# Lexical creativity and the organization of the lexicon*

Teresa Vallès
Universitat Internacional de Catalunya
Universitat Oberta de Catalunya

The aim of this paper is to show that lexical creativity and the organization of the lexicon are "mutually enlightening" realities. Using of Catalan neologisms, I take advantage of Bybee's (2001: 109) insight that "any multimorphemic word or sequence is highly embedded in connections with other words containing at least one of the same morphemes." I argue for the validity of Bybee's model over and beyond its use in inflectional processes, by developing an analysis of neologisms in which derivation and analogy play a key role. In the light of the theory on paradigmatic relations framing the lexicon (cf. van Marle 1985), I advance the hypothesis that lexical creativity is the external manifestation of the lexicon's intrinsic organization. Specifically, I show the dynamic role that two main types of paradigmatic relations (derivational categories and word families) play in lexical creativity. Because both of them are based on morpheme-to-morpheme relations among words, I emphasize this kind of relations in network representations. In particular, I show that low-level network patterns used in neology are suggestive evidence of a fuzzy boundary between derivation and analogy.

**Keywords:** lexical creativity, paradigmatic relations, morphological relations, analogy, derivation

## Introduction

One of the main goals of morphological theory is to characterize the speakers' capacity to use words that they have not heard previously. In this sense, generative morphological models attempt to make explicit the *mechanisms* of linguistic competence that allow speakers to generate all and only possible

*Annual Review of Cognitive Linguistics* 1 (2003), **137–160.**
ISSN 1572–0268 / E-ISSN 1572–0276 

words (i.e., lexical units that are grammatically well formed, whether or not they actually exist in that particular language). In such an approach to lexical creativity, the lexicon has no internal structure based on word formation patterns (Di Sciullo and Williams 1987: 15). Moreover, the grammatical model need not take into account speakers' *use* of these mechanisms, but only their knowledge of the lexical-creation rules. In short, speakers' use of actual words is not relevant grammatical data. As I will show in the next section, generative approaches to lexical creativity suffer from this paradoxical lack of grammatical relevance of the neologisms. Moreover, once performance is excluded, the model runs the risk of no longer being susceptible to empirical evidence.

In this article, I shall explore an alternative approach to lexical creativity, namely that of a usage-based model of actual words (Langacker 1987, 1988, 1991, 1999; Bybee 1985, 1988, 1995, 2001). Instead of emphasizing possible words at the expense of performance, this alternative takes as its starting point the actual words speakers use, so as to elucidate the interaction between competence and performance. This stands in stark contrast with classical generative accounts, such as the one outlined above, in which lists (i.e., the lexicon) and rules (i.e., morphology) are cleanly distinguished. Following Bybee, my account argues that the patterns used for the lexical creation contribute decisively to the structuring of the lexicon: "they are patterns that emerge from the intrinsic organization of the lexicon" (Bybee 1988: 125).

The present study defends Bybee's (1985, 2001) version of the network model. Because this author has focused almost exclusively on inflection and phonology, some authors have cast their doubts about the validity of her proposal for word formation. For example, Beard (1998: 50) voices his opinion that "many of the processes vital to Bybee's model remain undefined, so it is not currently possible to determine this theory's efficacy in accounting for derivational data." Moreover, according to Beard (1998: 46), in this model "derivation is a set of statistic paradigmatic lexical relations." My goal is to develop Bybee's model within the scope of neology by derivation and analogy in Catalan, providing further evidence for the validity of the model beyond inflectional morphology.

In the following section, I outline briefly the approach to lexical creativity in models of possible words. In order to explore lexical creativity from an alternative, usage-based, framework, I will examine in Section 2 the different kind of paradigmatic relations among words sharing morphemes that give structure to the lexicon. In Section 3, the relationship between lexical creativity and the morphological structure of the lexicon are introduced, and the role of

word families and low-level patterns is also stressed. I conclude by pointing out that low-level patterns used in lexical creativity represent an evidence for the absence of a well-defined line between analogy and derivation.

## 1. Lexical creativity in models of possible words

Over the past 25 years, research on lexical creativity has been guided by the belief that explaining the speakers' ability to use new words must start from a theory of possible words. According to this widely accepted theoretical assumption, basing morphology solely on possible words "frees up the investigator to study principles that govern the ability to create and understand new words, rather than stick to superficial differences in the material nature of the data investigated" (Toman 1998:308). From this standpoint, focusing on possible words entails two main advantages: on the one hand, lexical creativity can be accounted for —by putting forward certain *mechanisms* that explain possible words— and on the other hand, the burden of having to describe actual words is removed.

These supposedly advantageous points are, in my view, a stumbling block to understanding lexical creativity. Let us analyze why. As for the ability of models of possible words to account for the ability to create and to understand new words, it is worth mentioning that once the model incorporates word-formation mechanisms supposed to licence possible words, it neglects the usage of such mechanisms in the actual coining of words (cf. Vallès 2000). For this reason, neology is not really a grammatical phenomenon in generative models of possible words. Di Sciullo and Williams (1987:7), for example, argue that "most views […] attach grammatical significance to the use of a new word — when a speaker makes a new word, he changes his grammar by adding the word to his lexicon […]. But in our view the listedness of a regular form is of no grammatical significance."

To be sure, a model of possible words will never consider neologisms as relevant data for, if they are regularly formed words (for example, *viceministre* 'vice-minister', *vicecomandant* 'vice-commander'), they are by definition contemplated by grammatical rules: they are possible words. However, if neologisms are somewhat irregular, they may be considered as idiosyncratic forms or analogical creations. For instance, no rule is able to neatly account for a neologism such as *vicelíder* 'vice-leader', which, as opposed to other words with *vice-*, does not designate a hierarcical position in an institution or business. From the perspective of a theory of possible words, such lexical items are but a

*question of performance* whose study escapes the methodological constraints of a competence model. Bauer (1983: 84) points out performance may become a rag-bag category, used to explain all variation that is not neatly captured by the theory of a particular analyst.

Another reason why in a model of possible words neology is an *extra-grammatical* phenomenon has to do with the fact that the very concept of neologism can only be defined as opposed to the lexicon used by a community of speakers in a previous period of time. Thus, we can claim that *ecotaxa* 'ecotax' is a neologism from the 1990s in Catalan, since it is new with regard to the Catalan lexicon, a fact that can be corroborated by consulting an older Catalan corpus, such as the IEC (Institute for Catalan Studies) Corpus de Català Contemporani (Contemporaneous Catalan Corpus). That is to say, a neologism is an actual word that is new when compared to the lexicon previously used in a linguistic community.[1] This is why neologisms are naturally accounted in models of actual words, but have no status in models of possible words.

The paradoxical ruling out of neology in a theoretical framework aiming to account for lexical creativity had openly been accepted by Rey (1976: 8), for whom "the transformation of that which is virtual (morphology) into something actual (lexicon; neologism) is a psycho-sociological issue, not a 'merely' linguistic one." It seems to me that, in order to provide an explanation for the formation of new words, we must overcome this narrow perspective on what is 'linguistic' and start from a theoretical framework that will allow as to articulate (instead of isolating) competence and performance: a model based on usage. The main reason for this is that neologisms are in fact a competence phenomenon and, at the same time, a performance phenomenon: the capacity to form neologisms is feature of the linguistic system, while neologisms created by the speakers are a result of the use of the language. It is necessary to integrate both dimensions in the grammatical model, though.

The second supposedly advantageous consequence of focusing on possible words is that the researcher's burden of describing actual words is removed. A methodological objection to this perspective is that a competence theory on word formation without a complementary theory on performance runs the risk of not being susceptible to empirical demonstration (Rainer 1987: 194). What is more, from a methodological point of view, it seems clear that, in order to provide an explanation for lexical creativity, we must use a model that incorporates performance —and, therefore, actual words— as grammatically significant data.

Still another assumption that characterizes the possible-word approach to the lexical creativity is the lack of a relationship between the word formation

patterns and the organization of the lexicon: "the lexicon has no structure that corresponds to the structure of words assigned by word formation rules" (Di Sciullo and Williams 1987: 15). In spite of this, there is an increased parallel to the existing relation between lexical creativity and the paradigmatic relations that structure the lexicon. As Bauer (2001) puts it, in recent years, a number of linguists have drawn attention to the paradigmatic forces which affect the coining of new words (cf. van Marle 1985, 1994; Bauer 1997; Becker 1993; Booij 1997). From perspective adopted herein, the word formation patterns emerge from paradigmatic relations and the function of rule-learning might be to help organize the lexicon, to give it structure; its role is to express generalizations about what is part of the lexicon, not to simplify it by removing redundant information (Derwing 1990: 251).

## 2. Morphological relations in models of actual words

In this section I will explore paradigmatic relations among words sharing productive morphemes. My aim is to develop Bybee's (2001: 109) hypothesis according to which "any multimorphemic word or sequence is highly embedded in connections with other words containing at least one of the same morphemes." From a morphological standpoint, the lexicon triggers a network based on similarity relations among words sharing morphemes: be they affixes or stems. Hence, this structure is articulated over two main axes: relations among words with common affixes (*derivational categories*) and relations among words sharing the same stem (*word families*).[2]

The diagram in Figure 1 represents those two axes of relations occurring simultaneously among words such as *setmana* 'week', *setmanal* 'weekly' (adj), *setmanalment* 'weekly' (adv); *seqüència* 'sequence', *seqüencial* 'sequential', *seqüencialment* 'sequentially'; *mort* 'dead', *mortal* 'deadly' (adj), *mortalment* 'deadly' (adv), etc.

In order for speakers to establish these kinds of morphological relations, morphemes must be transparent from a morphosemantic standpoint; that is to say, a *recurrent association between some semantic content and a phonological form* must be observable. Morphosemantic transparency is a requirement in order to acknowledge this paradigmatic relations (whose productive use eases their way to entrenchment in grammar, thus prompting future eventual activations of new formations). All of this leads us to deduce the existence of etymologically justifiable relations that do not comply with this *sine qua non*

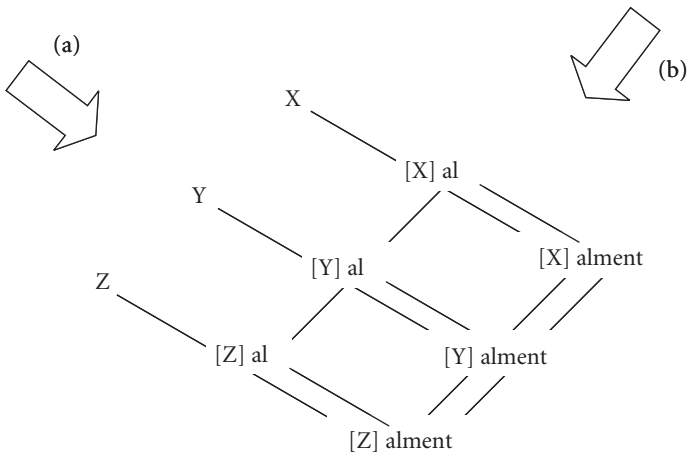


**Figure 1**. The two axes of the morphological structure of the lexicon.

requirement. This is the case with the Catalan *-ei* suffix (*homei* 'homicide', *servei* 'service') or even *-oll* (*grumoll* 'clot, lump', *manoll* 'tuft, bunch').

The relation among words in the same lexical family can also be equally opaque, such as *admetre* 'to admit', *cometre* 'to commit', *emetre* 'to emit', etc. and *advertir* 'to advise', 'to warn', *convertir* 'to convert', *divertir* 'to amuse', etc. (cf. Fig. 2). Although these words —most of which are learned words loaned from Latin— are etymologically interrelated, the relationship is opaque for many speakers, because it is hard for them to identify the base form with a previously known association of some semantic content and a phonological form. I have drawn those etymologically opaque relations in a dotted line. (For the sake of clarity, I have not made the relations of the *-ir* verbal ending explicit in this figure).

We should ponder the fact that words with opaque morphemes could be related from a phonological perspective. According to the phonological relations put forward by Bybee (1985, 1988, 1996a, 2001), words such as *reduir* 'to reduce' and *deduir* 'to deduce' would be related by all phonemes but the initial one. Nevertheless, unlike morphological relations, these phonological relations are not usually involved in word formation. In general terms, constituents that synchronically are only related from a phonological standpoint are unproductive: words with *-vertir,* for instance, are members of a *dead* word family. The paradigmatic relations used productively by derivation in lexical creativity are morphological relations. For this reason, I will adapt Bybee's usual representations of lexical relations in the following section.

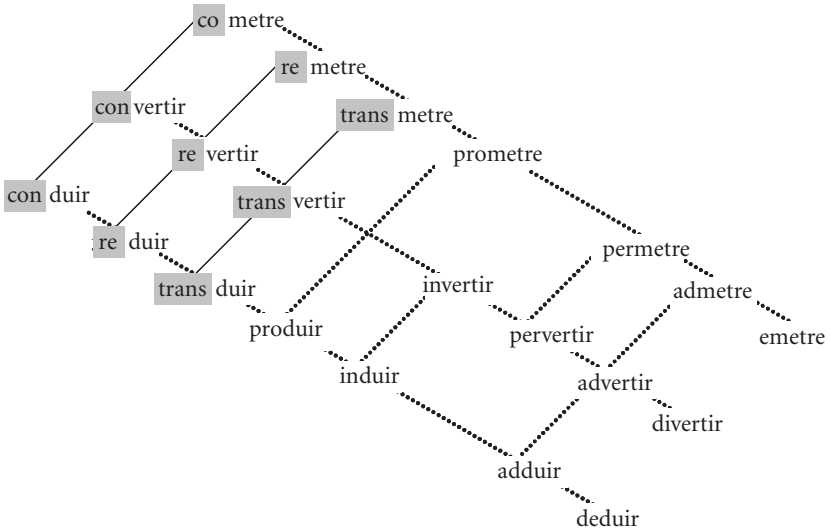


**Figure 2.** Dead word families.

## 2.1 Lexical relations and morphological boundaries

The emphasis on paradigmatic relations among words with a transparent morphological structure will lead us to explicitly express the morphological structure of these words. Here I deviate from Bybee's usual representations, insofar as the emphasis will no longer fall on the phonological relations but on the morphological relations. In Bybee's networks, limits between morphemes are not expressed explicitly; they are rather implicit —inferable by comparing the base form and the derivative, for instance— in the representation of a set of relations. In Figure 3, I compare Bybee's connections —which explicitly express phoneme-to-phoneme relations[3]— and the ones I will use, which are morpheme-to-morpheme relations.

The absence of explicit limits between morphemes in Bybee's representations is due to her belief of this segmentation to be unnecessary and, besides, problematic. The main reason why Bybee (1985: 127–128; 1996b: 63–65) rules out segmenting words in morphemes is the existence of problematic cases like the following:

i. Morphemes merged to the extent that they could only be separated arbitrarily. For instance, the stem and the morpheme in Spanish irregular past tenses *hube* 'I had (aux.), *tuve* 'I had' (poss.), *puse* 'I put'…

ii. Vowels or consonants that could be part of both the stem and the affix. They are ambiguous segments appearing between two morphemes, but they are not necessarily part of any such morpheme. For instance, the velar /g/ appearing in Spanish in some verbal tenses: *tenga* 'have' (sg. subj.), *ponga* 'put' (sg. subj.), *salga* 'leave' (sg. subj.)…
iii. Words with a part that is clearly a morpheme but with another part that has no straightforward meaning, that is, a morpheme apparently void of meaning (*cranberry morph*). For example, the diminutive suffix *-illo (a)* in words such as *semilla* 'seed', *anillo* 'ring'.

Some of these problems —such as the existence of fuzzy boundaries and elements of difficult categorization— reveal the shortcomings of the classical categorization theory. In other cases, the problem lies in the representation *on paper* of the morphological analysis of words with discontinuous morphemes or overlapped elements. It is worth mentioning that none of these problematic cases is an argument questioning the *mental* representation of the morphological structure of partially or totally transparent words.

We should also bear in mind that, in a cognitive model, morphematic segmentation of complex words in the lexicon does not have to be black or white —either a comprehensive segmentation of all complex words, or a total absence of limits between morphemes— but rather shades of gray. Speakers, for instance, may notice the recurrence of the initial element *di-* in *dilluns* 'Monday', *dimarts* 'Tuesday', *dimecres* 'Wednesday', etc. and associate *di-* with *dia* 'day', whereas the remaining part of these words is still morphologically opaque.

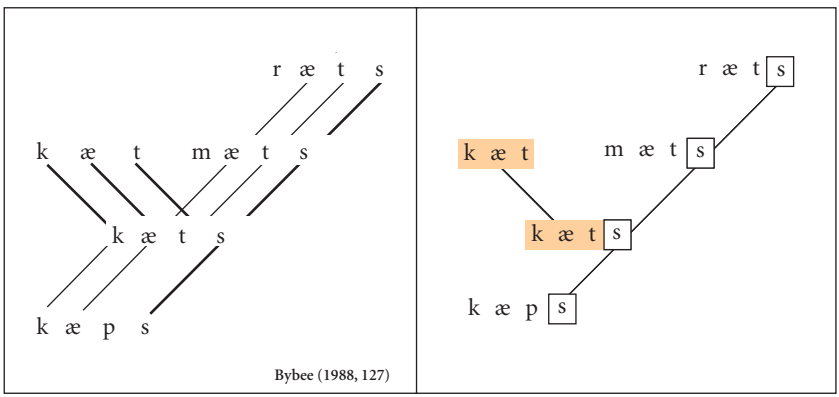


**Figure 3.** Comparison of phoneme-to-phoneme and morpheme-to-morpheme connections.

Segmentation can be partial, limited to recurrent and transparent morphemes, and among these especially to those productive. That is why words such as *anillo* and *semilla* —see above— , would not have an internal morphological structure (with the exception of inflection).

Regarding representation *on paper,* we could find a way to make the morphological structure of words with transparent and productive morphemes surface and at the same time account for problematic morphemes. For instance, the thematic vowel preceding the suffix *-ble* in *culpable* 'guilty' or *acceptable* 'acceptable' is a vowel between two morphemes which can be categorized in different ways —in a usage-based grammar there is no need for one single analysis. It could even be member of both morphemes *at the same time*, with no need to explicitly define an absolute limit (see Fig. 4).

I have pointed out that the other reason why Bybee (1985: 127–128) avoids explicit morphological segmentation is her considering it as unnecessary. However, morphological segmentation could be significant on other grounds: it allows for expressing the symbolic *unit* that constitutes each morpheme (a specific meaning associated to a particular phonologic form). Moreover the prominence of the internal structure of words is a relevant feature affecting the productivity of word formation processes (Ravid 1990: 328). Furthermore, from the perspective of lexical creativity, morphological relations among words in the lexicon are doubtlessly more salient than those solely phonological.[4]

The main reason why I emphasize morphological relations and I put forward an explicit morphological segmentation of complex words with transparent and productive morphemes is that lexical connections that speakers use to form neologisms are precisely relations among words with such morphemes. The productive use of morphemes forming the words in the lexicon could mirror the internal structure of such words and also the structure of the

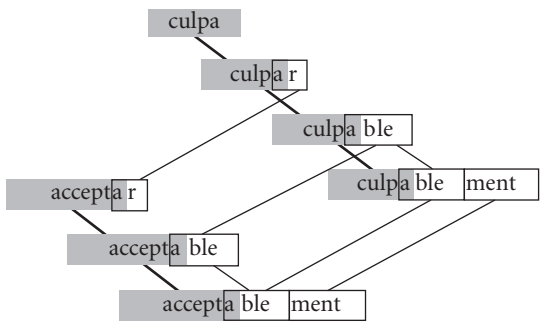


**Figure 4.** Example of overlapped elements in internal word structure: A thematic vowel.

very lexicon.[5] According to this, the morphological organization of the lexicon —the paradigmatic relations among words with common morphemes— results from the projection of the internal morphological structure of complex words.

## 3. Lexical creativity and morphological relations

In this section I explore the active role of morphological relations in lexical creativity. Firstly, I consider the paradigmatic relations between actual words sharing affixes and/or stems and, secondly, I analyze the way lexical creativity reveals this morphological structure of the lexicon. Paradigmatic relations[6] are thus key to the understanding of the dynamic aspects of the lexicon.

### 3.1 Lexical creativity and derivational categories

A prevalent case of derivational category is the paradigmatic relation among derivatives formed by means of a given affix. What does this paradigmatic relation have to do with lexical creativity? By observing the common prefix in words such as *superheroi* 'superhero' and *superministre* 'superminister', speakers manage to abstract the pattern and use it productively to coin the neologism *supertaxa* 'supertax'. In general terms, in order to codify and decode neologisms, it is necessary to compare similar conventional words and to observe the same morphemes in advance, in absence of which neologisms would be impossible to decipher. The relation among words with a common affix is expressed by means of morphological rules or *patterns emerging from the intrinsic organization of the lexicon* (Bybee 1988: 125).

Less studied is the paradigmatic relation between words that share two or more affixes. For example, as we will see below, there are reasons to think that a word with *-ització* like *calendarització* 'calendarization' (in wich the affixes *-itza* i *-ció* maintain a sintagmatic relation) is paradigmatically related to others words in *-ització*. The extreme case of such relation is an affix appearing systematically next to another. For instance, it is quite commonplace in Catalan that, when the base of adverbs formed with suffix *-ment* is a double-ending qualifying adjective (*dolç, dolça* 'sweet', the adverb is always formed on the feminine flectional form of the adjective. It is a systematical relation between the derivational category of derivative adverbs ended in *-ment* (*dolçament* 'sweetly', *eternament* 'eternally', etc.) and the derivational category of the feminine form of double-ending adjectives (*dolça* 'sweet', *eterna* 'eternal', etc.).[7]

In a model of actual words, this relation between derivational categories can be expressed by means of low-level patterns (Langacker 1987, 1991), which represent subregularities and are characterized by showing a high degree of specificity. As shown by Figure 5, the pattern for adverbs derived from adjectives with the suffix *-ment* has two low-level patterns, one for each morphologically conditioned base type: the feminine form of double-ending adjectives and the single form of invariable adjectives.

Another example of low-level pattern representing this kind of relations is the one specifying that the adjectival suffix *-ble* always appears after the basal thematic vowel: usually *-a* in first conjugation verbs (*agradar* 'to like', *agradable* 'pleasant'; *recomanar* 'to recommend', *recomanable* 'recommendable').[8] When two derivational categories always occur interrelated, affixes may even merge into one: *-ist* + *-ic*: *-istic* (*turístic* 'touristy', *periodístic* 'journalistic') or *-er* + *-i(a): -eria* (*adrogueria* 'cleaning materials store', *rellotgeria* 'watchmaker's').

Oftentimes relations between derivational categories reveal themselves in the tendency for two affixes to co-occur, that is to say, in the *preference* of an affix for bases formed with another affix. Such categories may correspond to two suffixes (*-itza* and *-ció*: *balearització* 'Balearization' = 'process of adopting/instilling Balearic features', *calendarització* 'calendarization' = 'adoption of a shedule'). The existence of low-level patterns with two affixes provides for and

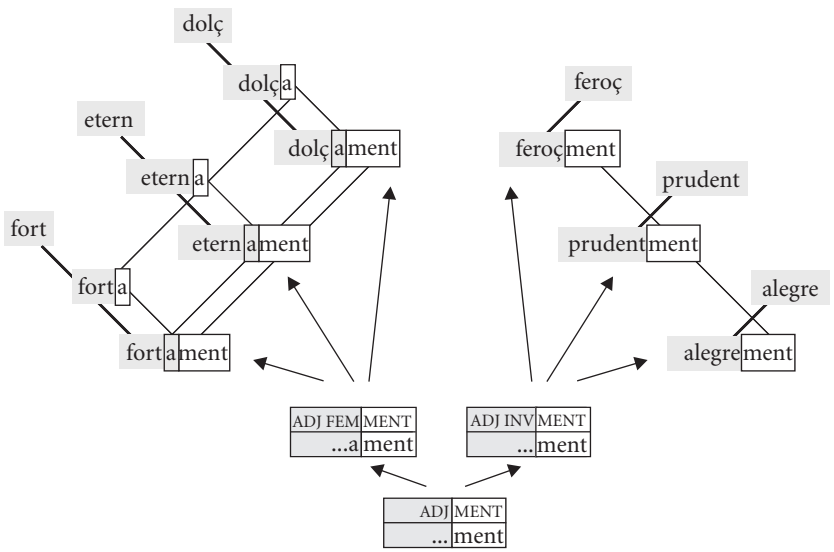


**Figure 5.** Transcategorial relation: *-ment* adverbs and adjectives in their feminine form.

accounts for neologisms with *-ització*, for example, even though the corresponding *-itza* verb may not be an established word (*precarització* 'precarization' = 'forced precariousness', *electoralització* 'electoralization'…). According to the data from the Observatory of Neology, the 57% of neologisms by suffixation formed on a non attested base are deverbal nouns formed with *-ització* (Freixa, Solé and Cabré 1998: 10). In English, the most often quoted paradigmatically related suffixes are *-ive* and *-ness* (*-iveness: effectiveness, expressiveness*), and *-ible* and *-ity* (*-ibility: separability, probability*) (cf. Bauer 1997: 248; Anshen and Aronoff 1988; for further examples, see Aronoff and Anshen 1998: 243).

On some occasions the derivational categories related do not correspond to two suffixes but to one prefix and one suffix. A well-known case is that of the *un-* prefix and the English past participle suffix (*-ed*). According to Zimmer's data (1964), the *un…ed* subscheme accounts for 50% of adjectives prefixed by *un-* (*unused, untrained, unsuspected…*). The interrelation between two derivational categories may therefore result in discontinuous low-level patterns. In Catalan, *re…ció, des…ció, anti…ista* and *inter…al* are some examples of this type of subschemes.

(1) The prefix *re-* and nouns ending in the suffix *-ció: reeducació* 'retraining', *repoblació* 'repopulation', *reindustrialització* 'reindustrialization', *reevangelització* 'reevangelization', *reinvenció* 'reinvention', *reaparició* 'reappearence', *reintroducció* 'reintroduction', *requalificació* 'requalification', etc.

(2) The prefix *des-* and nouns ended in the suffix *-ció*: *desinformació* 'disinformation', *descoordinació* 'incoordination', *desestabilització* 'destabilitization', *desestructuració* 'destructuring', *desmilitarització* 'demilitarization', *desmotivació* 'demotivation', etc.

(3) The prefix *anti-* and nouns or adjectives ended in the suffix *-ista*: *antiavortista* 'anti-abortionist', *antiinflacionista* 'anti-inflationist', *antifranquista* 'anti-Francoist', *antiracista* 'anti-racist', *antisandinista* 'anti-Sandinista', *antimilitarista* 'anti-militarist', etc.

(4) The prefix *inter-* and adjectives ended in the suffix *-al*: *interanual* 'interannual', *interconfederal* 'interconfederal', *interestatal* 'interstate', *interdepartamental* 'interdepartmental', *interpersonal* 'interpersonal', *interregional* 'interregional', etc.

A different type of discontinuous morphological pattern can be found in modern Hebrew, which uses nonlinear affixation as its most important word-

formation process (Ravid 1990: 293). In this Semitic language, roots are exclusively composed of three or four consonants and never form words by themselves. In order to form new words, roots combine themselves with affixation patterns, which contribute with vowels and accentuation. For instance, the root *z-m-r* 'to sing' unifies with CéCeC and tiCCóCet patterns to form *zémer* 'song' and *tizmóret* 'orchestra' respectively (Ravid 1990: 293).

So far the examples of relations between derivational categories could be expressed by means of subschemes of patterns allowing for the addition of affixal elements. There are, however, other cases where the relation between categories becomes apparent not in the addition but in the replacement of one element by another. Examples of this morphological strategy are the suffixes *-isme* and *-ista* (*racisme* 'racism', *racista* 'racist'; *franquisme* 'Franco era and government', *franquista* 'related to or pro-Franco'), *-òleg* and *-ologia* (*microbiòleg* 'microbiologist', *microbiologia* 'microbiology'),[9] combining forms *-cidi* and *-cida* (*etnocidi* 'ethnocide', *etnocida* 'person who commits an ethnocide'), *-crata* and *-cràcia* (*gerontòcrata* 'gerontocrat', *gerontocràcia* 'gerontocracy') or *macro-* and *micro-*.

This last example is of special interest, because there are reasons to believe that the productivity of *micro-* has crucially determined the productivity of its antonym *macro-* and that at the present time this influence could work both ways. The first *macro-* words attested in the IEC corpus (*macroscòpic* and *macrosisme, macrosísmic*) may have been formed by analogy to the equivalent *micro-* words, documented prior to them (*microscòpic*) or contemporaneously (*microsisme, microsísmic*). Moreover, 52% of *macro-* complex words in this corpus have a *micro-* equivalent (*microcosmos* v. *macrocosmos*) and more than 40% are found either contemporaneously or else subsequently to their *micro-* equivalent.

The phonological similarity between the two morphemes (differing only by the vowel in the first syllable) further bolsters the paradigmatic relation between the two antonymous combining forms. The power of this paradigmatic relation is a decisive factor when explaining their productivity, in the sense that the productivity of one form favors that of the other. Again, lexical creativity is thus associated with the power of paradigmatic relations among words in the lexicon.

In brief, I have drawn a distinction between three types of paradigmatic relations among actual words in the lexicon that have effects on lexical creativity:

i. The paradigmatic relation among words with a common affix, which can result in the formation of neologisms by attaching the affix at issue. It is therefore a relation among members of the same derivational category.

ii. The paradigmatic relation among words with at least two common affixes. The most frequent case of this relation between associated derivational categories is the attachment of either two suffixes or one prefix and one suffix. This relation can be expressed through low-level patterns.
iii. The paradigmatic relation among words with semantically related affixal elements, such as *-isme/-ista, macro-/micro-*…, where an affixal element of a complex word may be replaced by another.

Paradigmatic relations among complex words in the lexicon have thus repercussions on lexical creativity. Along these lines, I agree with van Marle (1985: 87) when taking the view that "investigation of the *processual aspect of word-structure* [that is, word-formation] can best be taken in hand within the framework of a theory in which the *entitative aspect of morphological structure* [the word as an entity related to other entities] is assigned a central position."

### 3.2 Lexical creativity and word families

The role word families might play in lexical creativity is a rarely explored domain. Yet the advantageousness of studying words within the family they belong to is an idea of old. Malkiel (1954), for example, suggests studying word etymology within the scope of lexical families. According to this author, the word family is to be considered "the most fitting unit in lexical research" (Malkiel 1954: 274). Many interesting problems —laments Motsch (1988: 25)— have no place in the framework of generative grammar, for example: are there paradigms in derivational morphology[10] and what is their influence on morphological processes?

The effects of the relation among word family members have recently been confirmed in experimental studies (Bertram *et al.* 2000, De Jong *et al.* 2000, Hay and Baayen 2002). Bertram *et al.* (2000) have verified the so-called *family size effect* in complex words, which means that words belonging to word families with numerous members activate more rapidly.[11] This facilitating effect is easily seen given the existence of a network of paradigmatic relations among morphologically transparent words formed on a common stem, where the activation of one node also activates nearby nodes. As such, the word-family effect can be viewed as a compelling piece of empirical evidence in favor of the existence of lexical interconnections based on morphologically transparent relations (Bertram *et al.* 2000: 402).

Thus, in the same way speakers observe several words sharing a given affix, there are reasons to believe they also recognize the similarity among words in

the same family. From this standpoint, the members in the following family: *edició* 'publishing', *editar* 'to publish', *reedició* 'reprint', *reeditar* 'to reprint', etc., are related by both their common affixes and their stem (see Figure 6).

It is worth stressing that the relation among words formed on the same stem *is of the same nature* that that of the relation among words formed with the same affix: they are both paradigmatic relations based on similarity when observing common morphemes. And yet, many morphological models contemplate only derivational categories. The reason for this seems to be that affixes, and not word families, are used productively. From this standpoint, the formation of a neologism like *autoeditor* 'self-publisher' is generally explained as a result of the high productivity of the prefix *auto-*. This productivity provides evidence that speakers have noticed the prefix *auto-* in different complex words (*autoengany* 'self-delusion', *autoconfiança* 'self-confidence', *autocontrol* 'self-control', *autocensura* 'self-censorship', etc.). The affix is therefore considered as the key to lexical creativity.

It is important to note, however, this neologism comprises two constituents —affix and base form— , so affix productivity occurs in parallel with word family productivity. From this view, the neologism *autoeditor* is at the same time a result of the productivity of the prefix *auto-* and that of the word family of the base form: *editor*. Figure 7 is a diagram of these relations among *autoeditor* (and other neologisms, highlighted in bold letters) and conventional words closer to the word family.

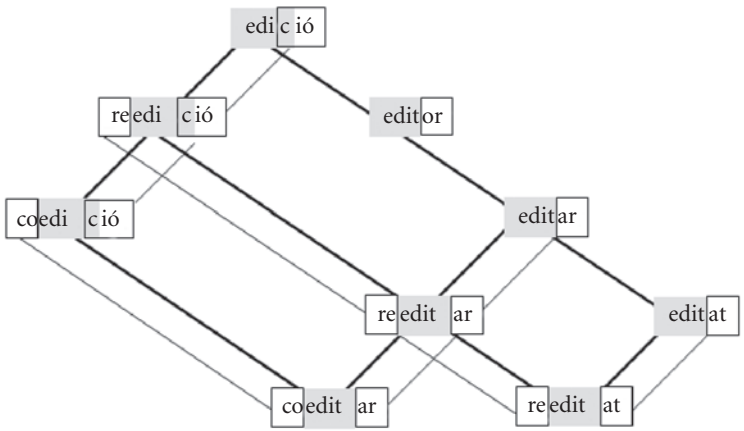


**Figure 6.** The *edició* word family (I)

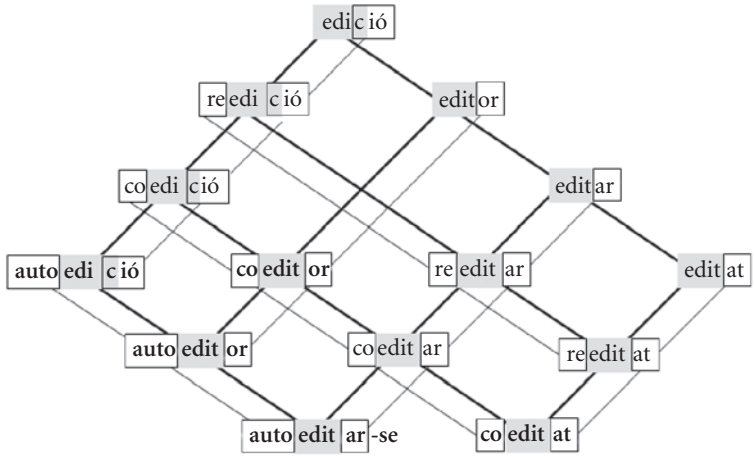


**Figure 7.** The *edició* word family (II)

The presence of conventional words formed by prefixation in this family (*reedició* 'reprint', *coedició* 'joint publication', etc.) certainly favors the incorporation of new prefixed words (*autoedició* 'self-print' = 'print of the publisher's own work', *autoeditor* 'self-publisher' = 'person who publishes his/her own work', etc.). What is more, the high productivity of the prefix *co-* should not be an obstacle to realizing that conventional words *coedició* 'joint publication' and *coeditar* 'to publish jointly' facilitate the formation of neologisms like *coeditor* 'joint publisher' = 'publisher in charge of a joint publication' and *coeditat* 'jointly published'.

The reason why it is sometimes difficult to see the active role that word-family relations play in lexical creativity has to do with the fact that their influence co-occurs with that of derivational categories. Still, this active role is sometimes more obvious. The influence of word family members reveals itself, for example, in analogical formations in which a conventional word is another word's *model* — e.g. in the so-called *back formation* or in affix replacement processes.

Consider a specific example such as this: the neologism *interdisciplina* 'interdiscipline'. Figure 8 shows a fragment of the word family where the word is included; dashed arrows mark the relations among the words likely to have been used for codifying and decoding the neologism.

The network of lexical relations shown here accounts for two possible processes in the coining of *interdisciplina*: (a) prefixation of *inter-* on *disciplina*, and (b) *back formation* from *interdisciplinari*, by suppressing the suffix *-ari*.

Given that *inter-* is not very productive with nouns (but with adjectives ending with the suffix *-al*), the back formation process could be more probable than the prefixation one.

We should bear in mind that the purpose of this grammatical model is not to spell out *the* rule that hypothetically *generates* the neologism, but rather to show the array of resources available to speakers when using it. In this case, there is more than one possible process. In a usage-based model, there are no reasons to believe that one analysis, categorization or explanation of a linguistic phenomenon must automatically exclude other possibilities (Langacker 1987: 28; Bybee 1999: 230).

As far as the neologisms *transdisciplinari* 'transdisciplinary' and *extradisciplinari* 'extradisciplinary' are concerned (cf. Fig. 9), they can at the same time be considered as: (a) complex words formed on *trans-* or *extra-* and *disciplinari,* and (b) analogical formations with regards to *interdisciplinari,* formed by replacing *inter-* with *trans-* or *extra-*. In this case, the close semantic relation among the neologisms and the analogy model supports the hypothesis of analogical formation. Similarly, there are two possible processes for *transdisciplinarietat* 'transdisciplinarity': (c) derivation from *transdisciplinari* and *-etat,* and (d) replacement of *inter-* with *trans-* starting from *interdisciplinarietat.* Insofar as the base for derivational process —*transdisciplinari*— is not a conventional word but another neologism, I think the replacement analogical process has the upper hand.

Beyond the units and the relations that make up a given family, what word families suggest is that neologisms are neither isolated units nor exclusively related to complex words in their same derivational category. On the contrary, neologisms join a network of paradigmatic relations based on common affixes and stems, and it is these relations that allow speakers to codify and decode new creations. Each lexical creation is therefore the result of a set of lexical relations.

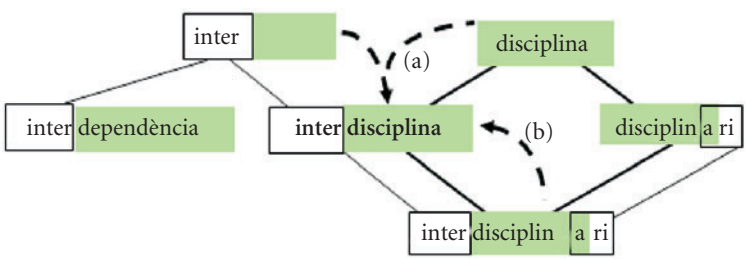


**Figure 8.** The *disciplina* word family (I)

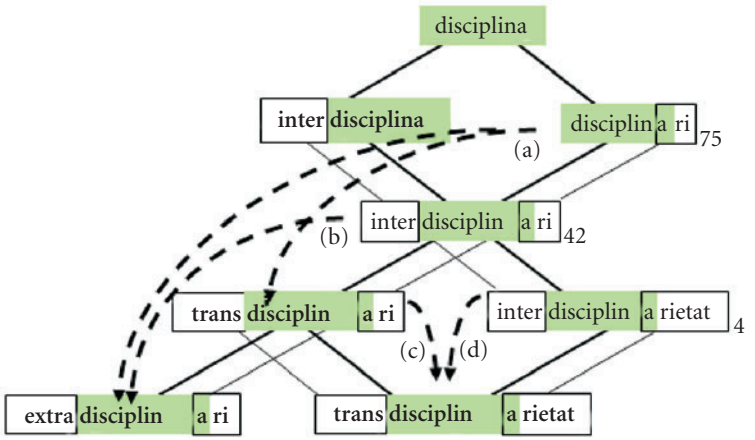


**Figure 9**. The *disciplina* word family (II)[12]

### 3.3 Lexical creativity as a manifestation of the morphological structure of the lexicon

The challenge for a model of actual words is to account for lexical creativity, demonstrating that —despite Beard's view (1998: 46)— it is not simply a set of static paradigmatic relations. We should therefore wonder: is a model of actual words just a hypothesis about the already consolidated lexicon in a language or is it rather able to account for lexical creativity? In other words, is a model of actual words just a static representation of the lexicon or can this static representation account for the dynamic dimension of the lexicon?

The analysis of neologisms is of interest for linguistics, because it gives us clues about how the dynamic and static aspects of the lexicon are related.[13] Specifically, neologisms are a fertile source of evidence for the existence of productive paradigmatic relations that hold between words sharing common morphemes. The key to explaining the interrelation between the static and the dynamic aspect of the lexicon lies precisely in the productive paradigmatic relations among words in the lexicon. Analyzing these relations from a static standpoint is studying the organization of the lexicon. Analyzing them from a dynamic perspective entails going further into the understanding of lexical creativity. The organization of the lexicon and lexical creativity are thus "mutually enlightening" realities.

On the one hand, the study of lexical creativity contributes with relevant data for the understanding of the organization of words. The use of productive resources of language mirrors the internal structure of the lexicon, in the sense

that it reveals the paradigmatic relations that are productively used by speakers. Out of all the family resemblance relations that could be established among lexical items from a morphological perspective, neologisms reveal those that have been noticed and used productively by speakers, i.e. the most salient ones. Moreover, the use of neologisms favors speaker's observation of similarities among complex words with similar morphemes, thus also facilitating recognition of the morphological structure of the words in the lexicon. Neology, therefore, helps maintain and at the same time reveals the prominence of the internal structure of complex words and paradigmatic relations among these words.

On the other hand, the study of the morphological organization of the lexicon —*viz.*, paradigmatic relations— constitutes an enriching approach to lexical creativity. In this paper, I have reformulated a hypothesis put forward by van Marle (1985). This author's core hypothesis is that morphological creativity can only be properly addressed within the framework of a carefully worked theory of the paradigmatic dimension of the morphological structure of the language (cf. van Marle 1985: 23). Though I have deviated quite far from both the theoretical framework and the morphological structure proposed by van Marle,[14] I have incorporated to cognitive morphology his hypothesis about the role of paradigmatic relations in lexical creativity.

The present analysis suggests that lexical creativity both pre-supposes and reveals the morphological structure of the lexicon. What is more, it is the manifestation of the intrinsic organization of the lexicon in paradigmatic relations among actual words sharing morphemes.[15] Once we have overcome the architectual dichotomy in which the lexicon is depicted as a "ragbag" of irregularities and rules are meant to economize storage through simplification, the organization of the lexicon from a morphological standpoint takes a central position. The very definition of grammar as a *structured* inventory of conventional linguistic units (Langacker 1987, 1991) and more specifically the conception of morphology as a *lexical organization* (Bybee 1988) are symptomatic of the increasing relevance of the morphological structure of the lexicon in linguistic theory. Along these lines, the organization of the lexicon may also turn out to be relevant for its contribution to the understanding of lexical creativity.

## 4. Final thoughts

Models of possible words treat new word formation as an *extragrammatical* phenomenon. Here I address lexical creativity within the framework of a model

of actual words. Specifically, Bybee's (1985, 1988, 1995, 2001) network model, that I have applied on the field of word formation by derivation and analogy. As we have seen, the morphological organization of the lexicon is a structure made up of two axes of paradigmatic relations: word families and derivational categories. These relations are based on the morphological structure of complex words with transparent and productive morphemes.

In word families and in derivational categories, two types of word formation processes can be seen: affix attachment and replacement. Traditionally, affix attachment is considered to be a rule-based derivational process, and replacement is supposed to be analogical. However, as Motsch (1988: 25) puts it, the clear-cut distinction between rule and analogy is due to methodological prerequisites of certain theoretical models. Derivation and analogy are both the result of paradigmatic relations —which speakers have noticed and used productively— among words with common morphemes (*-ment* '*-ly*', *-ització* '*-ization*'…) and/or with semantically associated morphemes (*-isme* '-ism'/-ista '*-ist*', *macro-/micro-, etc.*).

From the perspective of a model of actual words, there is no absolute boundary between derivation (by means of an affix) and word formation by analogy (where a word or a group of words are used as a model to form other words). The difference between both types of processes is a matter of degree. The typically derivative patterns have a high type frequency, they are not too specific and may have a high degree of entrenchment. On the contrary, analogical patterns are basically characterized for having a lower type frequency, so they are little or no fixed in the grammar, and they are very specific: they represent a great number of common traits for a reduced number of words.

In this context, low-level patterns that express the paradigmatic relation between words with more than a common affix (*-ització, re…ció, anti…ista*) are an argument in favor of a fuzzy boundary between derivation and analogy. These subpatterns are found half way between one and the other type of process since, in spite of representing derivative affixes, they share similarities with analogical patterns. They are more specific than the corresponding general pattern (*-ció* '-tion', *re-*, *anti-*…), and they have a lower type frequency and represent a great number of common traits within a group of words. For this reason, as Langacker (1999: 145) points out, words formed by means of low-level patterns could easily been considered as analogical. In conclusion, low-level patterns used in neology provide evidence for the absence of a well-defined line between derivation and analogy, two types of traditionally opposed word formation processes.

## Notes

* This paper was given at the III Congress of the Spanish Cognitive Linguistics Association, Universitat de València, Valencia, 15–17 May 2002. I would like to thank M. Teresa Cabré, Dirk Geeraerts, Joe Hilferty, M. Josep Jarque and an anonymous reviewer for his valuable comments on the written version. This research has been partially supported (FIAP 95/4709).

**1.** Bearing this in mind, the words contained in a corpus including media texts from the 1990s either absent or observed only once in the IEC *Diccionari de freqüències* corpus have been considered to be neological —nonconventional. The IEC dictionary of frequencies is based on a 50 million word corpus including texts ranging from 1833 to 1988. See the Introduction to the *Diccionari de freqüències* (Rafel i Fontanals 1996).

**2.** As we will see later, there are also paradigmatic relations among semantically associated elements (*-isme/-ista, -òleg/-ologia, macro-/micro-*…). About paradigmatic semantic relations (hyponymy: 'type of', meronymy: 'part of', metaphor, specialization, etc.) studied according to prototype theory, see for example Cruse (1994) and Geeraerts (1995, 1997, 2002).

**3.** Bybee tends to highlight in bold phonological connections that are also morphological, a very straightforward standard especially for single-phoneme morphemes (such as the plural morpheme *-s* in Fig. 3).

**4.** And yet, unproductive stems and affixal elements may be used in neologisms formed by analogy. For example, the *-gènia* morpheme appears in the neologism *telegènia,* formed by analogy to *fotogènia* (Vallès 2000).

**5.** This argument runs parallel to Cutler's (1980: 50), according to whom the preference for word boundary derivations may actually reflect the structure of the internal lexicon itself.

**6.** Following the Saussurean view, syntagmatic relations are associations among linguistic expressions existing *in praesentia*, whereas paradigmatic relations refer to associations *in absentia.* For instance, the relation among the constituents of a complex word is syntagmatic (*eco-* and *turisme* in the complex word *ecoturisme*), as opposed to the relation between base and derivative (*turisme* and *ecoturisme*), as well as the relation among several derivatives (*ecoturisme, ecoauditoria…*), which are paradigmatic.

**7.** The flectional morpheme *-a* as in *dolçament* cannot be interpreted as an epenthetic vowel, because, in invariable adjectives with a similar phonological context (*feroç* 'ferocious') the *-ment* suffix is directly attached to the stem (*feroçment* 'ferociously'). For a complete overview on word formation in Catalan, see Mascaró (1986), Cabré (1994, 2002) and Solà *et al.* (2002: 731–932).

**8.** Sometimes the thematic vowel is an -i- (*obtenir* 'to obtain', *obtenible* 'obtainable'*; desmentir* 'to deny', *desmentible* 'deniable'). The subscheme *-ible* is not as productive as *-able* (*edificable* 'buildable', *embargable* 'impoundable', *filmable*).

**9.** Notice that a *microbiòleg* '*microbiologist*' is not a 'small biologist', but an expert in *microbiologia* '*microbiology*'. Therefore, the name of the expert (shorter in Catalan) is derived from that of the domain (longer) (Rainer 1993: 81; quoted by Iacobini 1996: 231). See in Iacobini (1996: 231–232) the different explanations of such phenomenon that have been put forward in generative morphology.

**10.** See Bauer (1997) on this. This author concludes that, even though derivational paradigms are not prototypical, derivation is similar enough to inflection as to be able to apply the notion of paradigm to derivation.

**11.** Schreuder and Baayen (1997) had previously noticed this effect in simple word recognition: words in large families were recognized faster and more accurately. See also Bradley (1980).

**12.** As an indicator to its usage frequency, this figure shows the number of occurrences in the *Diccionari de freqüències* for the conventional words.

**13.** See Lehrer (1996) for other reasons why neologisms are important to study.

**14.** Van Marle (1985) focuses on, on the one side, the relation between the base and the derivative and, on the other side, the relation among affixes with a base of the same grammatical category —that is to say, among deadjectival, denominal and deverbal suffixes. The author does not provide for a set of paradigmatic relations among words in the same family or the interrelation among derivational categories. Van Marle (1994) does not address these relations either.

**15.** This approach to lexical creativity from the morphological point of view must be complemented with other analitical perspectives, such as semasiology and pragmatics. It would also be interesting to take a thorough look at the impact of exclusively semantic paradigmatic relations on lexical creativity.

## References

Anshen, F. & Aronoff, M. (1988). Producing morphologically complex words. *Linguistics,* 26, 641–655.

Aronoff, M. & Anshen, F. (1998). Morphology and the lexicon: Lexicalization and productivity. In A. Spencer & A.M. Zwick (Eds.), *The Handbook of Morphology* (pp. 237–247). Oxford: Blackwell.

Bauer, L. (1983). *English Word-Formation.* Cambridge: Cambridge University Press, 1993.

Bauer, L. (1997). Derivational paradigms. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology, 1996* (pp. 243–256). Dordrecht: Kluwer.

Bauer, L. (2001). *Morphological Productivity.* Cambridge: Cambridge University Press.

Beard, R. (1998). Derivation. In A. Spencer & A.M. Zwick (Eds.), *The Handbook of Morphology* (pp. 44–65). Oxford: Blackwell.

Becker, T. (1993). Back-formation, cross-formation, and 'bracketing paradoxes' in paradigmatic morphology. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology, 1993* (pp. 1–25). Dordrecht: Kluwer.

Bertram, R.; Baayen, H. & Schreuder, R. (2000). Effects of family size for complex words. *Journal of Memory and Language,* 42, 390–405.

Booij, G. (1997). Autonomous morphology and paradigmatic relations. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology, 1996* (pp. 35–53). Dordrecht: Kluwer.

Bradley, D. (1980). Lexical representation of derivational relation. In M. Aronoff & M.L. Kean (Eds.), *Juncture* (pp. 37–55). Saratoga (Calif.): Anma Libri.

Bybee, J.L. (1985). *Morphology. A study of the relation between meaning and form.* Amsterdam: John Benjamins.

Bybee, J.L. (1988). Morphology as lexical organization. In M. Hammond & M. Noonan (Eds.), *Theoretical Morphology. Approaches in modern linguistics* (pp.119–141). San Diego: Academic Press.

Bybee, J.L. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes,* 10(5), 425–455.

Bybee, J.L. (1996a). Productivity, regularity and fusion: How language use affects the lexicon. In R. Singh (Ed.), *Trubetzkoy's Orphan. Proceedings of the Montréal Roundtable Morphonology: Contemporary Responses* (pp.247–357). Amsterdam: John Benjamins.

Bybee, J.L. (1996b). Modelo de redes en morfología. In *Actas del XI congreso internacional de la Asociación Lingüística y Filología de la América Latina, Vol 1* (pp.59–74). Las Palmas de Gran Canaria: Universidad de Las Palmas de Gran Canaria.

Bybee, J.L. (1999). Usage based phonology. In M. Darnell; J.M.E. Moravcsik; F. Newmeyer; M. Noonan & K. Wheatley (Eds.), *Functionalism and Formalism in Linguistics, Vol. 1: General papers* (pp.211–242). Amsterdam: John Benjamins.

Bybee, J.L. (2001). *Phonology and Language Use.* Cambridge: Cambridge University Press.

Cabré, M.T. (1994). *A l'entorn de la paraula, Vol. 2: Lexicologia catalana.* València: Universitat de València.

Cabré, M.T. (2002). La derivació. In J. Solà; M.R. Lloret; J. Mascaró & M. Pérez Saldanya (Eds.), *Gramàtica del català contemporani, Vol. 1: Introducció. Fonètica i fonologia. Morfologia* (pp.731–775). Barcelona: Empúries.

Cruse, D.A. (1994). Prototype theory and lexical relations. *Rivista di Linguistica,* 6(2), 167–188.

Cutler, A. (1980). Productivity in word formation. *Papers from the Regional Meeting of the Chicago Linguistic Society,* 16, 45–51.

De Jong, N.H.; Schreuder, R. & Baayen, H. (2000). The morphological family size effect and morphology. *Language and Cognitive Processes,* 15, 329–365.

Derwing, B.L. (1990). Morphology and the mental lexicon: psycholinguistic evidence. In W.U. Dressler *et al.* (Eds.), *Contemporary Morphology* (pp.249–265). Berlin: Mouton de Gruyter.

Di Sciullo, A-M. & Williams, E. (1987). *On the Definition of Word.* 2nd ed. Cambridge (Mass.): MIT Press, 1988.

Freixa, J.; Solé, E. & Cabré, M.T. (Coords.). (1998). *Descripció quantitativa dels neologismes documentats durant l'any 1995 a la premsa en català* (Sèrie Informes, 23). Barcelona: Observatori de Neologia, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

Geeraerts, D. (1995). Especialització i reinterpretació en les expressions idiomàtiques. *Caplletra,* 18, 45–63.

Geeraerts, D. (1997). *Diachronic Prototype Semantics: A contribution to historical lexicology.* Oxford: Clarendon Press.

Geeraerts, D. (2002). The interaction of metaphor and metonymy in composite expressions. In R. Dirven & R. Prings (Eds.), *Metaphor and Metonymy in Comparison and Contrast* (pp.435–465). Berlin: Mouton de Gruyter.

Hay, J. & Baayen, H. (2002). Parsing and productivity. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology, 2001.* Dordrecht: Kluwer.

Iacobini, C. (1996). Il principio di direzionalità nella morfologia derivazionale. *Lingua e Stile,* 31(2), 215–237.

Langacker, R.W. (1987). *Foundations of Cognitive Grammar, Vol. 1: Theoretical prerequisites.* Stanford: Stanford University Press.

Langacker, R.W. (1988). A usage-based model. In R.W. Langacker, *Concept, Image, and Symbol: The cognitive basis of grammar* (pp. 261–288). Berlin: Mouton de Gruyter.

Langacker, R.W. (1991). *Foundations of Cognitive Grammar, Vol. 2: Descriptive application.* Stanford: Stanford University Press.

Langacker, R.W. (1999). A dynamic usage-based model. In R.W. Langacker, *Grammar and conceptualization* (pp. 91–145). Berlin: Mouton de Gruyter.

Lehrer, A. (1996). Why neologisms are important to study. *Lexicology,* 2(1), 63–73.

Malkiel, Y. (1954). Etymology and the structure of word families. *Word,* 10, 265–274.

Mascaró, J. (1986). *Morfologia.* Barcelona: Enciclopèdia Catalana.

Motsch, W. (1988). On inactivity, productivity and analogy in derivational processes. *Linguistische Studien,* 179, 1–30.

Rafel i Fontanals, J. (Dir.). (1996). *Diccionari de freqüències.* Barcelona: Institut d'Estudis Catalans.

Rainer, F. (1987). Produktivitätsbegriffe in der Wortbildungstheorie. In W. Dietrich; H.-M. Gauger & H. Geckeler (Eds.), *Grammatik und Wortbildung romanischer Sprachen* (pp. 187–202). Tübingen: Narr Verlag.

Rainer, F. (1993). *Spanische Wortbildungslehre.* Tübingen: Niemeyer.

Ravid, D. (1990). Internal structure constraits on new-word formation devices in modern Hebrew. *Folia Linguistica,* 24(3/4), 289–347.

Rey, A. (1976). Néologisme: un pseudo-concept? *Cahiers de Lexicologie,* 28(1), 3–17.

Solà, J.; Lloret, M.R.; Mascaró, J. & Pérez Saldanya, M. (Dirs.). (2002). *Gramàtica del català contemporani, Vol 1: Introducció. Fonètica i fonologia. Morfologia.* Barcelona: Empúries.

Schreuder, R. & Baayen, H. (1997). How complex simplex words can be. *Journal of Memory and Language,* 37, 118–138.

Toman, J. (1998). Word syntax. In A. Spencer & A.M. Zwick (Eds.), *The handbook of morphology* (pp. 306–321). Oxford: Blackwell.

Vallès, T. (2000). La innovació lèxica a la premsa: un repte per a la lingüística. In M.T. Cabré; J. Freixa & E. Solé (Eds.), *La neologia en el tombant de segle* (pp. 43–53). Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.

van Marle, J. (1985). *On the Paradigmatic Dimension of Morphological Creativity.* Dordrecht: Foris.

van Marle, J. (1994). Paradigms. In R.E. Asher (Ed.), *The Encyclopedia of Language and Linguistics, Vol. 6* (pp. 2927–2930). Oxford: Pergamon Press.

Zimmer, K.E. (1964). Affixal negation in English and other languages. *Word,* 20(2), 21–45.

## *Author's address*

tvalles@cir.unica.edu