluster, (ii) in a field where regular polysemy was definitely present and (iii) that was barely developed in the fr-LN at the time the experiment started.

This section is organized as follows: first, we give essential information on the actual generic term we worked on that is a specific sense of the French vocable FLEUR (Section 3.1); second, we propose a definition template for French nouns that are names of flowers and we detail the method

we followed to weave the *Gener* cluster and build the definition template (Section 3.2); finally, we present the patterns of regular polysemy we extracted (Section 3.3).

3.1 Picking the right *fleur* ‘flower’

Any lexicographic work performed on a “word” must first consider a **vocable** and its potential polysemy, that is, its different senses. In line with this cardinal principle, our initial task is to clearly identify the generic term we are working on and, for this, identify the polysemy of the vocable it belongs to.

In the version of the fr-LN we used, three lexemes (i.e., word senses) make up the vocable FLEUR. They are listed below with approximate lexicographic definitions expressed in English: the definienda – see, Section 2.1, above – are highlighted in gray and bold parentheses indicate weak/optional components. Each definition is accompanied by illustrative examples; copolysemy relations are also indicated between square brackets.

- I.1 *fleur I.1 of X* ‘part of a plant X, (that is colorful,) that participates in its reproduction process’ – Ex.: *On admire les magnifiques **fleurs** blanches des géraniums.* ‘We admire the beautiful white **flowers** of the geraniums’
- I.2 [**Metonymy** of I.1] *fleur I.2 in X (used by Y)* ‘(part of a) plant growing in X that comprises its flower(s) I.1 (used by Y for decorative or similar purposes)’ – Ex.: *Il a disposé les **fleurs** dans un vase.* ‘He arranged the flowers in a vase’
- II [**Metaphor** of I.1] *fleur II of X* ‘part of X that is at the very surface of it, like flowers I.1 at the extremity of a plant’ – Ex.: *J’admirais sa délicate fleur de peau.* ‘I was admiring the delicate **surface** of her/his skin’

The last lexeme, FLEUR II, does not have a corresponding sense in the English vocable FLOWER_(N) and does not translate easily; *very surface* is only an approximate translation. The presence of this sense in the structure of the vocable may seem odd even to native speakers of French as it is mainly used not to express its individual meaning but as lexical component of some idioms, such as in ‘FLEUR DE SEL’ ‘high-grade sea salt’, ‘FINE FLEUR [de N]’ ‘elite [of a group of people N]’, etc. The only relevance of FLEUR II to the present discussion is that it highlights the presence of a semantic component in the meaning of FLEUR I.1 that has to do with the position of flowers, on branches or stems. When writing a complete definition for FLEUR I.1, the lexicographer shall include a specific ‘position’ component to which the metaphorical sense can be hooked. Otherwise, the last sense does not concern us here and we have to concentrate on the first two. Among them, only the second, FLEUR I.2, is the *Gener* we are going to exploit. Botanical terminology may include a class of terms for specific kinds of FLEUR I.1, but it is surely not what we talk about when we say in non-specialized language:

- (3) des roses, des tulipes, des œillets et autres fleurs
 ‘roses, tulips, carnations and other flowers’

Clearly, the generic term used in (3) is FLEUR I.2; it is the node in the fr-LN that will be at the center of the *Gener* cluster we target (visualized earlier in Figure 2, Section 2.2).

Now that we have zeroed in on the precise sense of the vocable FLEUR around which the *Gener* cluster will be built, we need to deal with another issue: the additional availability of PLANTE¹⁹ ‘plant’¹⁹ as *Gener* for names of flowers. Some names of flowers in French (and in English) are lexicalized as names of plants as well. For instance, in French, *tulipe* ‘tulip’ can rightfully be used to designate both a flower I.2 and a plant; *rose*, on the other hand, designates only a flower I.2, and a part of something bigger: a plant called *rosier* ‘rose tree’. This means that for each flower name we process while weaving the *Gener* cluster of FLEUR I.2, we need to examine if one or more values of *Gener* have to be woven.²⁰

3.2 Definition template for flower names

Any regular polysemy is hooked to the definition of the basic lexical unit of vocables it applies to. For instance, the well-known regular metaphorical copolysemy relation L_1 ‘animal’ \mapsto L_2 ‘person having a characteristic analogous to a characteristic of L_1 ’ (see SNAIL I \mapsto SNAIL II ‘slow person’) is by essence connected to/triggered by a component of L_1 ’s definition. The identification of regular polysemy in the paradigm of flower names is therefore interwound with that of building a template for the definition of flower names. Types of regular polysemy we identify among flower names have to be connected to “regular” components found in the corresponding definition template. Though we start, in the present section, with the presentation of the definition template for flower names, the two tasks of building it and identifying types of regular polysemy (see Section 3.3 below) are interdependent.

The definition template is presented here in three steps: (i) definiendum, (ii) central component of the definiens and (iii) peripheral components.

DEFINIENDUM. It is not unfounded to base the definiendum of the definition of flower names on that for the generic term FLOWER I.2, whose drafted definition was given in Section 3.1. The definiendum of this definition is phrased as follows: *fleur* I.2 in X (used by Y). It features a first actant slot X

¹⁹ The superscript in the name of the lexeme PLANTE¹ originates from the presence of a homonym in French: PLANTE² ‘sole [of someone’s foot]’.

²⁰ Another *Gener* we have encountered is ARBUSTE ‘bush’, for nouns such as MIMOSA ‘touch-me-not’ for which we did not find any evidence that two separate senses should be distinguished: one for the bush and one for the flower.

corresponding to the place where flowers grow. There are many indications in French of the existence of this actant slot; for instance, the following adjectival lexicalizations of X (we use fr-LN's lexicographic numbers):

$$(4) \quad A_1(\text{fleur I.2}) = \text{fleuri I.b, en}_{(\text{Prép})} \text{III.1 } [\sim(s)]; \text{couvert}_{(\text{Adj})} \text{I.1 } [\text{de } \sim(s)], \text{recouvert 2 } [\text{de } \sim(s)]^{21}$$

The second actant slot Y, for the “user” of the flower, is indicated as optional because a vegetal can very well be designated as being *une fleur* even if it is not associated with any typical utilization. On the other hand, there is a proliferation of linguistic evidence pointing to the fact that FLEUR I.2 is conceptualized as denoting something that can be used (to decorate, to serve as a gift, etc.). Note that the fact that FLEUR I.2 controls two actant slots does not imply that lexical units for flower names all control the same actantial structure, but it is an indication that they have this potential.

CENTRAL COMPONENT OF THE DEFINIENS. The central component of the definiens for flower names is of course ‘fleur I.2’. This means that, recursively, the central component of the definiens for FLEUR I.2 given above (Section 3.1) – ‘(part of a) plant’ – is **by default** inherited by definitions of flower names. It indicates that what is called *fleur I.2* in French is a plant: either a rooted stem with flower(s) I.1 (e.g., *tulipe* ‘tulip’) or a more complex vegetal structure (e.g., *géranium* ‘geranium’).²² Additionally, the weak component ‘part of’ indicates that so-called ‘cut flowers’ (Fr. ‘*fleurs coupées*’) can be designated by flower names as well (tulips in a garden vs. tulips in a vase).

PERIPHERAL COMPONENTS OF THE DEFINIENS. We have been able to identify, through the lexicographic process summarized at the end of the present section, 8 recurrent peripheral components listed below. They are the basis for the establishment of regular polysemy patterns among French flower names presented in Section 3.3.²³

A) Color of the flower(s) I.1 (obligatory?). This is the first component in our list that provides a characterization of the flower(s) I.1 part of the flower I.2. It is an important component (see, Section 3.3) that may have to be flagged as obligatory. Even for types of flowers I.2 that can have

²¹ Translation of the values: *flowery*, *in bloom* (lit. *in flower(s)*); *covered* [with ~s], *recovered* [with ~s], as in *a flowery garden*, *a garden in bloom*, *a garden covered with flowers*, etc.

²² Some dictionary definitions try to reflect this distinction. For instance, the online *Petit Robert* defines *tulipe* with the double-shot formula *Plante à bulbe dont la fleur renflée à la base est évasée à l'extrémité ; cette fleur* ‘Bulbous plant whose flower, with a bulging base, is flared at its extremity; **this flower**’ while *géranium* (in non-specialized language) is simply defined with *Plante à feuilles arrondies et velues, à fleurs en ombelles roses, blanches ou rouges* ‘Plant with roundish and fuzzy leaves, with flowers grouped in pink, white or red umbels’ – <https://dictionnaire.lerobert.com>.

²³ In order to formulate our comments on the peripheral semantic components with precision, we use the English term *flower* accompanied by a lexicographic number – either I.1 or I.2 – borrowed from the fr-LN.

flowers I.1 of many different colors, this information about diversity should be included in the corresponding lexicographic definitions: for instance, with a component ‘of various colors’.

B) Scent of the flower(s) I.1. This is the second most important component concerning the specification of the flower(s) I.1 part of the flower I.2. We do not think it should be obligatory, though: in the absence of (a specific) scent, we propose to simply omit Component B.

Remark on polarity in Components A and B. Speaking in the “language of lexical functions” (see 1.2 above), color and scent definiens components for flower names are **by default** oriented towards *Bon* ‘good’ rather than *AntiBon* ‘bad’. A flower name with a lexicographic definition containing *AntiBon* of color or scent²⁴ (we did not encounter any) would definitely qualify as denoting an *AntiVer* flower I.2: a flower I.2 that is not as it is expected to be.

C) Size of the flower(s) I.1. The lexicographic definition of the lexeme *FLEUR* I.1 has to specify that it denotes **small** entities (relative to the size of the human body) that are part of a plant. However, specific types of flowers I.2 are often characterized by the fact that their flower(s) I.1 are relatively small or, on the contrary, relatively big.

D) Form of the flower(s) I.1. See the presentation of polysemy Pattern 4 in Section 3.3 below.

E) Disposition of the flower(s) I.1. We did not find any evidence of the relevance of this component in the polysemy of the vocables we examined. It is obvious, however, that many types of flowers I.2 are characterized by the disposition of their flower(s) I.1 on their stem. We include in this component the (approximate) indication of the number of flower(s) I.1 for each flower I.2: a single one, several, clusters, etc.

F) Environment in which the flower I.2 grows. It can be the immediate physical environment – for instance, it is essential to mention in the definition of *NÉNUPHAR* ‘water lily’ that it denotes a flower I.2 found on waterbodies – and/or it can be a geographical area – for instance, *ÉDELWEISS* ‘edelweiss’ that denotes a mountain flower I.2.

G) Seasonality of the flower I.2. This is important, for example, for flowers I.2 that are associated with springtime and whose corresponding lexical unit can carry associated connotations – see *le temps du muguet* ‘the time of lilies of the valley’ as a metaphor for springtime in a famous French song.²⁵

H) Utilization of the flower I.2. See the presentation of polysemy Pattern 3 in Section 3.3 below.

²⁴ i.e., ugly color or bad scent.

²⁵ Francis Lemarque, *Le temps du muguet*.

For lack of space, we cannot go deeper into the topic of patterns of lexicographic definitions for flower names and, more generally, lexical units sharing identical generic terms. The reader interested in a Natural Semantic Metalanguage perspective on this question should consult Wierzbicka (1985, ch. 4 & 5), in particular the notion of “mental questionnaire” proposed by the author (Section 5.3, pp. 332-336). For “practical” (i.e., non-theoretical) lexicography, one should examine the definition component of *template entries* presented in Atkins and Rundell (2008, Section 4.5).²⁶

At this stage, we can summarize the three-step method we applied in order to build the targeted *Gener* cluster, displayed earlier in Figure 2 (Section 2.2).

1. Gathering of an extensive (though not exhaustive) list of close to 150 flower names – we started from the list of hyponyms of *fleur* in the dictionary of synonyms embedded in the *Antidote* linguistic software (Druide, 2021).
2. For each element of this list, creation of the corresponding vocable, with its polysemy, if it is not already present in the fr-LN – often, when the vocable preexisted in the fr-LN’s wordlist, its polysemy had to be completed.
3. For each lexicographic article of a lexical unit that is created or updated, introduction of minimal lexicographic information: grammatical characteristics (part of speech, gender ...), inflections (singular and plural forms), semantic label, propositional form (actant structure), *Gener* and other essential lexical function relations, lexicographic examples.

Note that, following our relational approach to polysemy (see Section 1.3 above), each copolysemy relation that participates in the structure of the vocables we worked on is explicitly encoded as a typed copolysemy link in the lexical graph.

In the process of executing the above tasks, we extracted interesting lexical generalizations by the sheer fact of being confronted to a mass of linguistic data forming the lexical galaxy of FLEUR I.2. Generalizations concerning lexicographic definitions have just been discussed; we are now examining patterns of regular polysemy.

²⁶ In Atkins and Rundell (2008), each template entry belongs to a “template category” such as: Academic Qualification, Animal, Bodypart, etc. (on p. 128). These template categories are akin to *Gener* lexical units our definition templates are associated to – here, FLEUR I.2.

3.3. Extraction of patterns of regular polysemy

Le parfum lourd des sauges rouges
Les dahlias fauves dans l'allée
Le puits, tout, j'ai tout retrouvé
Hélas
*Barbara, Mon enfance*²⁷

We postulate the existence of a pattern of regular polysemy among flower names if at least two distinct copolysemy relations of the same type are present in the lexical network. Based on this minimalist interpretation of the notion of “pattern,” we identified four polysemy patterns that are explicated below. Each pattern description starts with an illustration, followed by comments. Remember that lexical units whose *Gener* is FLEUR I.2 frequently possess PLANTE¹ ‘plant’ as *Gener* as well; their derived copolyseme can therefore be anchored in either of their semantic facets. We decided to order the patterns according to their degree of relevance to the *Gener* cluster of FLEUR I.2, not PLANTE¹. Based on copolysemy relations we found and on common qualifiers attached to names of flowers in texts, it seems clear that flower names (at least, in French) are about first color, then odor, etc. Our polysemy patterns are ordered accordingly.

PATTERN 1. Metonymy-based metaphor: ‘L’ ↔ ‘of a certain color as if it were the color of L’

- (5) CYCLAMEN_(N) I ‘cyclamen’ ↔ CYCLAMEN_(Adj) II ‘that is mauve, similarly to the color of C. I’

The high productivity of this pattern highlights the fact that the ‘color’ peripheral component in the definition of flower names – see Component A (Section 3.2 above) – is probably the most significant one. Some comments are needed in order to explain (i) the name/conceptualization of Pattern 1 and (ii) the peculiar inter part of speech copolysemy relation it implements (N→Adj).

The notion of *metonymy-based metaphor* and its role in the modeling of copolysemy relations are discussed in Polguère (2018, Section 3.4.1). There are two flavors in copolysemy relations such as the one illustrated in (5) above. They contain some form of **metonymy** because a color is named after the name of a flower prototypically associated with this color: this mauve is one of the prototypical colors of cyclamens. However, the metonymy is approximate and has to be formulated as an **analogy** – ‘of a certain color as if it were the color of cyclamens’ – because, firstly, cyclamens can be of other colors (white, etc.) and, secondly, it is not necessarily the case that the mauve the

²⁷ *The heavy scent of red sages / The tawny dahlias in the alley / The well, everything, I got back everything / Alas* (Barbara, *My childhood*).

Speaker is referring to is precisely the mauve of cyclamens. This is why Pattern 1 is a case of metonymy-based metaphor and not a “pure” metonymy, or metaphor.²⁸

There are grammatical issues with Pattern 1. The actual lexical creation process in French, in such cases, is first the derivation of a **nominal expression** denoting a color with the name of the flower (or fruit, etc.) as appositive noun: *la couleur cyclamen* ‘the cyclamen color’. It is from this structure that an **adjective** is produced: *un rideau cyclamen* ‘a cyclamen curtain’. The appositive nominal nature of the derivational source manifests itself through remarkable grammatical characteristics of the derived adjective: it is invariable in all its inflections (see the unchanged feminine plural form in *des robes cyclamen* ‘cyclamen dresses’) and it is mainly postposited (which is the position for appositive nouns). In spite of this particular behavior, the derived copolyseme is indeed an adjective; for instance, it can be modified by an adverb: *une robe intensément cyclamen* ‘an intensely cyclamen dress’.²⁹ Of course, as for all adjectives, it is also possible to derive on-the-fly a corresponding noun – e.g., *J’aime le cyclamen*_[couleur] ‘I like the cyclamen_[color]’ – but this is perceived as a lexical creation on the part of the Speaker. To make the situation even more complex, such nominal derivatives produced by ricochet from the metonymy-based pattern can end up being lexicalized, probably because the type of color they denote is very popular (for clothing, etc.). Such is the case of the color noun AMARANTE_(N) II.b ‘purple red color’:

- (6) Alice et Bob vont maintenant choisir chacun une autre couleur secrète. Après maintes hésitations, Alice s’arrête sur de l’**amarante**.
 ‘Alice and Bob are now going to choose each of them a secret color. After much hesitation, Alice decides on the **amaranth color**’.³⁰

This color name is undoubtedly a lexicalized noun in French: an S₀ derivative (in terms of lexical function) of AMARANTE_(Adj) II.a; this latter is itself created from the flower name AMARANTE_(N) I,

²⁸ As one can see, we take *metonymy* in a narrow sense such that if X is a metonymy of Y, the state of affairs corresponding to X **necessarily** includes the state of affairs corresponding to Y. For instance, when one says that the 3rd violin of the orchestra is bald, one is at the same time necessarily talking about an actual musical instrument: the musical instrument usually played by the bald musician. This situation is totally different from the copolysemy relation in Pattern 1: if one talks about a beautiful fuchsia dress, this corresponds to a state of affairs where a given color is present, but in the absence of actual flowers with their actual color. Analogy (metaphor) is prominent here, shadowing contiguity (metonymy).

²⁹ The agreement of French color adjectives is a known issue, systematically mentioned in grammars, language courses, etc. Interestingly, it is almost never considered from the perspective of polysemy. On the topic of invariable adjectives in French, in the context of a reflection on the part of speech of nominal and prepositional idioms, see Mel’čuk (2006, Section 6).

³⁰ [https://www.arsouyes.org/blog/2017/09/Diffie Hellman couleurs](https://www.arsouyes.org/blog/2017/09/Diffie%20Hellman%20couleurs).

following Pattern 1.³¹ Finally, a last grammatical goody: color nouns in French are always masculine while flower nouns can be either masculine or feminine. In the sequence of copolysemy relations found in the vocable AMARANTE, the flower noun numbered I (basic lexical unit of the vocable) is feminine, and the resulting color noun, numbered II.b, is masculine. This can get some dictionaries quite confused as they are more reluctant than us to put under the same vocable entry feminine and masculine nouns, and an adjective as well. Such is the case of the French dictionary of *Antidote* (Druide, 2021) that contains three separate vocable entries: for the feminine flower noun, the color adjective and the masculine color noun. The online *Petit Robert* (<https://dictionnaire.lerobert.com>) puts the flower noun and the adjective under the same entry and ignores the lexicalization of the color noun. Just like *Antidote*, the online *Larousse* (<https://www.larousse.fr/dictionnaires/francais-monolingue>) has three separate vocables (no sense numbering) and offers a grammatical warning under a *DIFFICULTÉS* ‘difficulties’ section translated below:

The noun *amarante* is feminine when it designates the plant with purple flowers and masculine when it designates the color: *une amarante en pot* ‘a_[fem.] pot amaranth’; *un amarante profond* ‘a_[masc.] deep_[masc.] amaranth’.

The adjective is invariable: *des draperies amarante* ‘amaranth_[inv. fem. pl.] draperies’.

We gave a rather detailed account of Pattern 1 because it has a strategic importance in the *Gener* cluster of FLEUR I.2 and it possesses interesting characteristics. Due to lack of space, we shall be much more concise with the three remaining patterns.

PATTERN 2. Metonymy: ‘L’ ↔ ‘fragrance product that has the scent of L’

- (7) LAVANDE_(N) I ‘lavender’ ↔ LAVANDE_(N) II ‘lavender fragrance that has the scent of L. I’

This pattern connects to what is, in our opinion, the second most significant peripheral component in the lexical definition of French flower names: the scent of the flower – see Component B (Section 3.2 above). Of course, not all flowers are characterized by their scent in the French language and, among those that are, only a limited number has triggered Pattern 2.

Note that Pattern 1 has also been triggered in the case of LAVANDE_(N) I, used in illustration (7): the French lexicon contains the metonymy-based metaphor LAVANDE_(N) I ↔ LAVANDE_(Adj) III ‘of a color similar to that of L_(N) I’. Our lexicographic numbering favors in this case Pattern 2 copolysemy relation over Pattern 1. This reflects the fact that the use of *lavande* as a scent name is very

³¹ The vocable FUCHSIA ‘fuchsia’ offers another illustration for the lexicalization of the color name in French: FUCHSIA_(N) I → FUCHSIA_(Adj) II.a → FUCHSIA_(N) II.b [a type of pink].

Lexique, 31 (2022), 27-54.

e-ISSN 2804-7397.

widespread, unlike its use as a color adjective; this latter can even be unknown to native speakers of French. (There is a similar situation with English *lavender*.)

PATTERN 3. Metonymy: ‘L’ \mapsto ‘food/drink made with L’

- (8) GENTIANE I ‘gentian’ \mapsto GENTIANE II ‘alcoholic drink made with the root of G. I’
- (9) a. CAMOMILLE I ‘chamomile’ \mapsto CAMOMILLE II.a ‘food substance made of C. I used for hot drinks’
 b. CAMOMILLE II.a \mapsto CAMOMILLE II.b ‘hot drink prepared with C. II.a’

This metonymic pattern is well known and productive with names of plants in general. It is related to the ‘utilization’ component in the definition of flower names – see Component H (Section 3.2 above). The productivity of Pattern 3 is motivated firstly by extra-linguistic reasons: many plants are used to derive products for human consumption as food, medicine or both. However, we find this pattern to be not as relevant as the first two as regards to the *Gener* cluster of FLEUR I.2. Pattern 3 addresses in most cases the ‘plant’ semantic facet of the noun. In the same vein, we decided to exclude altogether from our list a similar pattern where names of materials are derived; for instance BRUYÈRE I ‘heather’ \mapsto BRUYÈRE II ‘wood material from B. I root’.

The second illustration above, (9a–b), shows that Pattern 3 can generate a double-shot copolysemic lexical creation. Because this dual creation is quite productive – e.g., CHICORÉE I ‘chicory’ \mapsto CHICORÉE II.a ‘food substance made with roasted root of C. I for hot drink’ \mapsto CHICORÉE II.b ‘hot drink prepared with C. II.a’ –, it should be modeled as such instead of being accounted for as the combination of two totally separate patterns (though those exist autonomously as well).

PATTERN 4. Form metaphor: ‘L’ \mapsto ‘something whose form evokes the form of L’

- (10) LOTUS I ‘lotus flower’ \mapsto LOTUS II ‘yoga position that evokes the form of a L. I’

This pattern is marginally activated in the French lexicon, at least in the set of vocables we analyzed. This came as a surprise because it is clear that the ‘form’ peripheral component – see Component D (3.2) – is relevant in the lexical definition of quite a few flower names, such as those that are metaphorical idioms alluding to a shape (and eventually color): e.g., *bouton-d’or* ‘buttercup’ (lit. ‘button made of gold’) and *gueule-de-loup* ‘snapdragon’ (lit. ‘maw of wolf’).

This concludes our presentation of polysemy patterns activated in the *Gener* cluster of FLEUR I.2. Our description is informal as our main target was to experiment with an exploratory methodology for polysemy pattern extraction. A more rigorous approach is required at the descriptive level in order to formalize these patterns – see for instance proposals made by Barque (2008, ch. 4), also within the framework of Explanatory Combinatorial Lexicology.

We are convinced that lexical regularities have to be explored and modeled by considering them as part of a **system**, rather than in isolation. For regular polysemy, in particular, the analyses presented above show that this phenomenon should be studied in parallel with regularities one can find in the lexicographic definitions for semantic classes of lexical units (Section 3.2). Ultimately, all these regularities that are encapsulated in the **language** manifest themselves as regular combinatorial patterns in **speech**; this is illustrated by the epigraph to the present section.

4. Conclusion

The purpose of this study was not the “discovery” of patterns of regular polysemy, and the four patterns presented in Section 3.3 are hardly new. Our aim was to propose and experiment with a novel approach to regular polysemy identification/characterization that displays some degree of systematicity. The key feature of this approach is, firstly, to be lexicography-based and to use non-taxonomic network lexical models, namely Lexical Systems. Secondly, it makes systematic use of the lexical function *Gener* and the relations it weaves within Lexical Systems, avoiding the recourse to taxonomically organized semantic labels (such as in Barque *et al.*, 2018). We strongly believe that the *Mental Lexicon* is not structured as a taxonomy and that lexicological studies that are based on or require a taxonomic structure for the lexicon are flawed in terms of cognitive adequacy (Polguère, 2014, 2016). For small-world models such as the fr-LN, we came to the conclusion that the extensive use of semantic labels to project a hierarchical structure onto the lexical graph is a dead-end, precisely because the lexicon resists semantic hierarchization – see non-nominal parts of speech and phenomena such as semantic ambivalence of lexical units (Milićević & Polguère, 2010; Smirnova & Tolochin, 2018). Therefore, semantic labels should be kept very coarse (to be used by grammar rules for instance) and the lexical function *Gener* should systematically be used instead.³²

Finally, it is important to highlight the fact that, within our *Gener*-based approach, regular polysemy is dealt with not in isolation but as a phenomenon that interacts with other lexical regularities (mainly, patterns of lexicographic definitions). It is our opinion that the outcome of our experiment with the *Gener* cluster of FLEUR I.2 is conclusive and that the proposed methodology should be explored further on the fr-LN and, eventually, applied systematically to all main generic terms of the French language.

³² See also remark in footnote 11. Of course, lexical resources that follow or are compatible with this approach are extremely scarce and, from a practical/implementation viewpoint, it is perfectly justified to exploit existing hierarchically organized lexical resources to extract polysemy patterns. As announced in the very first sentence of Section 1, our perspective is that of lexicographers, i.e., builders rather than users of lexical resources.

Bibliography

- Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- Barque, L. (2008). *Description et formalisation de la polysémie régulière du français* [Description and formalization of regular polysemy in French]. PhD dissertation, Université Paris 7.
- Barque, L., Haas, P., & Huyghe, R. (2018). Polysémie régulière et néologie sémantique : constitution d'une ressource pour l'étude des sens nouveaux [Regular polysemy and semantic neology: construction of a resource for the study of new senses]. *Neologica*, 12, 91-108.
- Barrios Rodríguez, M. A. (2008). *El dominio de las funciones léxicas en el marco de la teoría sentido-texto* [The domain of lexical functions in the framework of the Meaning-Text theory]. PhD dissertation, Universidad Complutense de Madrid.
- Barrios Rodríguez, M. A. (2009). Domain, Domain Features of Lexical Functions, and Generation of Values by analogy according to the MTT approach. In D. Beck, K. Gerdes, J. Milićević (Eds), *Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'09)* (pp. 45-53). Observatoire de linguistique Sens-Texte (OLST).
- Barrios Rodríguez, M. A., & Boguslavsky, I. (2019). A Spanish e-dictionary of collocations. *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)* (pp. 160-167). Association for Computational Linguistics.
- Berlin, B., Breedlove, D., & Raven, P. H. (1973). General Principles of Classification and Nomenclature in Folk Biology. *American Anthropologist*, 75(1), 214-242.
- Delaite, C. (2012). *Combinatoire grammaticale et lexicale des collocatifs en français : les cas de LANCER et LOURD* [Grammatical and lexical combinatorics of collocates in French: the case of LANCER 'to throw' and LOURD 'heavy']. MA dissertation, European Master in Lexicography (EMLex), Université de Lorraine.
- Druide (2021). *Antidote 10*. Druide informatique inc.
- Gader, N., Lux-Pogodalla, V., & Polguère, A. (2012). Hand-Crafting a Lexical Network With a Knowledge-Based Graph Editor. In M. Zock & R. Rapp (Eds), *Proceedings of the Third Workshop on Cognitive Aspects of the Lexicon (CogALex III)* (pp. 109-125). The COLING 2012 Organizing Committee.
- Gaume, B. (2004). Balades aléatoires dans les Petits Mondes Lexicaux [Aleatory strolls in Lexical Small Worlds]. *I3 Information Interaction Intelligence*, 4(2), 39-96.
- Gaume, B., Duvignau, K., Prévot, L., & Desalle, Y. (2008). Toward a cognitive organization for electronic dictionaries, the case for semantic proxemy. *Proceedings of the workshop on Cognitive Aspects of the Lexicon (COGALEX 2008)* (pp. 86--93). Association for Computational Linguistics.

- Gries, S. T. (2015). Polysemy. In E. Dąbrowska & D. S. Divjak (Eds), *Handbook of Cognitive Linguistics* (pp. 472-490). De Gruyter Mouton.
- Goddard, C., & Wierzbicka, A. (2007). Semantic primes and cultural scripts in language learning and intercultural communication. In F. Sharifian & G. B. Palmer (Eds), *Applied Cultural Linguistics: Implications for second language learning and intercultural communication* (pp. 105-124). John Benjamins.
- Mel'čuk, I. (2006). Parties du discours et locutions [Parts of speech and idioms]. *Bulletin de la Société de linguistique de Paris*, 101(1), 29-65.
- Mel'čuk, I. (2007). Lexical Functions. In H. Burger, D. Dobrovolskij, P. Kuhn & N. Norrick (Eds), *Phraseologie/Phraseology: Ein Internationales Handbuch zeitgenössischer Forschung/an International Handbook of Contemporary Research* (pp. 119-131). De Gruyter.
- Mel'čuk, I. (2013). *Semantics: From Meaning to Text. Volume 2*. John Benjamins.
- Mel'čuk, I. (2015). *Semantics: From Meaning to Text. Volume 3*. John Benjamins.
- Mel'čuk, I., & Polguère, A. (2018). Theory and Practice of Lexicographic Definition. *Journal of Cognitive Science*, 19(4), 417-470.
- Mel'čuk, I., & Polguère, A. (2021). Les fonctions lexicales dernier cri [Lexical functions, the latest]. In S. Marengo (Ed.), *La Théorie Sens-Texte. Concepts-clés et applications* (pp. 75-155). L'Harmattan.
- Mel'čuk, I., & Polguère, A. (in preparation). *Lexical Functions: The Naked Truth*.
- Mel'čuk, I., & Wanner, L. (1994). Lexical Co-occurrence and Lexical Inheritance. Emotion Lexemes in German: A Lexicographic Case Study. *Lexikos*, 4, 86-161.
- Mel'čuk, I. et al. (1984, 1988, 1992, 1999). *Dictionnaire explicatif et combinatoire du français contemporain. Recherches lexico-sémantiques. Volumes I-IV* [Explanatory Combinatorial Dictionary of contemporary French. Lexico-semantic research. Volumes I-IV]. Les Presses de l'Université de Montréal.
- Mikhel, P. (2018). *Collocational Lexical Units in Russian*. MA dissertation, European Master in Lexicography (EMLex), Université de Lorraine.
- Milićević, J., & Polguère, A. (2010). Ambivalence sémantique des noms de communication langagière du français [Semantic ambivalence of French nouns of linguistic communication]. In F. Neveu, V. Muni Toke, J. Durand, T. Klingler, L. Mondada & S. Prévost (Eds), *Actes de la section "Lexique et morphologie" du 2^e Congrès Mondial de Linguistique Française (CMLF'10)* (pp. 1029-1050). Institut de Linguistique Française (ILF).
- Miller, G. A. (1990). Nouns in WordNet: A Lexical Inheritance System. *International Journal of Lexicography*, 3(4), 245-264.

- Miller, G. A., Beckwith, R., Fellbaum, C. D., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An On-line Lexical Database. *International Journal of Lexicography*, 3(4), 235-244.
- Ollinger, S., & Polguère, A. (2020). *Guide de navigation Spiderlex (version du 19 mai 2020)* [Guide to Spiderlex Browsing (version 19 May 2020)]. Technical report, ATILF CNRS, Nancy. [<https://lexical-systems.atilf.fr/wp-content/uploads/2020/05/Spiderlex-doc-FR-19v20.pdf>].
- Polguère, A. (2011). Classification sémantique des lexies fondée sur le paraphrasage [Semantic classification of lexical units based on paraphrase]. *Cahiers de lexicologie*, 98, 197-211.
- Polguère, A. (2014). From Writing Dictionaries to Weaving Lexical Networks. *International Journal of Lexicography*, 27(4), 396-418.
- Polguère, A. (2015). Lexical Contextualism: The Abélard Syndrome. In N. Gala, R. Rapp & G. Bel-Enguix (Eds), *Language Production, Cognition, and the Lexicon* (pp. 53-73). Springer.
- Polguère, A. (2016). La question de la géométrie du lexique [The question of the geometry of the lexicon]. *SHS Web of Conferences*, 27, 01002, DOI: 10.1051/shsconf/20162701002.
- Polguère, A. (2018). A Lexicographic Approach to the Study of Copolysemy Relations. *Russian Journal of Linguistics*, 22(4), 788-820.
- Polguère, A. (to appear). Les modificateurs redondants et leur rôle en terminologie : de l'amour physique aux masques de protection [Redundant modifiers and their role in terminology: from physical love to protective masks]. In P. Frassi (Ed.), *Phraséologie et Terminologie*. De Gruyter.
- Smirnova, A., & Tolochin, I. (2018). Terrible Angels: Semantic Ambivalence and Polysemy. *GEMA Online*, 18(3), 153-169, DOI: doi.org/10.17576/gema-2018-1803-09.
- Watts, D. J., & Strogatz, S. H. (1998). Collective Dynamics of 'Small-World' Networks. *Nature*, 393, 440-442.
- Wierzbicka, A. (1984). Apples are not a "kind of fruit": the semantics of human categorization. *American Ethnologist*, 11(2), 313-328.
- Wierzbicka, A. (1985). *Lexicography and Conceptual Analysis*. Karoma.
- Žollkovskij, A., & Mel'čuk, I. (1967). O semantičeskom sinteze [On semantic synthesis]. *Problemy kibernetiki*, 19, 177-238.