

Lexicalization and Spanish derivational morphology

Carlos Benavides
University of Massachusetts Dartmouth / USA

Abstract – In this study lexicalization refers to derivation where an idiosyncratic component of meaning has been acquired. Being non-compositional, lexicalized items are usually considered irregular. In accordance with an emerging view that irregularity should take a place as one of the central issues in linguistic theory, this article deals with lexicalized derivatives in Spanish within the framework provided by the dual-route model. On the basis of intuitive speculation and an exploratory search of a Spanish corpus, the hypothesis was formulated that a significant majority of derivatives in Spanish are compositional; therefore, lexicalization is a secondary process in Spanish word formation. A corpus study comparing results from two large Spanish corpora was conducted to test the hypothesis. The results, based on an analysis of over 10,000 derivatives confirm the hypothesis, supporting the author's intuitions and providing additional support for the dual-route model. In addition, the corpus findings suggest that metaphor in Spanish derivation is not as common as may previously have been thought.

Keywords – lexicalization, derivation, irregularity, dual-route, corpus, metaphor

1. INTRODUCTION

The term 'lexicalization' has been defined in a variety of ways (see Bauer 1983, 2001; Lang 1990; Lipka 1990; Brinton and Traugott 2005; Alonso Calvo 2009). In the present study, lexicalization refers to derivation of words whose meaning is not transparent (see Aronoff and Anshen 1998), that is, the formation of items that have acquired an idiosyncratic component of meaning. Being non-compositional, lexicalized items are usually considered irregular, or at least semi-regular. In accordance with Jackendoff's (2010) view that irregularity (or semiproductivity) should take a place as one of the central issues in linguistic theory, this article explores lexicalized derivatives in Spanish within the framework provided by the dual-route model (or 'words and rules theory'; Pinker 1999, 2006; Pinker and Ullman 2002), employing two Spanish corpora. In addition to supporting a dual mechanism approach, the corpus findings suggest that metaphor in Spanish derivation is not as common as may previously have been thought.

Before moving on to a discussion of the dual-route model, it is important to mention two other salient senses of the term 'lexicalization'. One of them is the instantiation of a concept (or concepts) as a lexical item. For example, while the notion of 'finding pleasure in the suffering of others' is lexicalized in German as the noun *Schadenfreude*, this concept is not expressed in English as a single word (although *gloating* comes close), and therefore is not lexicalized in English. Conversely, while English lexicalizes the concept of 'walking' in the word *walk*, German does not have a specific single word for 'walk' and lexicalizes the concept in the word *gehen*. Talmy (1985) is an influential work that frames its analysis within this definition of lexicalization. Another sense, which has strong diachronic implications, is construed as the opposite of grammaticalization; that is, lexicalization is seen as the conversion of a bound or grammatical morpheme into a full lexical item (eg. using the suffix *-ism* as the noun *ism*, as in *Sociopolitical discussion is full of isms*) or the conversion of a syntactic structure into a lexical item, as when the sentence (Sp.) *No me olvides* 'do not forget me' is converted into *nomeolvides*, a type of flower or a type of bracelet (Moreno Cabrera 1998;

Lehmann 2002; Brinton and Traugott 2005; Blasco Mateo 2006; Buenafuentes De La Mata and Sánchez Lancis 2012). The present article is couched within the definition provided in the first paragraph of this section (non-transparent, idiosyncratic meaning) and approaches lexicalization from a synchronic point of view, without making a connection to grammaticalization. Thus, neither of these two latter definitions of lexicalization, nor others employed in the literature, is directly relevant to the present study.

Pinker and Ullman (2002) provide a clear picture of what the dual-route model entails. Employing various forms of evidence, they “defend the theory that irregular past-tense forms are stored in the lexicon, a division of declarative memory, whereas regular forms can be computed by a concatenation rule, which requires the procedural system. Irregulars have the psychological, linguistic and neuropsychological signatures of lexical memory, whereas regulars often have the signatures of grammatical processing”. Thus, according to the dual-route model, while regular inflected forms (eg. *walk+ed*) are computed by a concatenative rule, irregulars (eg. *sang*) must be stored in memory, as part of an associative network.¹ Pinker and Ullman (2002) base the model on the English past tense, but there is evidence that English regular and irregular words formed by derivation are accounted for by the dual-route model as well. For example, Alegre and Gordon (1999), through corpus and experimental studies, show that derivational morphology, much like inflectional morphology, manifests dissociations between rule-based and associative generalization mechanisms. They found that words formed with certain suffixes (*-ion, -al, -ity, -ous, -ic*) exhibit cluster (or gang, ie. associative) effects, just like irregular inflected words, while words formed with other suffixes (*-ize, -en, -ness, -able, -ment, -er*), much like regular inflection, do not display such effects. Furthermore, Vannest et al. (2005) found that decomposable (ie. regular) derived words in English (formed with the suffixes *-ness, -less, -able*) showed increases in activity in regions of interest (Broca’s area and the basal ganglia) relative to nondecomposable (ie. irregular) suffixed words (formed with *-ity, -ation*), suggesting that, in accordance with the dual-route model, while regular forms are accessed from the mental lexicon as separate morphemes (base and affix), irregulars are accessed as whole units.

2. LEXICALIZATION AND THE DUAL-ROUTE MODEL

Following Lipka (1990), lexicalization occurs when a lexical item (whether a base or a derivative) undergoes semantic drift, which causes the meaning of the whole to become unpredictable, and can no longer be derived from the meanings of its components (see Aronoff and Anshen 1998; Bauer 2001). The item has now become an irregular form. Once coined, the lexical item tends to become an unanalyzable lexical unit (see Bauer 1983). For example, in one of the meanings of *acompañar+miento* ‘accompaniment’ (from *acompañar* ‘to accompany’) the derived form is compositional because the suffix *-miento* adds a meaning of ‘effect’ or ‘action’ to the base (see Lang 1990; Varela Ortega 2005). However, aside from this compositional meaning, *acompañar+miento* has three additional specialized meanings, namely, ‘group of people that accompany’, ‘musical accompaniment’, and ‘accompanying food’, all of which are considered lexicalized forms (see example with *naturalize* in Aronoff and Anshen 1998). In a similar way, the adjective *confianza+udo* (from *confianza* ‘trust’), which should mean ‘having a lot of trust’, has undergone a shift in meaning so that the notion of ‘feeling entitled to’ has been added, resulting in the lexicalized meaning ‘feeling entitled to a lot of trust’, which can be conveyed in English as ‘overfamiliar’, ‘fresh’, or ‘forward’.

Given this notion of lexicalization, in the present study any derivative that has undergone any type of semantic drift or shift is considered lexicalized. Therefore, metaphors,² figurative language, and any other type of extension are also considered here forms of lexicalization. This definition of lexicalization is quite broad and may thus seem vague. However, as seen in Section 3, it allows for the corpus analysis that tests the hypothesis to be as conservative as possible and therefore to encompass the largest number of possible lexicalized items. Since, as we have seen, lexicalized words are irregular, according to the dual-route model (see Section 1) they should be memorized and stored as whole units (see Bauer 2001), in contrast to (derived) compositional words, which are computed by rule. It would thus be interesting to find out what percentage of derivatives in Spanish are lexicalized versus those that are compositional.

This led me to formulate the following research question (and associated possible answers), which serves as the basis for the study’s hypothesis (see below): what proportion of derived words in Spanish is lexicalized? If a significant proportion of derived words is lexicalized, that would be an indication of the robustness of lexicalization and the relative weakening of the compositional rule in Spanish derivation. On the other hand, if compositional word formation is found to be dominant, lexicalization may well turn out to be only a marginal phenomenon within derivation. In other

¹ See Jackendoff (2010) for an alternative view in which both regular and irregular (or semiproductive) forms are produced by rules, but productive rules are marked (by stipulation) with the diacritic feature [+productive] and the semiproductive rules are left unmarked.

² As Lieber and Baayen (1993) suggest, metaphorical use is a type of lexicalization.

words, if lexicalization were much more frequent than regular composition, that could be an indication that its use is extending over the lexicon and that it is becoming the dominant process for generating new words, over and above the regular rule. This could in turn lead to the possible consequence that the regular rule may be used only to create new words, with lexicalization then taking over in creating (new) meaning. However, a robust regular rule and a very low level of lexicalization is what the dual-route model would predict; lexicalized items, having a non-predictable meaning are stored as irregulars, while regular forms are created as usual by the concatenation rule and retain their compositional meaning.

On the basis of the above research question, in addition to intuitive speculation and an exploratory search of the *Corpus del Español* (CDE; Davies 2002), the following hypothesis was formulated: at least 80 percent of derivatives in Spanish are compositional (that is, at the most 20 percent of derivatives in Spanish are lexicalized); therefore, lexicalization is a secondary process in Spanish word formation. A corpus study employing two corpora was conducted to test the hypothesis.

To my knowledge, similar studies have not been conducted so far. Although lexicalization has been addressed in earlier works using the same or a very similar definition to the one adopted here, these studies either focus on non-derivational processes, such as the lexicalization of prepositional phrases (eg. Sp. *en seguida* becoming *enseguida* ‘right away’; Elvira 2006), or on expressive (appreciative or affective) affixation (Montero Curiel 2008; Juliá Luna and Prat Sabater 2013), which is not considered strictly derivational because it never changes the grammatical category of the base and because its semantic content is emotive rather than referential (see Scalise 1984; Lang 1990; Lázaro Mora 1993; Varela Ortega 2005). Other studies approach lexicalization from a historical rather than a synchronic point of view (Montero Curiel 2008). Although some of these works provide data from online corpora or dictionaries, they do not involve a corpus study that measures the degree of lexicalization in derivation across a range of affixes, as the present study does. Thus, the current article represents a significant contribution to the literature on lexicalization.

The study is based on an analysis of 10,046 derivatives (ie. tokens), formed with 5 suffixes and 3 prefixes, with varying degrees of productivity, for a total of 63 word forms or types (33 types formed with suffixes, 30 formed with prefixes). The results yield 77 percent of compositional items for the suffixed words, and 99 percent for prefixed words, for an 87 percent overall level of compositionality: 1,350/10,046 lexicalized items (13 percent) (for further details see Section 3), thus confirming the hypothesis put forward in the preceding paragraph and therefore supporting the author’s intuitions. In addition, these results provide support for the dual-route model: there seems to be a robust concatenative, compositional rule at work in Spanish derivation that generates large numbers of regular, semantically transparent forms. Lexicalized items, not being computable by rule, need to be stored in the mental lexicon (in an associative network).

I believe that the number of derivatives (embedded in sentences) analyzed in this article (slightly over 10,000) is a significant amount of material from which to draw solid conclusions, especially since the analysis consisted of the manual classification and interpretation of the 10,000+ derivatives. The 8 specific affixes analyzed were chosen due to their differing degrees of productivity. As shown on the list below, their productivity indices ranged from .31 percent to 4.81 percent, a relatively wide range of productivity. The measure of productivity (using CDE items) was calculated using the following formula (from Baayen 1991; notation slightly adapted):

$$P = HL/N$$

where P is the index of productivity, HL is the number of hapax legomena, that is, the number of words formed by a given morphological process occurring only once, and N is the total number of tokens formed by that morphological process that are found in the corpus. The following are the indices of productivity for the 8 affixes:

- iento*: Productivity cannot be calculated because no HL were found. This indicates a very low degree of productivity; the more HL, the more productive a given morphological process.
- izo*: $P = 2/644 = .0031 = .31$ percent
- ero*: $P = 36/1,409 = .0255 = 2.55$ percent
- azo*: $P = 3/801 = .0037 = .37$ percent
- era*: $P = 185/27,495 = .0067 = .67$ percent
- anti*–: $P = 429/10,901 = .0393 = 3.93$ percent
- multi*–: $P = 108/2,241 = .0481 = 4.81$ percent
- sub*–: $P = 368/7,938 = .0463 = 4.63$ percent

It is important to keep in mind that the hypothesis stated above is not merely about the preponderance of regular forms. Again, it is the proportion of lexicalized versus compositional forms that is key. Note also that for certain constructions there does not always exist a preponderance of regular forms in a particular language, and the regular rule need not apply to the most numerous forms either. Pinker (1999), for example, shows that German –s is the least common of the plural suffixes, yet it is used as the regular default (eg. *Cafés*, *Autos*). Nevertheless, it could be plausibly argued that the compositional rule in derivation functions as a default mechanism that produces the most natural meaning; as with the English plural (–s) and past tense (–ed), regular derivation in Spanish appears to yield a larger number of forms than irregular processes such as lexicalization.

In order to make comparisons with data from a corpus in addition to the CDE, all the derived words examined in the present study (10,046 tokens) were also analyzed using the CREA (*Corpus de referencia del español actual*; Real Academia Española 2011–2013). As shown in Section 3, the results of that analysis correspond relatively closely to those obtained from the CDE, providing further strong support for the conclusions of this study (see Tables 1 and 2). Moreover, since, as mentioned above, metaphorical uses are considered lexicalized forms, these corpus results provide some evidence against the notion that metaphor is prevalent in language and an essential part of thought, as argued (especially within Cognitive Linguistics) since Lakoff and Johnson (1980) (see Pinker 2007). At least in Spanish derivation, metaphor does not seem to be such a dominant force (see Section 3). This of course does not mean that metaphor is an insignificant part of thought or that it plays only a minor role in Spanish (or language in general). As Pinker (2007) notes, some metaphors can express truths about the world and can help us capture aspects of reality. The observations made above regarding metaphor are not meant to discredit Lakoff and Johnson's work or Cognitive Linguistics in general. Rather, they are plausible conclusions that follow from the evidence gathered in this study.

Frequency is an important factor when it comes to the issue of word storage as related to lexicalization. Highly frequent words tend to be stored in the mental lexicon (Plag and Baayen 2009) and even forms that are semantically transparent may be stored, especially if they are high frequency items (Lindquist 2009). In turn, according to Plag et al. (2008), higher frequency items undergo a higher degree of lexicalization. On the other hand, as Pinker (1999) and Jackendoff (2010) observe, citing psycholinguistic evidence, forms derived by regular morphology do not need to be stored, since they are built up by free combination, just like phrases. In contrast, non-compositional forms, which have unpredictable meanings, need to be stored (see also Plag and Baayen 2009). However, both Pinker and Jackendoff acknowledge that some regular, compositional forms may be stored, especially, again, if they have a high degree of frequency.

It must be stressed that this is not an article about the mental processing of morphology or an analysis of the strengths and weaknesses of the dual-route model, but primarily a corpus-based work. Accordingly, one of the goals of the article is to show how a corpus study can contribute evidence that is relevant to a particular theoretical framework, in this case the dual-route model. More specifically, as the research question above suggests, the main goal of the article is to determine the degree of lexicalization (and therefore of compositionality) in Spanish derivation through a corpus study. Furthermore, no claim is made here that these findings are key to the dual-route model. Rather, evidence is presented that may be seen as lending additional support for the dual-route model.

Before moving on to the description of the corpus study and the analysis of results, I briefly discuss the reasons for employing a corpus study. Stubbs (2002) suggests that native speaker intuition should be combined with the observable data that corpora provide, since corpus studies can confirm (or contribute to refute) intuitions, in addition to providing more detailed data than introspection by itself. Davies (2008) notes that these databases have become a necessity to verify linguistic introspection. In addition, as Aronoff and Anshen (1998) observe, counts based on large corpora are reliable sources of linguistic information because they measure actual use in authentic texts. On the basis of these observations, I decided to undertake a corpus study for the linguistic analysis of lexicalization in Spanish word formation.

3. CORPUS STUDY

As noted in Section 2, the corpus study was conducted using the *Corpus del Español* (Davies 2002), an online database consisting of more than 100 million words from more than 20,000 Spanish texts from the 1200s to the 1900s. From 1900 onwards the corpus texts are distributed in four registers, namely, spoken, fiction, newspaper, and academic. Only texts from the twentieth-century section of the corpus were searched because of the availability of register information and because they represent the most modern usage. Such texts consist of over 20.5 million words, with approximately equal amounts of words per register. Examples from all four registers were used in the study. The corpus used for comparison is the CREA (Real Academia Española 2011–2013), an online Spanish corpus containing over 150 million words from texts published between 1975 and 2004, belonging to four registers (or sources): books (fiction and non-fiction), magazines, press, and oral transcription. As with the CDE, concordance lines with examples from all four registers in the CREA were taken into account.

The CDE is a tagged corpus, so it allows searches for lemmas, parts of speech, and grammatical features, or any combination of these. Although the CREA is not tagged, it was chosen for comparison because of the large amount of material from the late twentieth and early twenty-first century it contains. Although CORPES XXI (*Corpus del español del siglo XXI*; Real Academia Española 2014) is a tagged corpus of about 174 million words, it was not chosen for the current study since, as a beta version, it is still under construction and, as noted on its Presentation page, some components still need adjustment. In addition, CORPES XXI does not yet contain transcriptions for oral texts, information about textual typology has been provided for only a small portion of the documents, and it has a very slow

interface as compared to the CDE and the CREA, especially after the first search. Only the CDE presents the results grouped by word type, as in (1) below, and shows a chart with sections that contain the number of results by century and by genre, which makes it easier to select the concordance for a particular century (see Section 3.2). In contrast, in corpora such as CORPES XXI and CNDHE (see description in footnote 10 below), results for all centuries are listed together in the concordance (in ascending or descending order by century) and the user has to separate the results per century visually.

3.1. Method

For each derived form, only concordances for items with at least 10 occurrences were considered for analysis.³ Of the derived words analyzed, 1,263 come from the CDE and 8,783 derivatives from the CREA, for a total of 10,046 derivatives (node words for purposes of the corpus study), formed with 5 suffixes and 3 prefixes, with varying degrees of productivity. As noted in Section 2, these 8 affixes were selected because they exhibit a relatively wide range of productivity, with indices going from .31 percent to 4.81 percent. In addition, as shown in Table 2 below, these 8 affixes cover a wide range of Spanish derivational processes, including those that produce denominal, deadjectival, and deverbal adjectives, nouns, and verbs.

The meanings of all node words (or KWICs: Key Words in Context) in the concordances were determined by hand by the author, who is a native speaker of Spanish. As Bauer (2001) notes, there is disagreement in the literature as to whether lexicalization is either semantic or phonological; that is, whether words may be phonologically lexicalized but semantically regular or vice versa. In this respect, Plag and Baayen (2009) observe that productive processes are both semantically and phonologically transparent. In turn, Mondorf (2009) gauges lexicalization (in compounds) in terms of spelling; words spelled together tend to be more closely integrated and less semantically transparent. For example, the compound *highrisk*, with no separation between the component words, would be considered more lexicalized than *high risk*.

In this study, semantics alone has been taken into account in determining the regularity of a derivative. For example, a search was done on the CDE of all the denominal adjectives ending in the suffix *-iento*,⁴ which yielded the ten word forms (ie. types) in (1), accompanied by the total number of occurrences of each type.

(1) Word forms for adjectival derivatives with *-iento* in the CDE

1	SANGRIENTO	58
2	HAMBRIENTO	55
3	POLVORIENTO	49
4	SEDIENTO	24
5	CENICIENTO	22
6	SOÑOLIENTO	19
7	GRASIENTO	14
8	MUGRIENTO	14
9	SOMNOLIENTO	11
10	HARAPIENTO	9
	TOTAL	275

The concordance for *hambriento* ‘hungry, ravenous’ in (2a) below illustrates the kinds of issues that have arisen in the determination of whether a given node word is lexicalized or, on the contrary, compositional. In the vast majority of cases the meaning of the node word is easy to determine – (2b) shows a portion of the concordance for *hambriento* from the CREA corpus, for comparison with the CDE format. For example, it is clear that in sentence 3 in (2a) *hambriento* has a literal meaning (a hungry animal), while in sentence 28 there is no doubt that *hambriento* is being used figuratively (or metaphorically), since it is modifying the noun *corazón* ‘heart’.

³ Although *harapiento*, with 9 occurrences, appears in (1), it is not used as an example in the corpus study itself, and is therefore not included in Table 2. See also footnotes 4 and 6 for relevant information about the characteristics of the concordance items (occurrences).

⁴ CREA, not being a tagged corpus, does not allow to search for specific grammatical categories, which makes it impossible to present results as in (1) from the CDE. Furthermore, for any search, the CDE presents the results grouped by word type, as in (1), something that is not possible with CREA either.

However, in some cases the dividing line between lexicalized and compositional items is not as clear-cut, especially when the semantic drift is only slight, or when there is some sort of ambiguity in the sentence. For example, in sentence 19 in (2a), it is a tumor that is hungry, not a human or an animal, so the meaning of *hambriento* appears to be figurative. However, in replicating uncontrollably, cells in tumors consume resources ravenously and therefore could be considered to be literally hungry. In borderline cases such as these,⁵ a conservative approach was taken and the decision was made to consider the derivative as lexicalized even when there was just a hint of some sort of semantic drift. By being as conservative as possible, the procedure for classifying corpus tokens (as either lexicalized or compositional) is thus also intended to be as objective as possible.

(2a) Concordance for *hambriento* in the CDE⁶

1	19-OR	Habla Culta: Bogotá: M38	únicamente iba. Inf.a. - Simpatizó conmigo y me dijeron que era sumamente hambriento , un rico hambriento, con joyas y con cantidad de cosas guardadas, y
2	19-OR	Habla Culta: Bogotá: M38	.a. - Simpatizó conmigo y me dijeron que era sumamente hambriento, un rico hambriento , con joyas y con cantidad de cosas guardadas, y entonces yo le dije
3	19-OR	Habla Culta: La Paz: M23	fiero, pero generalmente ese animal huye del hombre, salvo que esté herido o hambriento , pero generalmente huye. O sea que no hay una razón. Yo creo
4	19-F	Hijo de ladrón	cordillerano, la ropa y los zapatos destrozados, los pies llenos de heridas, hambriento y sucio, estaba Ipinza. Lo metí en la carpa del capataz, como
5	19-F	Hijo de ladrón	grupo hablaban muy fuerte y reían con más fuerza aún. Me sentía cansado, hambriento y desanimado. Nunca me había sentido más incapaz de nada. Allí no había
6	19-F	Hijo de ladrón	?, no lo sabía; de cualquier modo; estaba solo, enfermo y hambriento y no podía elegir; fuera de ellos no había allí más que el mar
7	19-F	Hijo de ladrón	en el mundo, flaco, además, y con cara de enfermo y de hambriento , debe tener, tiene que tener algo que contar. Me miró y como
8	19-F	Hijo de ladrón	; se hacía de noche, encendía una luz y leía; por fin, hambriento y cansado, me dormía hasta la mañana siguiente. No se podía seguir así
9	19-F	Hijo de ladrón	Mientras encuentra dónde acomodarse - advirtió. Una semana después, convertido en sirviente, hambriento , mal tratado, sucio y rabioso, comprendí que existía algo peor que perder
10	19-F	Hijo de ladrón	una vasta sala, ¿ y para qué quieres luz?; estás cansado o hambriento y sólo necesitas obscuridad y descanso, dormir o pensar; no sabes quién duerme
11	19-F	Hijo de ladrón	miseria y el hambre no tienen olfato; más aún, el olfato estorba al hambriento . La corteza, es la palabra más exacta, que la recubría, sonaba
12	19-F	Hijo de ladrón	diez o veinte presas y sólo tenía dinero para una y un panecillo. Estaba hambriento y comía y miraba. El pescadero, que parecía hecho de un material semejante
13	19-F	Cuentos de muerte y de sang...	una daga como de una brasada de largor y dio comienso a tragar a lo hambriento . En eso, y Dios parece que sirviera las miras del inglés, se
14	19-F	Xaimaca	encierra como una empalizada. Un roto nos mira desde el alambrado, como chimango hambriento de ojos. Espera en el Curacaví de los bueyes blancos y las casas lunares
15	19-F	Seis relatos	, como de una brasada de largor, y dio comienso a tragar a lo hambriento . »en eso, y Dios parece que sirviera las miras del inglés, se
16	19-F	De barro somos	llegaron hasta la bañera. Angélica derramó la leche en el agua. Mareado y hambriento , Raimundo se echó detrás. Los dedos nudosos hicieron el resto. No se
17	19-F	De barro somos	No sé... Algo me quema. Tengo ganas de orinar. Estoy hambriento . Creo que me voy a dar un buen baño antes de la cena.
18	19-F	Desde el encendido corazón ...	era el cielo, con su danza circular de estrellas, la aparición del tigre hambriento de luna, la marcha habitual de las constelaciones. Entre la intimidad de los
19	19-F	El destino, el barro y la c...	usted a mí de comprensión y comprensión... ¿ Comprende acaso el tumor hambriento que me está devorando los riñones?, ¿ lo comprende?, ¿ comprende

⁵ The number of borderline cases as described here constitutes a small percentage of the total, around or below 5 percent.

⁶ Note that the searches were not done for the lemma for each derivative, but rather on the basis of the grammatical category of the whole derivative (eg. those ending in *-iento* or beginning with *multi-*), as explained in footnote 4. Therefore, the search results were restricted to either singular masculine or singular feminine forms. For example, (2a) shows only *hambriento*, not *hambrientos* or *hambrienta*, and for *flojera* 'laziness', the form *flojeras* was not examined (the form *flojero* does not exist). Moreover, since CREA does not allow to search for the grammatical category of the derivative (see footnote 4 above), searches were done for the full word, eg. *hambriento* or *flojera*, with results as shown in (2b).

20	19-F	Donde ladrón no llega	mejor que estar solo, en medio de la oscuridad, tan cerca del Caañabé <u>hambriento</u> . Acurrucado entre los matorrales miró a los viajeros encendiendo la fogata, cuidando los
21	19-F	Micro cuentos para soñar en...	pedazo de su alfajor. - Claro que no te voy a invitar, osito <u>hambriento</u> - le dijo -. Vos jamás le invitás a nadie. [16] Luisito lloró
22	19-F	Tierra de Nadie-Ningúem	días bastarían para cumplir su cometido. En cambio, Gaúcho debía estar debilitado y <u>hambriento</u> tras su huida, pues no tuvo tiempo de llevar nada más que la ropa
23	19-F	El baile de tambor	. Lo había pensado infinitas veces. Se lo había imaginado constantemente mientras se ocultaba <u>hambriento</u> entre los bosques y bajaba por la noche a beber a los ríos o a
24	19-F	Casa de campo	Rosamunda y de Avelino. Cosme se sentó ceremoniosamente a la mesa. Pero, <u>hambriento</u> debido al tono de austeridad de las comidas de aquellos que no vivían en el
25	19-F	Casa de campo	fin su cuerpo a Mauro como quien entrega un trozo de carne a un perro <u>hambriento</u> , para que lo todo caso, que mis lectores estén tranquilos, porque Wenceslao
26	19-F	Casa de campo	Su gemelo había muerto al nacer dejándolo incompleto, de pestañas demasiado claras, siempre <u>hambriento</u> pese a que nunca dejaba de tener los bolsillos llenos de trozos de pan húmedo
27	19-F	De dónde son los cantantes	von Kopf bis Fuss auf Liebe eingestellt, abriendo en corazón sus boquitas de pez <u>hambriento</u> , en equilibrio, con los brazos extendidos, sobre una barra de hierro.
28	19-F	Maladrón: epopeya de los An...	cometas fosforescentes alumbren la sogá de nuestros pasos mágicos alrededor de esta pirámide, corazón <u>hambriento</u> de corazones de cautivos!!! - - vocean otros jefes. - -
29	19-F	Maladrón: epopeya de los An...	no las máscaras fugaces de la desvergüenza!... XVI Duero Agudo, <u>hambriento</u> , realmente hambriento, cayó en aquel galeón destinado a las Indias, con tan
30	19-F	Maladrón: epopeya de los An...	fugaces de la desvergüenza!... XVI Duero Agudo, hambriento, realmente <u>hambriento</u> , cayó en aquel galeón destinado a las Indias, con tan poco equipaje que
31	19-F	La muerte de Artemio Cruz	al alba: los chirríos de los pájaros escondidos, un grito agudo de niño <u>hambriento</u> , ese martilleo extraño de algún trabajador del pueblo, ajeno al estruendo invariable,
32	19-F	Tiempo de silencio	menguado pasto para los gusanos a través de cualquiera de las complicadas formas del morir <u>hambriento</u> (tuberculosis, escrófula, latirismo, erupciones de sangre, temblor progresivo de los
33	19-F	Manuel de historia	tenían inmovilizado mientras lo devoraban. Quería zafarse y no podía. Un monstruoso animal <u>hambriento</u> se había arrojado sobre él y le clavaba los colmillos, las pezuñas, lo
34	19-F	Miramar: La gesta del pez	ramas secas y las acomodó en la entrada para encender un fuego. No estaba <u>hambriento</u> , el almuerzo había sido abundante. El fuego ardía cuando sacó su cuaderno de
35	19-F	Papelucho	creo, porque Javier es muy hombre. Ahora no me puedo dormir de puro <u>hambriento</u> , porque a la comida tocó pescado, y me revienta, así es que
36	19-F	Calamares	la cocina impidió a Mauricio enterarse de lo que ahí estaba pasando. Fastidiado y <u>hambriento</u> , conectó el último polo, bien pelado, a la entrada " + "
37	19-F	Cenicientos o el Infarto de...	me duermo.) Aquél era el primero. Me sentía y estaba drogado y <u>hambriento</u> . Lorena recibió a mi mamá y la acompañó a acostarse. Rubí había llegado
38	19-F	Crisol del Olvido, El	la zona para buscarlo. A eso del medio día encontré a Irigoyen agotado y <u>hambriento</u> . No habíamos tenido suerte, ni siquiera habíamos encontrado huellas. - ¿ Tú
39	19-F	Espacios Vacíos	lacerada, un absceso supurante, el ardiente silencio orgánico rojo negro blanco del líquido <u>hambriento</u> abismo de tus sensaciones de euforia y poder inagotable.....en
40	19-F	Gibraltar 1933	el hidroavión alcanzó la velocidad adecuada se levantó de la superficie prendido de espuma, <u>hambriento</u> de aire, como un albatros que sale de caza. - Manuela, habrá
41	19-F	Cortesana, La	emigrado abandonando incluso el establo. Si llegaba un caballero cansado de trajinar cabalgadura, <u>hambriento</u> , sediento o padeciendo otros males, se encontraba que allí no había posibilidades de
42	19-F	Criatura, La	especie. La voz del animal lo sobresaltó. Se veía a leguas que estaba <u>hambriento</u> . No lo dudó un instante. Subió a su departamento y abrió una de
43	19-F	Deuda, La	, con el estómago vacío y la esperanza de una prórroga. A pie, <u>hambriento</u> , pero ya sin ninguna esperanza, seguía camino a la prisión. Sinencio -
44	19-F	Otro Lado de la Niebla, Al	hombrunas de la mecanógrafa se aplicaron sobre el teclado de la máquina Underwood cuyo picoteo <u>hambriento</u> empezó a sonar en el quimérico despacho. Antes de reabsorberse en la penumbra,
45	19-F	Luz	blanco y negro. Inconscientemente, anhelaba sus " buenos días " como el perro <u>hambriento</u> y maltratado la caricia del carnicero. Su sangre hervía esperando el desayuno, la
46	19-F	Oyendo Llover	cuando Euralio estuviera dormido, si no se le subía. Es como un coyote <u>hambriento</u> de perra. Bajan la cabeza cuando se cruzan en mi camino. De seguro
47	19-F	Secta Secreta se Esconde en...	qué tipo de ambigües se habían dado en su primera exposición. Recordaba cómo, <u>hambriento</u> y desesperado, recurrió con estado depresivo, casi suicida, a su desprecio por

48	19-F	Tiempo Seco	regresa. Su ausencia ronda la habitación: cualquier ruido se convierte en el maullido hambriento del retorno, los pasos de una mujer y su hija, en el tercer
49	19-F	El regalo	. El miedo invadía el pueblo y sus alrededores. Se hablaba de un ser hambriento de quién sabe qué, que acechaba entre los caminos esperando la oportunidad de aparecersele
50	19-N	España:ABC	, museos e iglesias. Tiene un carácter en general pacífico, pero cansado, hambriento , sediento y de mal humor, condiciones muy habituales, puede atacar a las
51	19-N	España:ABC	, que Pritchett describe en Unamuno, se reduce aquí a la lucha entre el hambriento viajero y una vendedora de tomates que no le quiere vender sus tomates. Las
52	19-N	Cuba:CubaNet:98May8	aquí ". La mano discreta pasó por allí y con una letra de un hambriento amarillo contestó: " Ni el miedo ni nadie ". A la entrada del
53	19-N	Cuba:CubaNet:98Jul10	viendo a manos de aquella señora si sonreía. Abrí la boca como un sinsonte hambriento . Ya la doctora venía hacia mí con sus instrumentos de tortura. La puerta
54	19-AC	Enc: Budismo	ésta puede reencarnarse en un ser humano, en un animal, en un fantasma hambriento , en un habitante del infierno o incluso en alguno de los dioses de la
55	19-AC	Enc: Psicología	el animal se encuentre en la misma situación. Por ejemplo, si un animal hambriento es recompensado con comida por girar a la derecha en un laberinto simple, tenderá

(2b) Concordance for *hambriento* in the CREA

Nº	CONCORDANCIA	AÑO	AUTOR
1	ios. Hombres y mujeres que han visto a su hermano	2000	PRENSA
2	s de ellos acaben devorados entre los pétalos del	1995	PRENSA
3	esbarcaron con gran pompa para salvar a un país	1994	PRENSA
4	dos en la más completa vorágine, porque un pueblo	1997	PRENSA
5	plota con facilidad. En cambio, agregó, un pueblo	1997	PRENSA
6	juzgará por cosas tales como si dimos de comer al	1990	PRENSA
7	vez una ciudad gobernada por un personaje avaro y	1996	PRENSA
8	al país político -mundillo de pajarilla- y no al	1996	PRENSA
9	as con ellos. Simultáneamente con el hijo pródigo	2000	PRENSA
10	elea del año", declaró Joppy. "La necesito. Estoy	2001	PRENSA
11	tarde o temprano lo logró. Todo el mundo esperaba	1997	PRENSA
12	so sí, la anoréxica se comporta como un organismo	2002	PRENSA
13	erte es un tipo de cambio o estado, como el estar	2000	PRENSA
14	a fortaleza que lo mantiene luchando, aunque esté	2000	PRENSA
15	o tengo pan y no tengo por qué compartirlo con el	1996	PRENSA
16	nos compete-deber de ciudadano - que no haya niño	2003	PRENSA
17	qua 2001 2001 10 111 P Regresa	2001	PRENSA
18	r Pérez R. wilder.perez@laprensa.com.ni Como un	2002	PRENSA
19	me ha reservado ese trono para mí". Rosendo sigue	1997	PRENSA
20	s esa inquieta barranquillera perdida en un París	1997	PRENSA
21	nio, a sus obras y a sus pompas y dan de comer al	1997	PRENSA
22	una. Es como si le sirven un fastuoso cocido a un	1997	PRENSA
23	ay recursos alimenticios y minerales que un mundo	1977	PRENSA
24	nial occidental, la Unión Soviética, como un lobo	1977	PRENSA
25	mirando desde aquí el mundo, como un niño pobre y	1988	PRENSA

Ir arriba Pantalla: 1 de 15. Siguiente 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 Ver párrafos

It is significant that the vast majority of lexicalized derivatives examined in the present study have bases which are themselves lexicalized, usually figurative or metaphorical. It could well be argued that these derived forms are actually compositional since the affix combines in a transparent way with the metaphorical base. For instance, in example 40 in (2a) above, the hydroplane is said to be "hungry for air". Since a hydroplane is a machine, it is clear that the base *hambre* 'hunger' is metaphorical (or figurative) and that the suffix *-iento* is added to form the compositional – and still metaphorical – *hambriento* 'hungry.' Even so, cases such as these have been counted as lexicalized items given that the meaning of the base has undergone semantic shift and that base is participating in a process of derivation. Although the affix and the base may be combining transparently, there is some lexicalization involved (that of the base and the meaning of the entire derivative), so a conservative approach has been taken here. Were these items to be tallied as compositional, the total number of lexicalized derivatives would be drastically reduced from the already low 13 percent yielded by the study.⁷ There are extremely few lexicalized derivatives that are formed with literal bases. For instance, as seen in Section 2, the adjective *confianz+udo* has the literal base *confianza* 'trust', and compositionally it should mean 'having a lot of trust'. Nevertheless, it appears that the combination of base and suffix has undergone a shift in meaning, so that the notion of 'feeling entitled to' has been added to the whole.

Finally, though few items in the concordances were proper names, for some types many or all of the tokens were proper names, and were therefore discarded as candidates for the analysis. For example, *varadero* 'dry dock' was initially considered, but since 44 out of the 45 occurrences of this form in the CDE corresponded to the proper name Varadero, a town in Cuba, the form was discarded.

⁷ A lower proportion of lexicalized items would further strengthen the conclusions of this study. However, the hypothesis has been confirmed with the data analyzed in this way.

Before moving on to the results, mention should be made of the relatively reduced number of types available in Spanish for each of the derivatives analyzed in the present study (see Table 2, Section 3.2). For example, out of the almost 1,500 word forms (types) ending in *-iento* in the *Diccionario Inverso de la Lengua Española* (d'Urgell i Rubió 2003), only a small fraction (25, representing 1.7 percent) are the kind of derivative of interest in this article, namely, denominal adjectives such as *sediento* (7 types are analyzed in the present study). The remaining types consist of a menagerie of items, including simplex words, such as *aliento* 'breath', and deverbal nouns ending in the suffix *-miento* (eg. *derribamiento* 'downing'), which constitute the vast majority of derivatives, among others. The same applies to the remaining derivatives analyzed in the present study, including those formed with prefixes. These proportions are similar to those obtained from other inverse dictionaries, such as GoodRAE and DiRAE⁸ (close to 1,530 and 1,100 word forms ending in *-iento*, respectively).

Furthermore, many of the 25 derivatives ending in *-iento* in the *Diccionario Inverso de la Lengua Española* have extremely low productivity,⁹ to the point that many of them are either hapax legomena or do not appear at all in either the CDE or the CREA. For example, neither *gargajiento* 'that produces much phlegm' nor *hediento* 'smelly' appears in the two corpora. The 7 types ending in *-iento* analyzed in this study, out of less than 25 relatively productive derivatives in the *Diccionario Inverso de la Lengua Española*, can well be considered representative word forms, as well as the remaining derivatives also analyzed in this article.

3.2. Results and discussion

Table 1 below presents a summary of the results of the corpus study by affix. It is immediately followed by Table 2, which shows the full results by derived form. Percentages have been rounded to the nearest whole number in both tables. As can be seen, only 10 percent of CDE derivatives (tokens) and 14 percent of CREA derivatives are lexicalized, which confirms our hypothesis (that at least 80 percent of derivatives in Spanish are compositional, which suggests that lexicalization is a secondary process in Spanish word formation), and provides additional support for the dual-route model.

Table 2 shows that few suffixed words have lexicalization levels of 30 percent or higher (the percentages for these items appear in boldface), and none of the prefixed forms has a lexicalization level higher than 27 percent (see below). Note also that for one derivative, *hervidero* 'seething/boiling mass, swarm, hotbed', all forms are lexicalized in both corpora. This is interesting in itself, but it is worth noting that in the definitions of *hervidero*, *atolladero*, and other highly lexicalized derived words, dictionaries such as the *Diccionario Clave*¹⁰ and the *Diccionario de la Real Academia Española* (DRAE) include the metaphorical sense in addition to their literal meaning. This suggests that these forms have indeed reached a high degree of lexicalization. In contrast, highly compositional words such as *heredero* 'heir' tend not to have a metaphorical sense in their dictionary definitions; both the *Diccionario Clave* and the DRAE list only three literal senses for this word.

In addition, *hervidero* is interesting in that it appears to have been significantly less lexicalized in the past. A search was done for this word in the CDE for the nineteenth century, yielding a 57 percent level of lexicalization (16/28 tokens) (... *el hervidero del volcán* 'the volcano's seething/boiling crater'). Searches for earlier centuries in the CDE yielded only one or no tokens. Since no searches of this kind can be conducted on the CREA, I resorted to CORDE (*Corpus diacrónico del español*; Real Academia Española 2011–2013), a historical online Spanish corpus containing over 230 million words from texts published between the time of the earliest written records of Spanish (no date given) and 1975. It contains texts from two major registers (or genres): fiction (eg. novels and poetry) and non-fiction (e.g. press and scientific writing). A search for *hervidero* in the nineteenth century material in CORDE yielded a level of lexicalization similar to the one obtained with the CDE: 61 percent (41/67 instances) (there was only one instance of *hervidero* in the eighteenth century and another in the seventeenth century). Although it seems that *hervidero* has been highly lexicalized since it first arose (probably by the seventeenth century), the results from the CDE, CREA, and CORDE suggest that at some point in the twentieth century its degree of lexicalization exploded to reach 100 percent.¹¹

⁸ GoodRAE is an inverse, lemmatized, and hypertextual online dictionary that uses the *Diccionario de la Real Academia Española* (DRAE) as its source, while DiRAE is the *Diccionario Inverso de la Real Academia Española*, also based on the DRAE.

⁹ GoodRAE and DiRAE yield similar results.

¹⁰ The *Diccionario Clave* was chosen for comparison with the DRAE because the author is well familiarized with it and because both dictionaries have approximately the same number of entries: 80,000 in the *Diccionario Clave*, close to 90,000 in the DRAE.

¹¹ The *Corpus del Nuevo diccionario histórico del español* (CNDHE; Instituto de Investigación Rafael Lapesa de la Real Academia Española, 2013), a historical online corpus containing over 355 million words, was not used for this search for several reasons. First, because its texts are taken largely from selected CORDE (and CREA) material. Moreover, although the CNDHE is a tagged corpus, it has several disadvantages: it has a very slow interface, especially after the first search, and though the results per page are numbered, unlike other corpora, the total number of results for a given search are not indicated. The *Diccionario del castellano del siglo XV en la Corona de Aragón* (DiCCA-XV), another historical corpus of Spanish, is restricted to the language used in the Crown of Aragon in the fifteenth century. Its interface, which is also rather slow after the first search, is not particularly user-friendly, since it does not have a search field; searches have to be done by scrolling. Since the DiCCA-XV only contains texts from

As can be seen in Tables 1 and 2, the overall CREA results correspond relatively closely with those obtained from the CDE, providing further support for the conclusions of this article. In addition, these results give a hint as to the predictive force of a corpus study. They suggest that the results from a sufficiently large corpus, such as the CDE, tend to predict that similar results will be obtained when data is gathered from a corpus of similar size or larger, such as the CREA. Moreover, as noted in Section 2, the results from both corpora provide evidence against the notion that metaphor is prevalent in language and an essential part of thought, as argued (especially within Cognitive Linguistics) at least since Lakoff and Johnson (1980) (see Pinker 2007). Metaphor does not appear to be such a productive device, at least in Spanish derivation. As mentioned in Section 2, this does not imply that metaphor is an insignificant part of thought or that it plays only a minor role in Spanish (or language in general). As Pinker (2007) notes, some metaphors can express truths about the world and can help us capture aspects of reality. Again, these observations regarding metaphor are not meant to discredit Lakoff and Johnson's work or Cognitive Linguistics in general. Rather, they are plausible conclusions that follow from the evidence gathered in this study.

At this point, the question could be asked whether metaphors could be much more productive in other derivations, which brings us back to the relatively low number of affixes studied. As noted in Section 4 below, a similar study could be conducted examining a wider range and number of suffixes and prefixes. This would of course expand the number of derivations available for study. However, it is unlikely that the trend identified with the analysis of these 10,000+ tokens will differ much from that resulting from the examination of additional derivation types, for three reasons. First, as mentioned above, over 10,000 derivatives (embedded in sentences) is taken to be a significant amount of material from which to draw solid conclusions. Second, the affixes studied in this article show varying degrees of productivity (see Section 2). Finally, and most importantly, as also noted in Section 2, though the preliminary (or exploratory) search was done only with the CDE, the CDE final results were compared to those from the CREA, and the degree of lexicalization corresponded closely between the two corpora (recall that all word types examined in the CDE were also analyzed using the CREA corpus, for a total of 10,046 tokens). Given that the trend identified in the CDE with a significant number of tokens was confirmed in the CREA, this is likely to happen if further prefixes and suffixes are added to the picture.

It is worth pointing out that although metaphors are often expressed as part of phrases or sentences, it is commonly one or two words (whether bases or derivatives) in the utterance that carry the metaphorical weight; that is, the metaphor is not necessarily composed from the meanings of all the elements of the phrase or sentence. For example, in the literary metaphor *It is the east, and Juliet is the sun*, the words *east* and *sun* are metaphors, but the rest of the words, including *Juliet*, have a literal meaning. The same applies to conceptual metaphors, that is, underlying metaphors that are implicit in a family of related expressions or figures of speech (see Pinker 2007). For instance, although the utterances *Your claims are indefensible* and *I demolished his argument* fall under the conceptual metaphor ARGUMENT IS WAR (Lakoff and Johnson 1980), it is the words *indefensible* and *demolished*, respectively, that carry the metaphorical meaning in the corresponding sentences. Crucially, *claims* in the first sentence and *argument* in the second one retain their literal meaning. Therefore, the results of the present study, which focuses on single words, are a valid basis to support the claim that metaphors are not prevalent in language and are not an essential part of thought. The fact that the study focuses on single words (as opposed to phrases or sentences, as explained above) offers further validation to the observation that metaphors are not prevalent in Spanish.

As can be seen in Table 2, for the 8 affixes studied here, compositionality in prefixed derivatives turned out to be much higher than in suffixed words. While some suffixed words have rather high lexicalization levels (eg. *asidero* 'handle' in the CDE with 65 percent), the highest degree of lexicalization for a prefixed word is 27 percent (*submundo* 'underworld'; CREA). Prefixes seem to combine with their bases in a more transparent way; they seem to be more 'separated' from their bases than suffixes. These results seem to provide support for the idea, summarized in Lang (1990), that the semantic cohesion between prefix and base is much looser than that between suffix and base; thus, compositionality in derivation with prefixes is stronger. This separation between prefix and base is manifested in phonology as well. As Varela Ortega (2005) observes, unlike suffixes, prefixes tend to preserve their phonological identity and do not fuse with their base, with few exceptions, even when two vowels are side by side (see *pre-escolar* 'pre-school'; *anti-inflamatorio* 'anti-inflammatory') (see also Varela Ortega 1990). This notion is also related to Mondorf's (2009) finding that compounds whose elements are more closely integrated (eg. *hard-nosed*) – an indication that the compound is lexicalized – are less likely to take the synthetic comparative form (eg. **harder-nosed*) than compounds whose component words are more separated (*hard nosed* – *harder nosed*), and thus are more likely to be compositional (see spelling-related examples in Section 3.1).

In addition, as also noted by Lang (1990), prefixes are usually less ambivalent than suffixes, with a tendency towards monosemy, with a clear and constant meaning. Lang adds that prefixed forms are ephemeral and less well-established than suffixed words. Suffixes are more semantically integrated to their bases, and therefore tend to become more

the fifteenth century, as might be expected from the CORDE results (no occurrences of *hervidero* before the seventeenth century), the word *hervidero* does not appear in this corpus.

established. In turn, more established items tend to be used more frequently and are therefore more likely to be stored and undergo lexicalization (see Section 2).

Finally, it is worth noticing that the number of suffixed words analyzed in this study (5,770) is 35 percent higher than that of prefixed words (4,276). Given the high level of compositionality evinced by prefixed words, had the number of suffixed words been about equal to that of prefixed forms, the overall level of lexicalization is likely to have been lower than 13 percent.

The full corpus results are shown in Table 2. Words appear in alphabetical order. Cognates (eg. *antiviral*) have not been provided with an English translation. Note that unlike the CDE, the CREA corpus does not allow the visualization of more than 1,000 tokens (concordance lines) as the result of a given search. Since the total number of tokens in the CREA for some of the derivatives under study was higher than 1,000 (eg. derivatives with *-dero* and *anti-*, as seen in Table 1), each word type (eg. *criadero*) was searched individually. This way, no single word form had more than 1,000 occurrences in the corpus (see Table 2 and related information in footnotes 4 and 6 above).

	Derivatives	CDE			CREA			Both corpora			
		Lex items	Total items	% lex	Lex items	Total items	% lex	Grand Total Lex items	Grand Total items	% lex GrandTotal	Compositional items
Suffixes											
1	<i>-iento</i>	39	196	20%	149	1,147	13%	188	1,343	14%	
2	<i>-izo, -dizo</i>	34	141	24%	267	988	27%	301	1,129	27%	
3	<i>-dero</i>	38	179	21%	581	1396	42%	619	1,575	39%	
4	<i>-azgo</i>	5	172	3%	135	996	14%	140	1,168	12%	
5	<i>-era</i>	4	88	5%	51	467	11%	55	555	10%	
	Total	120	776	15%	1,183	4,994	24%	1,303	5,770	23%	77%
Prefixes											
1	<i>anti-</i>	5	165	3%	3	1,382	0.2%	8	1,547	0.5%	
2	<i>multi-</i>	2	108	2%	12	806	1%	14	914	2%	
3	<i>sub-</i>	2	214	1%	23	1,601	1%	25	1,815	1%	
	Total	9	487	2%	38	3,789	1%	47	4,276	1%	99%
	Grand total	129	1,263	10%	1,221	8,783	14%	1,350	10,046	13%	87%

Table 1: Summary of results from the *Corpus del Español* (CDE) and the *Corpus de Referencia del Español Actual* (CREA) (Lex items = lexicalized items; % lex = % lexicalized items)

	Derivatives	CDE			CREA		
		Lex items	Total items	% lex	Lex items	Total items	% lex
Suffixes							
	<i>-iento</i> N → A						
1	ceniciento 'ash colored'	3	22	14%	8	116	7%
2	grasiento 'greasy'	2	14	14%	5	79	6%
3	hambriento '(very) hungry'	11	55	20%	49	366	13%
4	mugriento 'filthy'	3	14	21%	5	92	5%
5	polvoriento 'dusty'	1	48	2%	6	228	3%
6	sediento 'thirsty'	15	24	63%	71	193	37%
7	soñoliento 'sleepy'	4	19	21%	5	73	7%
	Total	39	196	20%	149	1,147	13%
	<i>-izo, -dizo</i> N → A A → A V → A	Lex items	Total items	% lex	Lex items	Total items	% lex
1	castizo 'of pure caste/descent'	3	22	14%	8	198	4%
2	cobrizo 'copper-colored'	0	16	0%	0	73	0%
3	escurridizo 'slippery'	8	20	40%	55	154	36%
4	huidizo 'evasive, elusive'	7	16	44%	40	96	42%
5	movedizo 'mobile, unstable'	6	17	35%	26	85	31%
6	plomizo 'leaden'	3	21	14%	34	117	29%
7	quebradizo 'brittle'	3	15	20%	20	90	22%
8	resbaladizo 'slippery'	4	14	29%	84	175	48%
	Total	34	141	24%	267	988	27%

	<i>-dero</i> V → N	Lex items	Total items	% lex	Lex items	Total items	% lex
1	aserradero 'sawmill'	0	15	0%	1	45	2%
2	asidero 'handle'	11	17	65%	184	207	89%
3	atolladero 'mire, bog'	11	14	79%	179	187	96%
4	bebedero 'trough'	0	12	0%	0	39	0%
5	bombardero 'bomber'	0	16	0%	3	75	4%
6	criadero 'breeding place'	1	14	7%	4	124	3%
7	desaguadero 'drain'	0	11	0%	9	14	64%
8	embarcadero 'jetty, wharf'	0	30	0%	0	174	0%
9	encomendero 'grocer/colonist'	0	11	0%	0	66	0%
10	fregadero 'kitchen sink'	1	13	8%	0	181	0%
11	hervidero 'seething/boiling mass, swarm, hotbed'	14	14	100%	188	188	100%
12	picadero 'exercise ring (for horses)'	0	12	0%	13	96	14%
	Total	38	179	21%	581	1,396	42%
	<i>-azgo</i> N → N	Lex items	Total items	% lex	Lex items	Total items	% lex
1	hartazgo 'being stuffed, gorged'	2	18	11%	120	141	85%
2	mecenazgo 'patronage'	0	96	0%	0	359	0%
3	noviazgo 'engagement'	3	58	5%	15	496	3%
	Total	5	172	3%	135	996	14%
	<i>-era</i> A → N	Lex items	Total items	% lex	Lex items	Total items	% lex
1	flojera 'laziness'	0	20	0%	0	142	0%
2	sordera 'deafness'	4	53	8%	51	304	17%
3	tontera 'silliness'	0	15	0%	0	21	0%
	Total	4	88	5%	51	467	11%
	Total for suffixes	120	776	15%	1,183	4,994	24%
Prefixes							
	<i>anti-</i> A → A N → A N → N	Lex items	Total items	% lex	Lex items	Total items	% lex
1	anticoncepción	0	10	0%	0	104	0%
2	anticorrupción	0	13	0%	0	205	0%
3	antidisturbios 'anti-riots'	0	10	0%	0	230	0%
4	antiesclavista 'antislavery'	0	10	0%	0	21	0%
5	antimateria	4	33	12%	3	82	4%
6	antimonopolio	0	13	0%	0	44	0%
7	antinatural	1	11	9%	0	125	0%
8	antirretroviral	0	10	0%	0	63	0%
9	antisemita	0	20	0%	0	106	0%
10	antisocial	0	10	0%	0	192	0%
11	antiviral	0	15	0%	0	87	0%
12	antivirus	0	10	0%	0	123	0%
	Total	5	165	3%	3	1,382	0.2%
	<i>multi-</i> N → A N → N A → A	Lex items	Total items	% lex	Lex items	Total items	% lex
1	multicolor	2	27	7%	12	275	4%
2	multicultural	0	16	0%	0	108	0%
3	multimillonario	0	25	0%	0	308	0%
4	multipartidista 'multiparty'	0	15	0%	0	32	0%
5	multirregional	0	15	0%	0	21	0%
6	multiuso 'multipurpose'	0	10	0%	0	62	0%
	Total	2	108	2%	12	806	1%

	sub- N → N V → V A → A	Lex items	Total items	% lex	Lex items	Total items	% lex
1	subcomisión	0	12	0%	0	210	0%
2	subconjunto 'subset'	0	12	0%	0	75	0%
3	subdesarrollado 'underdeveloped'	0	25	0%	0	216	0%
4	subdividir	0	12	0%	0	33	0%
5	subdivisión	0	44	0%	0	150	0%
6	subempleo 'underemployment'	0	17	0%	0	147	0%
7	subespecie	0	17	0%	0	209	0%
8	subgerente 'second- level manager'	0	19	0%	0	79	0%
9	subgrupo	0	13	0%	0	202	0%
10	submundo 'underworld'	2	12	17%	23	86	27%
11	subproducto 'byproduct'	0	18	0%	0	169	0%
12	subunidad	0	13	0%	0	25	0%
	Total	2	214	1%	23	1,601	1%
	Total for prefixes	9	487	2%	38	3,789	1%
	Grand total	129	1,263	10%	1,221	8,783	14%

Table 2: Complete results from the *Corpus del Español* (CDE) and the *Corpus de Referencia del Español Actual* (CREA)
(Lex items = lexicalized items; % lex = % lexicalized items)

4. CONCLUSION

The starting hypothesis of this article was that at least 80 percent of derivatives in Spanish are compositional and that, therefore, lexicalization is a secondary process in Spanish word formation. The results of the corpus study conducted to test this hypothesis yielded 87 percent of compositionality, and thus confirmed the hypothesis and the author's intuitions, also providing additional support for the dual-route model. Although only 13 percent of the items under study here were lexicalized, the data contained several derived forms with levels of lexicalization higher than 30 percent, with a few derivatives even reaching levels above 60 percent. To use a metaphor, it seems as if the largely compositional derivational landscape is dotted here and there with small pockets of lexicalization.

A similar study may be conducted examining a wider range and number of suffixes and prefixes as well as a larger set of tokens, not only in Spanish but also in other languages. Future studies may also compare the lexicalization levels of underived (or simplex) bases – where lexicalization is supposedly more prevalent – to the lexicalization of derived forms. All such studies would help to further test whether or not metaphor is prevalent in language.

Finally, as seen in Section 3, derivatives that are currently highly lexicalized, such as *hervidero*, have apparently seen their degree of lexicalization grow over time. An exploration of how lexicalization levels of derived forms in Spanish have changed through the centuries would represent an important contribution to historical linguistics.

REFERENCES

- Alegre, Maria and Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40: 41–61.
- Alonso Calvo, Raquel. 2009. Lexicalización y colocaciones: una introducción a su estudio diacrónico. *Onomázein: Revista de Lingüística, Filología y Traducción* 19: 33–56.
- Aronoff, Mark and Frank Anshen. 1998. Morphology and the lexicon: lexicalization and productivity. In Andrew Spencer and Arnold Zwicky eds. *The handbook of morphology*. Oxford: Blackwell, 237–247.
- Baayen, Harald. 1991. Quantitative aspects of morphological productivity. In Geert Booij and Jaap van Marle eds. *Yearbook of morphology*. Dordrecht: Kluwer, 109–149.
- Bauer, Laurie. 1983. *English word-formation*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 2001. *Morphological productivity*. Cambridge: Cambridge University Press.
- Blasco Mateo, Esther. 2006. Acerca de la lexicalización de las secuencias en las que interviene *so* en el siglo XV. In José Luis Girón Alconchel and José Jesús de Bustos Tovar eds. *Actas del VI Congreso Internacional de Historia de la Lengua Española*. Madrid: Arcos Libros.

- Brinton, Laurel J. and Elizabeth Closs Traugott. 2005. *Lexicalization and language change*. Cambridge: Cambridge University Press.
- Buenafuentes de la Mata, Cristina and Carlos Sánchez Lancis. 2012. Procesos de gramaticalización y lexicalización a la luz de los corpus académicos. In Tomás Jiménez Juliá, Belén López Meirama, Victoria Vázquez Rozas and Alexandre Veiga eds. *Cum corde et in nova grammatica. Estudios ofrecidos a Guillermo Rojo*. Santiago de Compostela: Universidade de Santiago de Compostela, 153–165.
- Davies, Mark. 2002–. *Corpus del Español*. <<http://www.corpusdelespanol.org>>
- Davies, Mark. 2008. New directions in Spanish and Portuguese corpus linguistics. *Studies in Hispanic and Lusophone Linguistics* 1: 149–186.
- d'Urgell i Rubió, Jaume. 2003. *Diccionario Inverso de la Lengua Española*. Madrid: JDR Ediciones.
- Elvira, Javier. 2006. Aproximación al concepto de lexicalización. *Diacronía, lengua española y lingüística*. In Javier Rodríguez Molina and Daniel Moisés Sáez Rivera eds. Madrid: Síntesis, 21–41.
- Instituto de Investigación Rafael Lapesa de la Real Academia Española. 2013. *Corpus del Nuevo diccionario histórico*. <<http://web.frl.es/CNDHE>>
- Jackendoff, Ray. 2010. *Meaning and the lexicon*. Oxford: Oxford University Press.
- Julià Luna, Carolina and Marta Prat Sabater. 2013. Derivación apreciativa y procesos de lexicalización en nombres de animales. Delivered at 27e Congrès International de Linguistique et de Philologie Romanes. Nancy. <<http://www.atilf.fr/cilpr2013/programme/resumes/732a6bbd6273a2faf755157aef71f9cb.pdf>>
- Lakoff, George and Mark Johnson. 1980. *Metaphors we live by*. Chicago: University of Chicago Press.
- Lang, Mervyn F. 1990. *Spanish word formation*. London: Routledge.
- Lázaro Mora, Fernando. 1993. Compatibilidad entre lexemas nominales y sufijos diminutivos. In Soledad Varela ed. *La formación de palabras*. Madrid: Taurus Universitaria, 303–315.
- Lehmann, Christian. 2002. New reflections on grammaticalization and lexicalization. In Ilse Wischer and Gabriele Diewald eds. *New reflections on grammaticalization*. Amsterdam: John Benjamins, 1–18.
- Lieber, Rochelle and Harald Baayen. 1993. Verbal prefixes in Dutch: a study in lexical conceptual structure. In Geert Booij and Jaap van Marle eds. *Yearbook of morphology*. Dordrecht: Foris, 51–78.
- Lindquist, Hans. 2009. A corpus study of lexicalized formulaic sequences with preposition + *hand*. In Roberta Corrigan, Edith A. Moravcsik, Hamid Ouali and Kathleen M. Wheatley eds. *Formulaic language. Volume I: Distribution and historical change*. Amsterdam: John Benjamins, 239–256.
- Lipka, Leonhard. 1990. *An outline of English lexicology*. Tübingen: Max Niemeyer.
- Mondorf, Britta. 2009. How lexicalization reflected in hyphenation affects variation and word-formation. In Andreas Dufter, Jürg Fleischer and Guido Seiler eds. *Describing and modeling variation in grammar*. Berlin: Mouton de Gruyter, 361–388.
- Montero Curiel, Pilar. 2008. El proceso de lexicalización del diminutivo *ardilla*. *Anuario de Estudios Filológicos* 31: 117–131.
- Moreno Cabrera, Juan C. 1998. On the relationships between grammaticalization and lexicalization. In Anna Giacalone Ramat and Paul J. Hopper eds. *The limits of grammaticalization*. Amsterdam: John Benjamins, 209–227.
- Pinker, Steven. 1999. *Words and rules*. New York: Basic Books.
- Pinker, Steven. 2006. Whatever happened to the past tense debate? UC Santa Cruz: Linguistics Research Center. <<http://escholarship.org/uc/item/0xf9q0n8>>
- Pinker, Steven. 2007. *The stuff of thought*. New York: Viking.
- Pinker, Steven and Michael T. Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Sciences* 6: 456–463.
- Plag, Ingo and Harald Baayen. 2009. Suffix ordering and morphological processing. *Language* 85: 109–152.
- Plag, Ingo, Gero Kunter, Sabine Lappe and Maria Braun. 2008. The role of semantics, argument structure, and lexicalization in compound stress assignment in English. *Language* 84: 760–794.
- Real Academia Española. 2011–2013. *Corpus diacrónico del español*. <<http://www.rae.es>>
- Real Academia Española. 2011–2013. *Corpus de referencia del español actual*. <<http://www.rae.es>>
- Real Academia Española. 2014. *CORPES XXI. Corpus del español del siglo XXI*. <<http://www.rae.es>>
- Scalise, Sergio. 1984. *Generative morphology*. Dordrecht: Foris.
- Stubbs, Michael. 2002. *Words and phrases: corpus studies of lexical semantics*. Malden, MA: Blackwell.
- Talmy, Leonard. 1985. Lexicalization patterns: semantic structure in lexical forms. In Timothy Shopen ed. *Language typology and syntactic description. Vol. 3: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press, 57–149.
- Varela Ortega, Soledad. 1990. *Fundamentos de morfología*. Madrid: Síntesis.
- Valera Ortega, Soledad. 2005. *Morfología léxica: la formación de palabras*. Madrid: Gredos.
- Vannest, Jennifer, Thad A. Polk and Richard Lewis. 2005. Dual-route processing of complex words: new fMRI evidence from derivational suffixation. *Cognitive, Affective & Behavioral Neuroscience* 5: 67–76.