How is the meaning of complex lexemes constructed?
A study of neoclassical compounds in *-cratie* / *-crate* and *-logie* / *-logue*

Marine Lasserre & Fabio Montermini

This paper proposes a categorial analysis of two connected pairs of so-called neoclassical compounds in French, namely lexemes in *-cratie* / *-crate* and in *-logie* / *-logue*. Its main goal is to investigate, via observation of newly constructed items, how the semantic information of complex lexemes is constructed and to show that it cannot be reduced to a simple function of the elements it contains. Rather, we claim that, from a semantic point of view, a word formation process does not consist primarily of the combination of two or more discrete units (e.g. a base and an affix) but of the inclusion of a new lexeme into a lexical network to which other words constructed by the same pattern also belong. The polysemy of complex lexemes is taken into account, and we show that it cannot be seen as a chaotic and unpredictable fact. In fact, the semantic variability of lexemes apparently constructed with the same pattern is often systematic and predictable, and it cannot be reduced to a – possibly very abstract – meaning from which all the attested meanings are derived.[*]

1. Introduction

The nature of word formation processes (WFPs) is an important topic of debate for models of morphological derivation. For a long time these processes have been viewed in terms of rules (cf. Aronoff 1976, among others), i.e. mechanisms that apply to an input in order to obtain a specific output. Under this view, rules are selective and deterministic. They are selective because an item (a base lexeme, in the case of morphological derivation) either is a suitable input for a rule, or it is completely excluded from its scope; they are determin-

istic because, when they are applied to a particular item, they can only give an output which is entirely predictable. In other words, a rule is not expected to have more than one acceptable output, nor to be applied to various inputs to different degrees. Both the above-mentioned properties of rules are problematic when they are submitted to the proof of data, and in particular when large amounts of real linguistic production are considered. It has often been shown, for instance, that when a morphological rule selects a particular type of base, this requirement can be violated under certain circumstances, thus suggesting that being a suitable input for a morphological process is a gradient rather than a discrete property (cf. Plag 2004, among others). Similarly, it is not uncommon that complex words apparently constructed by means of the same morphological process do not have homogeneous properties. They can, for instance, display a wide range of different meanings, for which it is sometimes hard to identify a common feature (a particularly insightful example is given by the treatment of the suffixes *-isme* and *-iste* proposed by Roché 2011b). That is why, in recent years, models of morphological derivation have undergone an evolution from rigidly rule-based to pattern- or schema-based. Patterns or schemas are considered to be more flexible than rules, and thus more suitable to account for the diversity observed in real languages. This means, for instance, that whereas rules can be described independently of the actual words that exemplify them, the identification of patterns is tightly linked with the observation of the existing lexicon.

The study we propose here is intended to give some hints towards a model of the construction of the meaning of complex words within a pattern-based approach to derivation (§ 2). The material on which the analysis is conducted is a particular class of complex lexemes in French, the so-called 'neoclassical compounds' (NCs). Although NCs are problematic in many respects, including their identification as a homogeneous class of lexical elements, their diachronic and synchronic status in the language makes them a good observation point of the manner in which the lexicon and morphology interact in determining the form and meaning of newly constructed lexemes. This issue is addressed in § 3. For our analysis, we selected four types of NCs in French, ending respectively in *-crate*, *-cratie*, *-logue* and *-logie*. A complete description of the material on which the analysis is conducted is given in § 4. The mechanisms that are responsible for the construction of the meaning of these NCs and their distribution are discussed in § 5. Finally, § 6 contains some concluding remarks.

2. The construction of the lexicon: a pattern-based approach

It is quite common for rule-based models of morphology to adopt a more or less explicit compositional view of the semantics of linguistic structures. As is well-known, the principle of compositionality in linguistics goes back at least to Frege (1980[1914]), and states that the meaning of a linguistic structure can be computed on the basis of the meaning of the parts that constitute it and of the rules that combine them (for morphology, see in particular Aronoff 2007: 803). As far as word formation is concerned, according to the compositional hypothesis, the meaning of a complex word can be entirely described as the result of a semantic operation performed by the word formation rule on the meaning of the base lexeme. However, it is quite easy to observe that, in the real world, complex words – even those constructed by means of very productive rules – may have a wide range of different meanings. Sometimes these meanings are clearly connected, but in many other cases the connection is harder to see. In order to account for this fact, morphologists have often adopted what Booij (2010: 78-79) calls a "monosemy approach", which consists in considering that, at least ideally, each word formation rule contains a unique semantic representation, and that all the variation observed with actual derivates can be reduced to this – possibly very abstract – meaning. In other words, the meaning of a productively constructed derivate is always predictable, and all the particular meanings observed in real language use should be attributed to extra-morphological factors such as pragmatics, stylistics, metaphorical or metonymic shifts, etc. This position is evident, for instance, in the distinction made by Corbin (1989) between the "attested meaning" and the "predictable meaning" of complex lexemes. The latter "has two fundamental properties: it is a compositional function of the morphological structure of a constructed word, and is not fixed by the attested meaning" (Corbin 1989: 32); as a consequence, "each morphological structure is associated with one and only one semantic interpretation that does not necessarily correspond to its attested meaning" (Corbin 1989: 45) (see also Corbin 1991).

Nevertheless, if this "distortion" – as Corbin calls it – between the predictable meaning of a derivational process and its attested meaning(s) really lay outside grammar, as the above-mentioned hypotheses seem to claim, there would be no predictability in the range of the possible referents of a complex word. In fact, it is rarely the case that the lexemes built by the same rule simply display complete semantic unpredictability. As we will see, the semantic types a

derivate can belong to are often limited in number and predictable from the meaning of the base, at least in a probabilistic way. That is why we prefer to adopt a model in which the various interpretations the outputs of a WFP can receive are clearly distinguished and potentially linked, even if the synchronic connection between them may be more or less transparent (what Booij 2010 calls a "regular polysemy approach").

Among the models which go beyond a purely compositional conception of the semantics of complex words we can include onomasiological approaches (cf. Štekauer 1998, Rainer 2012), and more generally lexical approaches (cf. Roché 2009, 2011a for an overview), which consider that the properties of complex lexemes can be accounted for only by taking into account the global lexical network they belong to, and Construction Morphology (cf. in particular Booij 2010), according to which patterns of word formation are constructional schemas that can be hierarchically ordered, thus accounting for the fact that complex words displaying the same exponent (e.g. the same prefix or the same suffix) can receive various, possibly connected, semantic interpretations. In a purely rule-based model, the construction of the meaning of complex lexemes can be described by simply observing the local relation between two items (a derivate and its base); in pattern-based approaches, like the ones just mentioned, on the other hand, the entire network of lexical relations a complex lexeme belongs to should be taken into account in order to properly describe how its actually attested meaning(s) have been constructed. In this kind of framework, we can give an explicit definition of what we consider to be a word formation process (or pattern). It consists, primarily, of inserting a new complex word into a lexical network, to which other words constructed by the same process also belong.

More generally, we defend a dynamic model of morphological derivation, in which WFPs are not given *per se*, but emerge from the generalisations made by speakers on the basis of global lexical relations. In other words, we consider analogy as being the main principle according to which complex lexemes are formed. In several linguistic approaches (including, e.g., the structuralist or the generative tradition) analogies are often opposed to rules. Under this view, rules are automatic mechanisms which, when applied to an appropriate input, do not tolerate any deviation, whereas analogies are considered as occasional, often unpredictable, epiphenomena. However, in much recent work (cf. Rainer 2003: 197-198; Blevins & Blevins 2009), the rule / analogy distinction has been questioned and reinterpreted as being more quantitative than qualitative, and rules are viewed as

very general analogies. Pattern-based approaches, such as the one we adopt, are more compatible with a non-discrete conception of the rule / analogy distinction, in that they allow considering that new lexemes are primarily formed by taking other words in the lexicon as models. These models can correspond to single words, which, by virtue of their frequency, their salience, etc., exert a strong attraction on the rest of the lexicon, but they can also correspond to entire sets of words, or, in the most extreme case, be so general that no precise model can be identified. The NC construction patterns we choose to analyse permit us to exemplify each of these cases.

3. Neoclassical compounds: a heterogeneous class of objects

NCs exist in all the major European languages, and very often the same compound appears, with slight variation due to phonological adaptation, in different languages with a similar semantic content:

(1)   a. *psychologie*   [psik'loʒi]   (French)
      b. *psychology*   [saɪˈkɒlədʒɪ]   (English)
      c. *psicologia*   [psikoloˈdʒiːa]   (Italian)
      d. *psychologie*   [psyçoloˈgiː]   (German)
      e. *psicología*   [sikoloˈxia]   (Spanish)
      f. *psychologia*   [psɨxologja]   (Polish)

Although NCs are quite easily identified on an intuitive basis, this class is hard to define on theoretical grounds, and it is difficult to determine precisely which elements it contains. Nevertheless, for many linguists who have dealt with the phenomenon the existence in various languages of a clear-cut class of NCs with specific properties seems uncontroversial. More specifically for French, a tradition that goes back to Corbin (2001; 2004) identifies NCs as compounds made up of lexemes which, unlike canonical lexemes, are not syntactically autonomous units. However, as lexemes, they have a syntactic category (only manipulated by morphological rules), a conceptual (lexical) meaning and, of course, a formal representation (for more recent accounts of NCs in French, see Amiot & Dal 2007; Namer & Villoing 2007; Villoing 2012). Among the main properties of NCs listed in the literature, we may cite at least the following (examples are taken from French):

(i) at least one of the constituents of NCs is of Ancient Greek or Latin origin (henceforth we will use the label CF for these constituents, as they are commonly referred to as "combining forms" in the literature in English):

(2) a. *anthropophage*   man+eat   'anthropophagous' (Greek)
     b. *homicide*   man+kill   'homicide'   (Latin)

(ii) in French (and in general in Romance languages), the head-determiner order is inverted with respect to 'native' compounds (which are left-headed, whereas NCs are right-headed):

(3) a. *hom<u>i</u>cide*   man+kill   'homicide'
   b. <u>*homme*</u>-*grenouille*   man+frog   'frogman'

(iii) they include a linking vowel, phonologically specified as /ɔ/ with Greek CFs and as /i/ with Latin CFs.

(4) a. *anthropophage*   man+eat   'anthropophagous' (Greek)
     b. *homicide*   man+kill   'homicide'   (Latin)

(iv) they often belong to formal, learned registers, especially in scientific and technical domains, and have been massively introduced in the lexicon of European languages from the 17th century on, and especially from the 18th century. Figure 1 shows the number of words in *-logy* / *-logist* introduced in English and of words in *-logie* / *-logue* introduced in French from the 12th century on (English data are taken from the OED, and French data from the TLFi and the Grand Robert).

In addition, two main properties are claimed to distinguish CFs from 'genuine' affixes:

(v) the position they occupy in the complex word is not fixed, i.e. they can appear in the first or in the last position, whereas an affix is generally either a prefix or a suffix.[1]
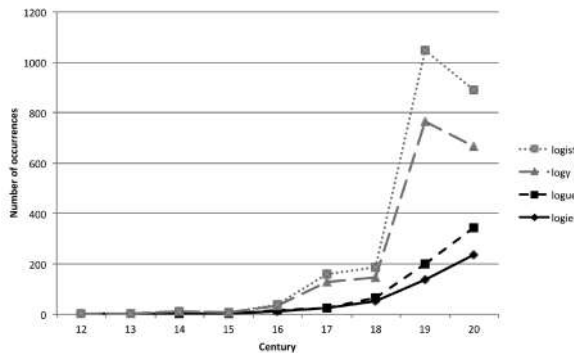
Fig. 1. Number of words in *-logy* / *-logist* introduced in English and of words in *-logie* / *-logue* introduced in French from the 12th to the 20th century.

(5) a. *graphologue*   write+word/study 'graphologist'
     b. *logographe*   word/study+write 'logographer'

(vi) they can be combined with canonical lexemes (6a), but also, very commonly, with other CFs (cf. examples (2-5) above) or with affixes (6b), whereas genuine affixes can never be combined with each other:

(6) a. *cancérologue*   cancer+discourse/study   'oncologist'
        insecticide   insect + kill   'insecticide'

     b. *phobique*   fear + *Suff*   'phobic'
        céphalée   head + *Suff*   'headache'

The features listed above can in fact be applied to a large number of lexemes that are generally classified as NCs in the literature and in dictionaries. There are some factors, however, that complicate this apparently straightforward picture. First of all, the most productive use for CFs today is to be combined with autonomous lexemes and not with CFs or other bound elements. This is not really surprising as lexemes, unlike CFs and affixes, are virtually infinite and constitute the largest reservoir of potential bases for any WFP. Table 1 shows the number of NCs in our corpus (see § 4 below) in which the elements *-logie* and *-logue* are combined, respectively, with an autonomous lexeme or with a bound element (typically another CF).[2]

Table 1. Percentage of autonomous vs. non autonomous elements combined with *-logie* and *-logue* in French.

	-logie		*-logue*	
Autonomous lexeme	624	63.03%	525	66.62%
Bound element	366	36.97%	263	33.38%

Moreover, CFs, as a general set of elements, can hardly be distinguished from other kinds of morphological units. For instance, the semantic criterion according to which CFs, as opposed to affixes, have a lexical and not an instructional meaning (cf. Corbin 2004, Villoing 2012: 33) does not hold in all cases, as in the following examples:

(7) a. *Nous n'avons rien à gagner à emboîter les pas des dictatures* pétroli-fères *wahhabites du Qatar et de l'Arabie Saoudite*
        'We have nothing to gain by in following the Wahhabi petro-dictatorships of Qatar and Saudi Arabia'
        [http://tempsreel.nouvelobs.com/la-revolte-syrienne/20130314.OBS1919/reactions/]

b. *Je fais régulièrement appel à vos connaissances* cinéphiles *pour pallier les miennes qui restent défaillantes dans bien des domaines du cinéma.*
'I regularly exploit your film knowledge in order to remedy mine which are failing in many areas of cinema'
[http://www.effets-speciaux.info/forum/viewtopic.php?p=19207]

In the sentences above, the replacement of the NCs highlighted in bold with a corresponding relational adjective (e.g., respectively, PÉTROLIER or FILMIQUE) would not determine any change in meaning, thus showing that, at least in some uses, the semantic contribution of a CF does not differ, in substance, from that of a canonical suffix.

Apart from affixes, there are at least two other kinds of elements with which CFs can be confused:

(i) the first group is constituted by alternative suppletive stems of autonomous lexemes: in various European languages some lexemes possess suppletive stems of neoclassical origin which are formally closer to the original Greek or Latin word (or in some cases may be reconstructed in synchrony with a supposedly 'classical' form) (for more examples see Roché 2010).

(8)	Lexeme	Stems			
		Native	Greek	Latin	Pseudo-learned
	CHEVAL ('horse')	ʃəval	ip (*hippique* 'equestrian')	ɛk (*équestre* 'equestrian')	
	EAU ('water')	o	idʁ (*hydrique* 'hydric')	akwa(t) (*aquatique* 'aquatic')	
	IVOIRE ('ivory')	ivwaʁ		ebyʁn (*éburnéen* 'eburnean')	ivɔʁ (*ivoréen* 'eburnean')

In the examples above, the native stem is the one which is used when the lexeme is an autonomous syntactic unit; the Greek and Latin stems are used in some learned derivates (e.g. with the affixes *-ique* or *-ien* / *-éen*). In addition, IVOIRE possesses a stem (/ivɔʁ/) for which some cues (e.g. de-diphthongisation) indicate that it has been recreated by speakers on the model of other learned stems from the autonomous one /ivwaʁ/. Amiot & Dal (2007) analyse such elements as *lud-*, *pyr-*, and partially *anthrop-*, as suppletive stems, respectively, of the lexemes JEU ('game'), FEU ('fire') and HOMME ('man'). However, it is not completely clear where the borders between these stems and pure CFs should be drawn. Is it legitimate, for instance, to consider the

French CFs *-graphe* and *-cide* as suppletive stems of the verbs ÉCRIRE ('write') and TUER ('kill')?;

(ii) the second group are CFs which are not of direct classical descent, but originated from autonomous lexemes, often by truncation. Examples of this type in French are *pétro-* (< *pétrole* 'oil'), *ciné-* (< *cinéma(tographe)*), but also *-gate*, which was, of course, borrowed from English but is nowadays widespread in French. Corbin (2004) calls these elements "fractoconstituants", as opposed to "archéoconstituants" (i.e. genuine CFs). However, the two groups are not necessarily distinguished in speakers' competence, and some elements may belong to both categories with different meanings, like *télé-* ('distance' in *télétravail* 'telecommuting' and 'television' in *téléreportage* 'television report') or *photo-* ('light' in *photosensible* 'photosensitive' and 'photograph' in *photojournaliste* 'photojournalist').

In our view, all the above-mentioned facts suggest that CFs can hardly be considered as a homogeneous class of elements and, consequently, that all NCs, in particular those which are productively constructed, cannot be considered as formed by a morphological process of the same type. Space is lacking here to go deep into this matter, and we just mention the hypothesis that different NCs may correspond to different construction patterns which are at different stages of grammaticalisation, from actual compounding to more affixation-like phenomena. What we want to show is that for some patterns, such as those considered below, the construction of the meaning of NCs does not differ substantially from that of constructs which are undisputedly formed by affixation. Although we are not explicitly addressing the question of the nature of these CFs as affixes or (stems of) lexemes, we do not see any serious argument for distinguishing them from other, more typical, examples of exponents of constructional schemas in Booij's (2010) sense.

4. The data

In what follows we consider in particular French lexemes containing CFs which are organised in two pairs: *-crate* / *-cratie* and *-logue* / *-logie*. These lexemes share several interesting properties. Both pairs include final neoclassical elements borrowed from Greek nouns: the former from κράτος 'power' and the latter from λόγος 'discourse'.[3] The series formed by means of the CFs in question are parallel. On the one side, *-logie* and *-cratie* (henceforth [+ie][4] constituents) mainly form [−human] abstract nouns. Most nouns in *-logie* designate

165

a speciality or a discipline, while nouns in *-cratie* contain the original notion of 'power'. On the other side, *-logue* and *-crate* (henceforth [–ie] constituents) mainly form [+human] nouns or adjectives. Most [+ie] and [–ie] lexemes are in correspondence, the latter designating, respectively, the specialist of a particular discipline and a person exerting some form of power.

One could consider that [+ie] lexemes are derived from [–ie] lexemes by suffixation of the element *-ie* (phonologically /i/). Ancient Greek and Latin possessed a suffix *-ia*, which constructed abstract nouns from human nouns or adjectives (cf. Greek φιλόλογος 'scholar' > φιλολογία 'reasoning'). Synchronically, in French some quality nouns (*fou*$_A$ 'mad' > *folie*$_N$ 'madness') and some nouns designating a place (*boucher*$_N$ 'butcher' > *boucherie*$_N$ 'butcher's shop') also contain the suffix *–ie*.[5] However, as we will see below, each of the CFs considered is the exponent of a construction pattern which has different, only partially overlapping, meanings. Moreover, not all [–ie] lexemes have a [+ie] correspondent, and *vice-versa*. These gaps may be fortuitous (as in (9)), or be a consequence of the semantic instruction attached to each word formation pattern (as in (10)):[6]

(9) a. *argencratie < argent* 'money' 'state or territory ruled by money'
    b. °*argencrate* 'member of such a territory'

(10) a. *macdocratie < macdo* 'McDonald's' + *-cratie* 'omnipresence of McDonald's restaurants'
    b. **macdocrate* 'Macdocrat'

In (10a) the derivate belongs to a class of lexemes having the semantic instruction 'omnipresence of X', which does not seem compatible with the *-crate* WFP. We consider, thus, that the non-attestation of °*argencrate* in our corpus is fortuitous, while that of **macdocrate* is not (of course, °*macdocrate* would be acceptable with the meaning 'someone who exerts power by means of McDonald's restaurants', a reading, however, which is pragmatically less plausible[7]). This irregular distribution of meanings between [–ie] and [+ie] lexemes constitutes a further argument supporting the hypothesis that we have here four CFs which are the exponents of four different word formation patterns, and not two CFs plus a rule of *–ie* suffixation.

As the main focus of this article is semantics, we do not linger over formal observations. We limit ourselves to observing that the emergence of /ɔ/ as a linking vowel represents a strong constraint on the output form of the NCs in question. On average, more than

87% of the lexemes in our corpus display an /ɔ/ in the expected position. This vowel can be already present in the constituent to the left, either because this element is borrowed from Ancient Greek (e.g. *psycho-* in *psychologie* 'psychology') or because it is a French lexeme that contains it (e.g. *escargologie* < *escargot* 'snail' + *-logie*). When the left constituent does not contain this vowel, the output form can display it (e.g. *rockologie* < *rock* 'rock' + *-logie*) or not. In the latter case, with few exceptions, the neoclassical compound displays a different vowel at the junction of the two constituents (e.g. *mafiacratie* < *mafia* 'mafia' + *-cratie*). If a lexeme longer than three syllables contains an /ɔ/ in the second or in the third syllable, it can be truncated at this point, so that the formal constraint in question is respected (e.g. *consocratie*[8] < *consommer* 'consume'), as well as size constraints.[9] From the data in our corpus it emerges, in fact, that *-crate* / *-cratie* are preferentially preceded by a disyllabic element, and *-logue* / *-logie* by a trisyllabic one (Figure 2). This difference can be explained by the number of constructions involving three elements, i.e. in which the base is constituted by two elements (for example *psychopathologie* 'psychopathology') which are much more important for *-logie* and *-logue* than for *-cratie* and *-crate*.

The left element can also be present in the form of a learned suppletive stem. For instance, in *gémellocratie*, we consider that the left element (phonologically /ʒemel/) is a learned stem of the lexeme JUMEAU ('twin'); note that the popular stem (/ʒymo/) would have respected both size and form constraints. Incidentally, the emergence of two competing forms is not excluded either, as in the pair *nivologie*
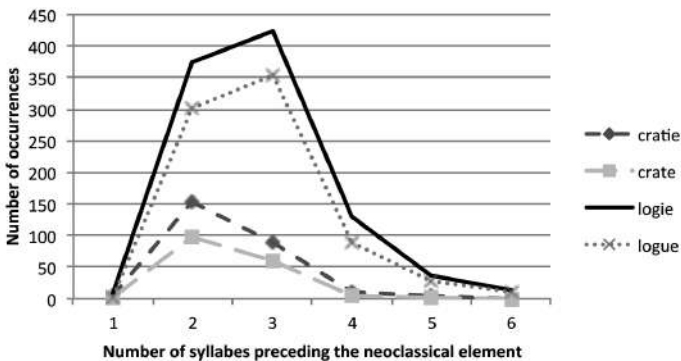

Fig. 2. Number of lexemes in *-cratie*, *-crate*, *-logie* and *-logue* according to the size of the base.

/ *neigeologie*, which are constructed, respectively, on a learned (/niv/) and on a popular (/nɛʒ/) stem of NEIGE ('snow'). The first (*nivologie*) corresponds to the denomination of a recognized scientific discipline, whereas the second (*neigeologie*) is more likely perceived as the name of an imaginary or mock specialty. The difference in stem origin here corresponds to a difference in register.

NCs were indeed originally created in order to enhance the technical vocabulary, in a period of great scientific and technical progress, between the 17th and the 19th centuries. Scientists needed new words to name new disciplines, methods, phenomena, objects, etc. Today, the creation of NCs is no longer restricted to learned technical registers, and several of these formation patterns have lost their original function and have been taken over by everyday language. Among the motivations for this spreading of NC formation patterns outside the technical sphere there are certainly stylistic factors, such as the desire of imitating, mostly in an ironic way, scientific or technical vocabulary. This is evident, for instance, in some constructs like *moustachologue* (< *moustache* 'mustache') or *chevillologue* (< *cheville* 'ankle'), which are constructed on the same schema as other specialists' names, like *dermatologue* ('dermatologist'), although they are implausible in the technical vocabulary, due to the lack, in our societies, of specialists specifically dedicated to moustaches or to ankles (see also the following example taken from the Web (< *chaussette* 'sock')):

(11) *Et pourtant il va bien falloir les différencier. C'est là qu'il faut avoir toute la finesse d'analyse et toute l'expérience du chaussettologue distingué (moi !), se basant sur d'infimes détails: nuance de noir, degré d'usure, épaisseur de l'élastique, souplesse du tissu…*
      'Yet, we have to differentiate them. This is where you need all the analytical finesse and all the experience of a distinguished sockologist (like me!), observing every tiny detail: the exact tone of black, the degree of wear, the thickness of the elastic, the flexibility of the fabric…'
      [http://community.weightwatchers.fr/Blogs/ViewPost.aspx?threadID=1047287]

The availability of NC formation patterns in everyday language makes both compounds that are listed in dictionaries and those that are not interesting for the analysis. In order to observe their variety, we constructed a corpus from different resources. First, we collected the NCs containing the elements *-logue*, *-logie*, *-crate* and *-cratie* in two dictionaries of French: Le Trésor de la Langue Française informatisé (TLFi) and Le Grand Robert (online version). Then, we used

the Google Ngrams dataset for French[10] in order to extract lexemes containing the same sequences not recorded in the above-mentioned dictionaries. Table 2 gives the number of lexemes for each type of NCs collected in the different resources.

Table 2. Number of lexemes collected in the different resources.

	[−IE]		[+IE]		[−IE]		[+IE]	
	-crate		*-cratie*		*-logue*		*-logie*	
Dictionaries	18	11 %	30	12 %	226	29 %	535	54 %
Google	146	89 %	230	88 %	562	71 %	455	56 %
Total	164		260		788		990	

The drawbacks of using the Web as a corpus for linguistic studies are well-known and well-described (cf. Kilgariff 2007, Lüdeling et al. 2007, Hathout et al. 2008). However, our goal here is to study lexical creativity, which can be accessed mainly by observing how the lexicon is constructed on line by speakers. As the data in Table 2 show, due to its size, the Web allows access to a number of constructed words, including neologisms, which cannot be equalled by any other lexical resource, not even the largest dictionaries available.

5. Complex lexemes: a semantic network

The examination of our corpus confirms that the meaning of the NCs in question is not constructed in a purely compositional way and supports the idea that it is constructed, rather, in connection with that of other lexemes in the lexicon. The following sections present the different ways in which this connection can be made.

5.1. Leader words: -cratie / -crate

The notion of 'leader word' belongs to the philological tradition, and it is intended to designate a form which is particularly salient for the analogical restructuring of other words (cf. Malkiel 1966, among others). More recently, some scholars have adapted this label to lexicological and morphological studies, in order to designate a lexeme which, in a lexical network, functions as a catalyst for a derivational pattern, by virtue of its origin, its frequency, its saliency, etc. (cf. in particular Rainer 2003, Roché 2011, Roché & Lignon 2011). Semantically, this means that when a new lexeme is created, it has to be placed in relation to the leader word. When a large number of new lexemes depart from the leader word, a new (sub)pattern can be born,

and it can in turn function as a catalyst for other constructions.

Let us take, for instance, the lexemes containing the elements *-cratie* and *-crate*. These derivates do not constitute a homogenous class. It is generally admitted, including in 'popular' metalinguistic consciousness, that *-cratie* means 'power'. A finer observation of the data, however, shows that two leader words exist, each of which is the kernel of a distinct semantic pattern. The first of them is *démocratie*, which means 'the power of people' and whose first appearance in French goes back to 1370. A Google search realized on May 3, 2013, yielded more than 3 million (estimated) indexed pages containing the word *démocratie* in French. This lexeme gave rise to the semantic pattern 'power of X'. The second leader word for *-cratie* lexemes is *aristocratie*, which was also first attested in French in 1370.[11] On May 3, 2013, more than 1.5 million pages in French contained this word. Although the original meaning of *aristocratie* ('power of the best') also contains a reference to the notion of 'power', in the late 18th century this lexeme started to designate the noble class, from which the meaning of a social group with a self-identified high value developed.

The two main patterns described above, in their turn, gave birth to new subpatterns. For instance, the pattern carrying the meaning 'power of X' presents different nuances in various cases (12):

(12)   a. *corruptocratie*   < corruption 'bribery'   'power exerted by means of bribery'
       b. *fricocratie*   < fric 'money'   'power exerted in order to get money'
       c. *thalassocratie*   < thalass- 'sea'   'maritime supremacy'

In each of these subpatterns the left element adds an argument to the predicate expressed by the construction, indicating an instrument, a goal or a place. Another pattern that emerged includes lexemes which are exocentric and designate a 'territory ruled by X', X being often the name of an individual (13-14):

(13)   a. *bushocratie*   < (George W.) Bush
       b. *sarkocratie*   < (Nicolas) Sarkozy

(14)   *Bienvenue en Sarkocratie*
       'Welcome to Sarkocracy'
       [http://www.pantheresroses.org/Pourquoi-pas-encore-plus-de.html]

The pattern 'social group' generated a new subpattern whose semantic instruction corresponds to 'omnipresence of X' (15-16) (see also *macdocratie* in (10) above).

(15)   *footocratie*   < foot(ball) 'football'   'omnipresence of football'

170

(16) *Ça fait 2 semaines qu'on se bouffe du foot à la télé, dans les bars, PARTOUT ! NON A LA FOOTOCRATIE !*
'For the last two weeks we've only seen football on TV, in bars, everywhere! No to footballocracy!'

Some of the patterns described for *-cratie* lexemes possess a corresponding pattern for *-crate* lexemes. For instance, the semantic feature 'power' (cf. (12)) is present in several *-crate* lexemes which designate a 'partisan of the state ruled by X', while other *-crate* lexemes designate the member of a group with a self-identified high value (i.e. belong to the *aristocratie* type). No pattern for *-crate* lexemes corresponds, on the other hand, to the meaning 'omnipresence of X'. A series of lexemes in *-crate*, which designate a person who grants him/herself an undeserved power, seems to lack a corresponding *-cratie* series.

Figure 3 summarizes the patterns we identified for *-cratie* and *-crate* lexemes; the semantic instructions corresponding to the former are indicated in bold, while those corresponding to the latter are indicated in italics:

It is obvious that for a precise description of the semantics of a complex lexeme, and consequently of a WFP, it is necessary to take into account the contexts in which it appears. We can observe, for instance, that the same construction may correspond to different patterns depending on the context. *Mafiacratie* ('mafia' + *-cratie*) can designate the 'power of the mafia' (17), a 'territory ruled by mafiosos' (18) and also a 'group of mafiosos' (19).[12]
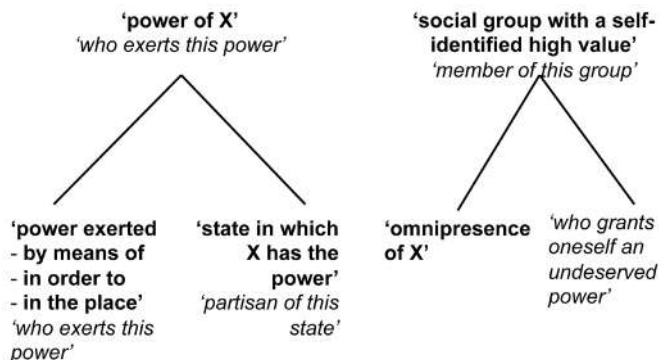

Fig. 3. Patterns identified for *-cratie* and *-crate* lexemes.

171

(17) *Quand on a des convictions, on les défend à visage découvert, pas caché derrière un masque: dans une vraie démocratie tout à fait d'accord. Hélas pas dans une mafiacratie ou un état qui ne reconnait pas son peuple après chaque élection, ce serait un geste suicidaire et stupide.*
'When one has convictions, one defends them overtly, not behind a mask – in a true democracy, I completely agree with that. Alas, in a mafiacracy or in a state that does not recognize its citizens after each election, it would be a suicidal and stupid act.'
[http://www.lepoint.fr/debats/anonymous-bandits-ou-justici-ers-18-02-2012-1432682_34.php]

(18) *La France est une vraie oligarchie et très près d'une mafiacratie et c'est ça la différence, le peuple n'a absolument rien a dire.*
'France is a true oligarchy, it is very close to a mafiacracy and this is the difference, people have absolutely nothing to say'
[http://plus.lefigaro.fr/article/la-france-perd-son-aaa-20120113-661943/commentaires/8397506]

(19) *En rapides tableaux, la caméra montre le 'casse légal du siècle' opéré par l'ex-président argentin, avec la complicité active de la 'mafiacratie' : ministres, parlementaires, industriels, syndicalistes...*
'In brief scenes, the camera shows the 'legal heist of the century' carried out by the Argentinean former president, with the active complicity of the 'mafiacracy': ministers, members of Parliament, businessmen, trade unionists...'
[http://web.radicalparty.org/pressreview/print_250.php?func=detail&par=9458]

For this reason, some lexemes from our corpus are counted several times in Figures 4 and 5, which give the number of complex lexemes for each pattern. In this work, we do not take a definite position whether the different interpretations of units like *mafiacratie* should be considered as the different meanings of a polysemous lexeme or are sufficient to identify several homonymous lexemes.

The patterns identified function as poles of attraction to which each actually observed NC can approach more or less closely.

5.2. -logie / -logue: A distributed organization

If we now look at the lexemes containing the CFs *-logue* and *-logie*, we observe that their meaning is constructed in the same way, i.e. on the basis of the existing lexicon, in connection with some attracting poles. Unlike the case of the NCs seen above, however, these attracting poles are not organised around individual leader words. Rather, these patterns display a wider range of differ-
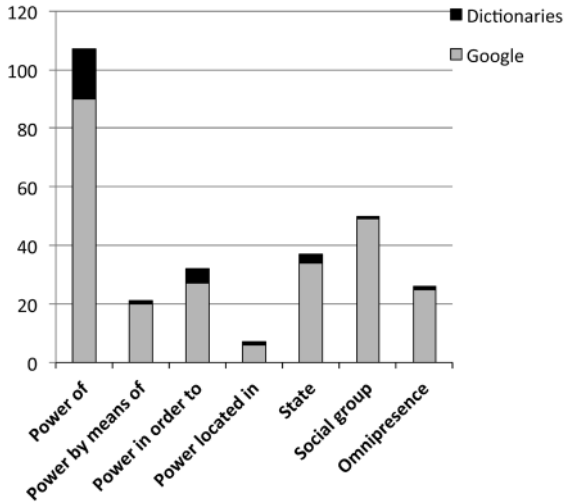
Fig. 4. Distribution of complex lexemes for each *-cratie* pattern.
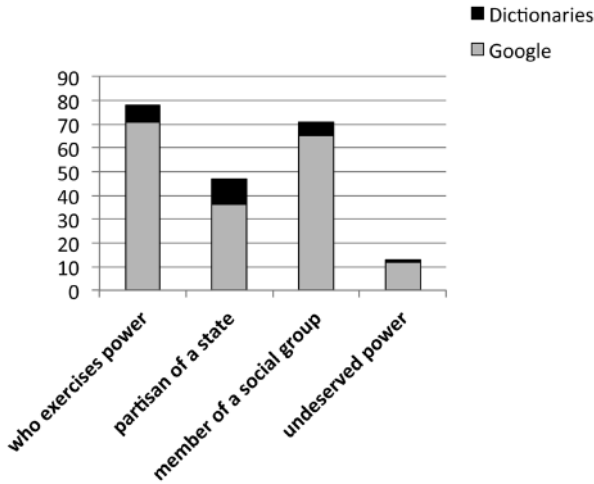
Fig. 5. Distribution of complex lexemes for each *-crate* pattern.

ent but connected meanings. The examination of our corpus allows us to identify four main classes of lexemes constructed with the element *-logie*. The most productive of these patterns is the one that forms
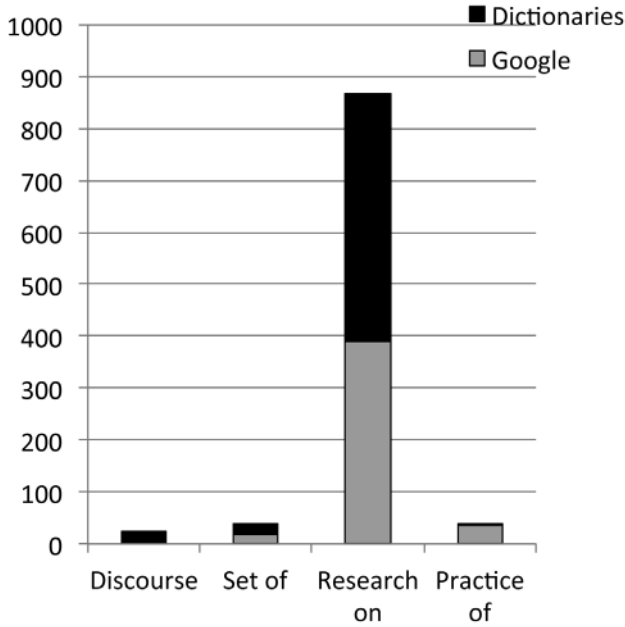
Fig. 6. Distribution of complex lexemes for each *-logie* pattern.

nouns meaning 'research on, study of X'. Besides this pattern, we find two other constructions which are available; one seems to be derived from the previous one and has the meaning 'practice of X'; the other one, carrying the meaning 'set of X', conserves the original Greek meaning 'collect'. Finally, a fourth pattern seems not to be productive anymore; its meaning is related to the original meaning of the Greek lexeme meaning 'discourse'. Figure 6 gives the numbers of occurrences for each of the main patterns.

Each of the identified semantic patterns can be divided into different subpatterns. Hence, for instance, for the meaning 'discourse' several different arguments can be instantiated by the left element: a modality (20a), an instrument (20b) or a content (20c).

(20) A. *hypologie*   'disease preventing saying words longer than two syllables'
      B. *dactylologie*   'typed up discourse'
      C. *nécrologie*   'death notices'

The semantic pattern 'set of X' is more concentrated around the leader word *anthologie*. Its right element is linked to the Greek verb λέγειν, ('collect') (see note 5 above). Thus, in Modern French we found a number of lexemes in *-logie* denoting a collection of objects (21). A more specific pattern linked to this one denotes a 'set of *n* artistic works' (22). For instance, the non-listed lexeme *septalogie* (22 b) fits into this pattern.

(21)   A. *anthologie*   'anthology'
       B. *critérologie < critère* 'criterion'   'set of criteria'

(22)   A. *trilogie*   'trilogy'
       B. *tétralogie*   'tetralogy'
       C. *septalogie < sept* 'seven'   'set of seven works'

Also, depending on the semantic value of X, a wide range of connected meanings is attached to the pattern 'research on X', as can be seen in Table 3.

Table 3. Semantic types found with the meaning 'research on X'.

MEANING OF THE LEFT ELEMENT	EXAMPLES	GLOSS
disease	*cancérologie* *ampoulologie*	< cancer 'cancer' < ampoule 'blister'
part of the body	*cardiologie* *chevillologie*	< cardio- 'heart' < cheville 'ankle'
concrete object	*sédimentologie* *chaussettologie*	< sédiment 'sediment' < chaussette 'sock'
abstract object	*irénologie* *problémologie*	< irén- 'peace' < problème 'problem'
phenomenon	*courantologie* *neigeologie*	< courant 'stream' < neige 'snow'
individual	*mozartologie* *batmanologie*	< Mozart < Batman
activity, discipline	*technologie* *régimologie*	< techn(o)- 'technique' < régime 'diet'
animal	*primatologie* *delphinologie*	< primate 'primate' < delphin- 'dolphin'
plant	*algologie* *orchidologie*	< algue 'algae' < orchidée 'orchid'
place, country, people	*américanologie* *balkanologie*	< américain 'American' < Balkans 'Balkans'

All these subpatterns are interconnected: for instance, the distinction between a disease and a phenomenon can be tiny. Is *calvitie*

'baldness' in *calvitologie* seen as a disease or as a phenomenon? Only context can help to decide and the same constructed lexeme can belong to different subpatterns.

This distributed organisation is dynamic and in this case too, a lexicalized compound may become the source of a new subparadigm. We can observe it with *œnologie* 'oenology', which is the source of a small subpattern of lexemes meaning 'study of an alcoholic drink':

(23) A. *vinologie*   < vin 'wine'   'study of wine'
      B. *biérologie*   < bière 'beer'   'study of beer'
      C. *cidrologie*   < cider 'cider'   'study of cider'

The fourth and last pattern, 'practice of X', is more recent and contains almost exclusively non-dictionary lexemes. Within it, lexemes are also organized into several subpatterns according to the semantic value of X, as in:

(24) (I) 'practice of an activity'
      A. *siestologie*   < *sieste* 'nap'
      B. *piscinologie*   < *piscine* 'swimming pool'

      (II) 'practice of a type of discourse'
      A. *excusologie*   < excuse 'apology'   'great practice of apologizing'
      B. *blablatologie* < blabla 'blahblah'   'great practice of talking for the sake of talking'

The subpattern (24.i) could be derived from the main pattern 'study of X', while the subpattern (24.ii) is very close to the pattern 'discourse characterized by X' and the nuance becomes clear only by virtue of the context. Therefore, the distinction between all patterns and subpatterns is not clear-cut. Once again, the same lexeme can be interpreted in different ways according to its context. *Piscinologie* can mean 'the practice of swimming' (25) or 'the study of swimming pools' (26):

(25) *La piscinologie m'avait été fortement déconseillée ces derniers mois par le corps médical, en des termes laissant peu de place à la fantaisie : "c'est la piscine ou la médecine".*
      'I had been strongly discouraged from swimming in the last months by the medical profession, with terms which left little room for imagination "it's the swimming pool or medicine" '
      [http://yaelleliane.blogspot.fr/2008/11/comme-une-sirne-sur-le-sable.html]

(26) *Un professeur en piscinologie pourrait il me dire quelle est la durée de vie du chlore lent?*
'Could a specialist of pool-ology tell me the lifespan of slow chlorine'
[http://www.eauplaisir.com/annuaire/forums/ftopic_piscine9553.html&highlight=]

The [–ie] constituent corresponding to *-logie* is *-logue*. When *Xlogie* designates a scientific discipline, *Xlogue* designates a specialist of this discipline (e.g. *cardiologie* 'cardiology' / *cardiologue* 'cardiologist'). When *Xlogie* designates a 'set of X', *Xlogue* designates a member of this set, like *conologue* (< *con* 'bloody idiot') in (27).

(27) *Sur les plateaux télés, chez nous, les politologues, les sociologues et autres conologues se déchaînent*
'In TV studios, in our country, political specialists, sociologists and others idiotologists are are going wild'
[http://www.paperblog.fr/4152796/plantu-fin-au-pays-des-droits-de-l-homme/]

6. Conclusion

The analysis we presented shows that the meaning of NCs cannot be seen simply as the combination of the meanings of their elements, i.e. that derivational morphology does not function in an additive manner, but consists primarily of inserting a new lexeme into a complex lexical network. This holds in particular for the semantic properties of complex lexemes, which can only be determined on the basis of their similarity to, or distance from, other lexemes in the lexicon. The analysis we conducted on two pairs of NCs shows that an apparently unique word formation pattern can actually correspond to different semantic schemas. Hence, the two parallel series of NCs we observed are similar in that the meaning of the lexemes they include can only be described as the result of the inclusion of a lexeme in a global lexical network. They differ, however, in the manner in which this inclusion is realized. New lexemes belonging to both patterns are modelled on the basis of the existing lexicon. However, for the *-cratie/ -crate* set the model is represented by individual leader words which serve as semantic catalysts; for the *-logie/ -logue* set the semantic models are 'distributed', as no individual lexeme emerges as the unique model for other words. The schemas that emerge should not be seen as closed classes but rather as poles of attraction; for each lexeme in a given context, we should be able to measure its proximity

to or its distance from an attracting pole. Moreover, we showed that the analysis of isolated lexemes is not sufficient for understanding the functioning of a morphological process. For this reason, we consider the set up and the realisation of large-scale corpus-based observations of constructed lexemes together with their contexts to be of great importance, as one of the currently most promising approaches to the study of the complex lexicon.

Addresses of the Authors

UMR 5263 CLLE-ERSS, Maison de la Recherche, Université de Toulouse II, 5, allées Antonio Machado, 31058 Toulouse Cedex 9, France
<marine.lasserre@univ-tlse2.fr>
<fabio.montermini@univ-tlse2.fr>

Notes

[1]   Note, however, that languages do not fully exclude the possibility that an element can function alternatively as a prefix or as a suffix. This includes cases of derivational and also of inflectional morphology, from which NCs are, by definition, excluded (see, for instance, the so-called 'Wackernagel affixes', whose position is determined with relation to the whole word in which they appear and not with the base they attach to (Nevis & Joseph 2007, Bonami & Stump ms.)).

[2]   We considered that a base was autonomous if it also had a realization as an autonomous syntactic atom in French. Thus, *théâtrologie* was analyzed as constructed on the lexeme THÉÂTRE rather than on the Greek CF *théâtr(o)-*.

[3]   Actually, the constituent *-logie* has two distinct origins in Ancient Greek: the noun λόγος 'discourse' and the verb λέγω 'collect, gather'. Both these meanings can be found in the French constituents *-logue* and *-logie*, although we consider that the two original meanings conflated into a single constituent.

[4]   The sole function of this label is to distinguish the two parallel series of NCs, and no assumption is made on the precise nature of the *-ie* sequence whose suffixal status is disputable (see below).

[5]   In fact, words like *boucherie* are ambiguous, as the sequence *-erie* is indisputably a unitary suffix in other cases such as *conciergerie* ('concierge's office'), *piraterie* ('piracy'), etc.

[6]   Following a tradition of the French literature on morphology, we use ° as a label for a non-attested but possible lexeme.

[7]   A reviewer suggested the interpretation "an individual who advocates the Macdonaldization of the planet, i.e. the expansion of the presence of MacDonald's on the planet" for *macdocrate*. This interpretation might of course be possible but does not exactly correspond, for us, to *macdocratie* as the 'omnipresence of MacDonald's restaurants' but rather to *macdonaldization* as a process.

[8]   *Consocratie* can also be considered as having been coined from the agentive noun *consommateur* or even from its clipped form *conso*, which is common in colloquial French. The process consisting of reducing a word to its first two or three syllables and adding a final *-o* is frequent in spoken French both for common and

proper nouns (*Sarko* < *Sarkozy*, *intello* < *intellectuel* 'intellectual', *proprio* < *propriétaire* 'owner') (cf. Kilani-Schoch 1996, Fridrichová 2013). These clippings are frequent, in our corpus, as bases for *-cratie*, *-crate*, *-logie* and *-logue* constructions, as they are similar to neoclassical bases in some respects (in particular bi- or tri-syllabic format and ending in /o/).

[9] For the role played by size constraints in the derivational morphology of French, see in particular Plénat (2009); Plénat & Roché (2003). Note that sensitivity to size constraints constitutes a further argument supporting the view of CFs as exponents of constructions similar to affixes.

[10] http://storage.googleapis.com/books/ngrams/books/datasetsv2.html.

[11] The TLFi dictionary gives as first attestations for both *démocratie* and *aristo-cratie* the work *Ethiques* by Oresme, published in 1370.

[12] As the example *mafiacratie* shows, taking the actual contexts of use of complex lexemes (not only those of neoclassical compounds) is necessary in order to precisely determine the range of their meanings and their interactions. In Lasserre & Montermini (to appear a, b) we propose an analysis of words constructed with the elements *-cide* and *-logique*, based on large-scale observations of the contexts of appearance of constructed lexemes in some corpora of French and on the Web.

Bibliographical References

AMIOT Dany & Georgette DAL 2007. Integrating neoclassical combining forms into a lexeme-based morphology. In BOOIJ Geert, Luca DUCCESCHI, Bernard FRADIN, Emiliano GUEVARA, Angela RALLI & Sergio SCALISE (eds.). *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5), Fréjus, 15-18 September 2005*. Bologna: Università degli Studi di Bologna. 323-336.

ARONOFF Mark 1976. *Word Formation in Generative Grammar*. Cambridge, MA: MIT Press.

ARONOFF Mark 2007. In the beginning was the word. *Language* 83.4. 803-829.

BLEVINS James P. & Juliette BLEVINS 2009. Introduction: Analogy in grammar. In BLEVINS James P. & Juliette BLEVINS (eds.). *Analogy in Grammar*. Oxford: Oxford University Press. 1-12.

BONAMI Olivier & Gregory T. STUMP ms. *Paradigm function morphology*. Manuscript. Université Paris-Sorbonne/University of Kentucky.

BOOIJ Geert 2010. *Construction Morphology*. Oxford: Oxford University Press.

CORBIN Danielle 1989. Form, structure and meaning of constructed words in an associative and stratified lexical component. In BOOIJ Geert & Jaap VAN MARLE (eds.). *Yearbook of Morphology 1989*. Dordrecht: Foris. 31-54.

CORBIN Danielle 1991. La formation des mots: structures et interprétation. *Lexique* 10. 7-30.

CORBIN Danielle 2001. Préfixes et suffixes: du sens aux catégories. *Journal of French Language Studies* 11.1. 41-69.

CORBIN Danielle 2004. Français (Indo-européen: Roman). In BOOIJ Geert, Christian LEHMANN, Joachim MUGDAN & Stavros SKOPETEAS (eds.). *Morphologie / Morphology. An International Handbook on Inflection and Word-Formation. 2.* Berlin/New York: de Gruyter. 1285-1299.

FREGE Gottlob 1980 [1914]. Letter to Jourdain. Philosophical and math-

ematical correspondence, in Gabriel GOTTFRIED, Hans HERMES, Friedrich KAMBARTEL, Christian THIEL & Albert VERAART. Chicago: Chicago University Press. 78-80.

FRIDRICHOVÁ Radka 2013. Quelques observations sur les mots tronqués dans le français contemporain. *Romanica Olomucensia* 25.1. 1-13.

HATHOUT Nabil, Fabio MONTERMINI & Ludovic TANGUY 2008. Extensive data for morphology: Using the World Wide Web. *Journal of French Language Studies* 18.1. 67-85.

KILANI-SCHOCH Marianne 1996. Syllable and foot in French clipping. In HURCH Bernhard & Richard A. RHODES (eds.). *Natural Phonology: The State of the Art. Papers from the Bern Workshop on Natural Phonology*. Berlin/New York: Mouton de Gruyter. 135-152.

KILGARIFF Adam 2007. Googleology is bad science. *Computational Linguistics* 33.1. 147-151.

LASSERRE Marine & Fabio MONTERMINI (to appear a). Une approche distributionnelle pour l'analyse des composés néoclassiques. *Verbum*.

LASSERRE Marine & Fabio MONTERMINI (to appear b). Neoclassical compounds as relational adjectives. *Proceedings of 9th Mediterranean Morphology Meeting, Dubrovnik (Croatia), September 15-18*.

LÜDELING Anke, Stefan EVERT & Marco BARONI 2007. Using Web data for linguistic Purposes. In HUNDT Marianne, Nadja NESSELHAUF & Carolin BIEWER (eds.). *Corpus Linguistics and the Web*. Amsterdam/New York: Rodopi. 7-24.

MALKIEL Yakov 1966. Genetic analysis of word-formation. In SEBEOK Thomas A. (ed.). *Current Trends in Linguistics. III. Theoretical Foundations*. The Hague: Mouton. 305-364.

NAMER Fiammetta & Florence VILLOING 2007. Have cutthroats anything to do with tracheotomes? Distinctive properties of VN vs. NV compounds in French. In BOOIJ Geert, Luca DUCCESCHI, Bernard FRADIN, Emiliano GUEVARA, Angela RALLI & Sergio SCALISE (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5), Fréjus, 15-18 September 2005*. Bologna: Università degli Studi di Bologna. 105-124.

NEVIS Joel A. & Joseph BRIAN 1992. Wackernagel's affixes: Evidence from Balto-Slavic. In BOOIJ Geert & Jaap VAN MARLE (eds.), *Yearbook of Morphology 1992*. Dordrecht: Foris. 93-111.

PLAG Ingo 2004. Syntactic category information and the semantics of derivational morphological rules. *Folia Linguistica* 38.3-4. 193-225.

PLÉNAT Marc 2009. Les contraintes de taille. In FRADIN Bernard, Françoise KERLEROUX & Marc PLÉNAT (eds.). *Aperçus de Morphologie du Français*. Saint-Denis: Presses Universitaires de VIncennes. 47-63.

PLÉNAT Marc & Michel ROCHÉ 2003. Prosodic constraints on suffixation in French. In BOOIJ Geert, Janet DECESARIS, Angela RALLI & Sergio SCALISE (eds.). *Topics in Morphology. Selected Papers from the Third Mediterranean Morphology Meeting*. Barcelona: Universitat Pompeu Fabra. 285-299.

RAINER Franz 2003. Semantic fragmentation in word-formation: The case of Spanish *-azo*. In SINGH Rajendra & Stanley STAROSTA (eds.). *Explorations in Seamless Morphology*. New Delhi/Thousand Oaks/London: Sage Publications. 197-211.

Rᴀɪɴᴇʀ Franz 2012. Morphological metaphysics: Virtual, potential and actual words. *Word Structure* 5.2. 165-182.

Roᴄʜé Michel 2009. Pour une morphologie *lexicale*. *Mémoires de la Société de Linguistique de Paris. Nouvelle Série* 17. 65-87.

Roᴄʜé Michel 2010. Base, thème, radical. *Recherches linguistiques de Vincennes* 39. 95-134.

Roᴄʜé Michel 2011a. Quelle morphologie? In Roᴄʜé Michel, Gilles Boʏé, Nabil Hᴀᴛʜoᴜᴛ, Stéphanie Lɪɢɴoɴ & Marc Pʟéɴᴀᴛ (eds.). *Des Unités Morphologiques au Lexique*. Paris: Hermès-Lavoisier. 15-39.

Roᴄʜé Michel 2011b. Quel traitement unifié pour les derivations en *-isme* et en *-iste*? In Roᴄʜé Michel, Gilles Boʏé, Nabil Hᴀᴛʜoᴜᴛ, Stéphanie Lɪɢɴoɴ & Marc Pʟéɴᴀᴛ (eds.). *Des Unités Morphologiques au Lexique*. Paris: Hermès-Lavoisier. 15-39.

Roᴄʜé Michel & Stéphanie Lɪɢɴoɴ 2011. Quelle distribution pour *-éen* vs. *-ien*? In Roᴄʜé Michel, Gilles Boʏé, Nabil Hᴀᴛʜoᴜᴛ, Stéphanie Lɪɢɴoɴ & Marc Pʟéɴᴀᴛ (eds.). *Des unités morphologiques au lexique*. Paris: Hermès-Lavoisier. 191-250.

Šᴛᴇᴋᴀᴜᴇʀ Pavol 1998. *An Onomasiological Theory of English Word-Formation*. Amsterdam: John Benjamins.

Vɪʟʟoɪɴɢ Florence 2012. French compounds. *Probus* 24. 29-60.