Beyond Frequency: On the Dictionarization of New Words in Spanish

Judit Freixa and Sergi Torner

Institut de Lingüística Aplicada, Universitat Pompeu Fabra

judit.freixa@upf.edu, sergi.torner@upf.edu

Abstract

The most recent literature on neology has discussed the criteria that must be taken into account in order to include new words in dictionaries (Metcalf 2002, Cook 2010, Ishikawa 2006, O'Donovan and O'Neill 2008, Freixa 2016, among many others). Although there are other factors that must be considered, such as morphologic features or semantic transparency (Adelstein and Freixa 2013, Bernal et al. 2018), authors broadly agree that frequency plays a central role, given that high corpus frequency may be taken as evidence of the institutionalization of a lexical unit. However, it has also been pointed out that frequency is a complex criterion in itself, and, therefore, aspects such as stabilization in use (Cook 2010) or a possible longitudinal change in frequency (Metcalf 2002, Ishikawa 2006) must also be taken into account when measuring frequency in corpora.

In this paper, we approach lexical frequency as a criterion to evaluate whether neologisms should be included in Spanish dictionaries from a new perspective. Specifically, we compare data concerning change in frequency of neologisms through time with speakers' perceptions about their novelty, known as 'neological feeling' in the specialized literature (Gardin et al. 1974, Sablayrolles 2003). Data about speakers' perceptions is obtained from online questionnaires carried out within the framework of the Neómetro project (Bernal et al. in press). A set of questionnaires was launched in which 100 subjects evaluated their perceptions of 130 neologisms in Spanish according to four different criteria (correct formation, frequency, novelty and necessity of

inclusion in dictionaries). Frequency data is taken from an extensive corpus of texts from the press, FACTIVA, which provides histograms of frequency through time.

For this study, we analyze the neologisms that were perceived as the most and the least frequent in the questionnaires. We analyze their frequency curve through time in Factiva to find correlations between stabilization in time and speakers' perceptions of their institutionalization. The data allows us to improve the predictive capacity of frequency as a measure to decide which neologisms should be included in dictionaries, as it introduces factors (formal, semantic or usage) that favor or hinder institutionalization into the equation.

Keywords

Spanish, neologism, frequency, histogram, institutionalization

The Dictionarization of New Words

To *dictionarize* a word is to incorporate it into the dictionary, although *dictionarization* usually refers to incorporation into a monolingual reference dictionary (Adelstein and Freixa 2013, Sánchez Manzanares 2013, Bernal et al. 2018). For a word to be included in the dictionary it must be *dictionarizable*, i.e., it must meet the inclusion criteria set by those in charge of compiling or updating the dictionary, and at the same time not fulfill the exclusion criteria. It should be noted that different lexicographic traditions adopt different approaches to the incorporation of new words or neologisms; for instance, Romance lexicography is generally more inclined towards regulating neologisms than English-language lexicography. In any case, words are more or less dictionarizable according to how many dictionarizing criteria are met or not, and *dictionarizability* is the quality of being dictionarizable.

The NEÓMETRO project[1] of Pompeu Fabra University in Barcelona investigates the possibility of measuring the dictionarizability and neologicity of Spanish new words[2]. *Neologicity* is the gradable quality of being more or less neological (Cabré et al. 2004, Auger 2010, Bouzidi 2010, Freixa 2010, Sánchez Manzanares 2013, Estopà 2015, Bernal et al. in press), a quality that the French literature has called *sentiment néologique* since the 1970s (Gardin et al. 1974, Sablayrolles 2003, Ben Hariz 2009). According to the research carried out, the feeling of neologicity is shaped by different parameters that make the speaker perceive a neologism as being more or less neological: within the NEÓMETRO project, Bernal et al. (in press) have already researched the impact of novelty, frequency of use and non-regular formation of the unit, and are currently working on the identification of other parameters such as type of neologism, use in special domains and register of use. It is clear that dictionarizability and neologicity have, at least to some extent, an inversely proportional relationship, although strictly speaking the former is an extrinsic and dependent quality, while the latter is an intrinsic, variable quality.

Reflection on the dictionarization of words is as old as dictionaries; however, the criteria for dictionarization have historically had an applied character and did not appear

---

[2] NEÓMETRO (https://www.upf.edu/web/neometro) works with data from the Observatori de Neologia (https://www.upf.edu/web/obneo) to which we refer in the data analysis.

in the literature as a solid line of research until the 1980s, with several works dealing

with the incorporation of neologisms into English dictionaries (such as Barnhart 1985,

Algeo 1993, Ishikawa 2006 and O'Donovan and O'Neil 2008) and more recently for

Romance lexicography (Freixa et al. 2003, Adelstein and Freixa 2013, Bernal et al.

2018, Freixa in press). In these works a proposal has gradually been developed in which

word-formation, semantic, documentary and pragmatic criteria are combined, with

frequency always having a prominent place[3].

In this study we analyze the relationship between the perception of the

dictionarizability of a number of Spanish neologisms and their real frequency. We also

attempt to describe this relationship with different frequency dimensions derived from

the analysis of histograms and to establish other factors that intervene in speakers'

perceptions of dictionarizability.

Relationship between Dictionarization and Institutionalization

Although, according to Hohenhaus (2005), dictionarization represents the official

institutionalization of a word, the two terms are not synonyms, since dictionaries do not

incorporate all institutionalized words. As we have stated in recent works (Bernal et al.

2018), among the most frequent neologisms in the Spanish press in recent years (and,

therefore, highly institutionalized) we find words that are not being dictionarized[4] for

---

[3] In Freixa (in press) there is a proposal for Catalan (NADIC project,
https://www.upf.edu/web/nadic) that combines and evaluates a total of 20 criteria
grouped into three blocks: linguistic, usage and documentary.

[4] They are not being dictionarized in the normative dictionary of the Real Academia
Española, but some of these words do appear in other general dictionaries with a more
descriptive function.

different reasons. This is the case, for example, for loans such as *selfie*, *celebrity*, *online* and *bullying*, which have been dictionarized under the Spanish forms of *autofoto*, *celebridad*, *en línea* and *acoso escolar*; derivatives based on the names of political figures by means of a suffix, such as *chavismo*, *uribista* or *kirchnerismo*; and compounds based on the colors of sports teams, such as *verdiblanco*, *rojinegro* and *blanquiazul* (*green-and-white*, *red-and-black*, *white-and-blue*).

The many studies on institutionalization, which have been carried out mostly in English, are very useful for the study of the dictionarization of new words in any language. In our work we agree with what Schmid (2008) and to some extent other authors (such as Lipka et al. 1994, Fischer 2014, Štekauer 2002, and Hohenhaus 2005) have asserted, namely that between the first appearance of a neologism (*nonce*) and its establishment in a language, three concomitant phenomena take place: at a structural (linguistic) level, *lexicalization* takes place, that is, the "emergence of specific semantic, orthographic, phonological or syntactic properties of the complex lexeme, which require it to be listed as an entry in its own right in the lexicon of the language" (Schmid 2008, 2); at a sociopragmatic level, its *institutionalization* takes place in the community of speakers when it spreads and becomes generally used; and, finally, from a cognitive perspective there is the *hypostatization* of the concept, understood as the "formation and entrenchment of a concept associated with the word in the minds of the members of a speech community" (Schmid 2008, 2). These three phenomena occur in a process in which three stages can be distinguished: creation, consolidation, and establishment.

For a new word to reach the last stage, establishment, requirements must be met at all three levels. For lexicalization, for example, the word must meet certain formation conditions. Thus, *occasionalisms*, those nonce occurrences in discourse with a marked contextual and co-textual character, which are consciously created to fulfill a mainly

5

expressive (humorous, critical, etc.) function, are not easily repeated in different types of discourse; they are usually creative and formally transgressive formations that tend not to become lexicalized[5]. For a new word to be established in usage, in addition to being – preferably – regularly formed, speakers must be exposed to it until they recognize it and become familiar with it, which leads to the hypostatization of the new concept. When speakers feel the conceptual need for it, its social use increases. This is clearly observed in loan words which somehow remain in the language even though there is already a commonly used word for the same concept. Through semantic or pragmatic redistribution, speakers restrict the space of the existing word and create a defined space for the new unity. In Spanish, for example, the word *resumen* is used to refer to a short text that contains the synthesis of a longer one, but with the arrival of the English loan *abstract*, understood as the short text for a journal or conference, many academics prefer to use *abstract* because it is more specific, while *resumen* keeps the rest of the semantic content. This phenomenon is just one example of semantic specialization demonstrating the strength of hypostatization once a conceptual need has been generated.

Frequency and its Dimensions

One aspect of both the dictionarization and institutionalization of new words on which there is general agreement is that frequency of use is a good indicator (Barnhart 1985, Ishikawa 2006, O'Donovan and O'Neill 2008, Cook 2010, Adelstein and Freixa 2013,

---

[5] There are striking exceptions to this rule, such as *mileurista*, an institutionalized and already dictionarized unit formed by non-prototypical suffixation.

Freixa 2016). Several authors, in fact, consider frequency a decisive factor[6], while others propose using it as a filter prior to the application of the other dictionarization criteria: all neologisms incorporated into a general dictionary must be frequent, but not all frequent neologisms must necessarily enter the dictionary unless they meet other dictionarizability criteria (Freixa 2016). Although total frequency of use is definitely a good indicator, we argue that it does not provide determining information on the degree of institutionalization of a unit for at least two reasons. Firstly, some words are used more frequently than others, as they deal with more commonly discussed subjects, at least in the corpus of press texts usually taken as a reference to measure social use. In the Factiva corpus[7], for example, the word *viernes* (*Friday*) has been documented 6,260,201 times, while *sábado* (*Saturday*) appears only 4,067,266 times (a 35% difference between two words referring to equally institutionalized days of the week). Textual corpora are the best source for measuring and attesting speakers' use, but they should not be taken as an infallible measure in all cases[8]. Second, as has already been

---

[6] Thus, for example, in his analysis of neology in the OED3, Ishikawa (2006, 41) concludes that "the only criteria for getting a word added is if it is being used or not. If it's out there, even if it's vulgar and non-grammatical, it will go in".

[7] https://global.factiva.com.

[8] "No matter how large a corpus is, there may be gaps in its coverage, and a word appearing low ranking in a frequency list can give a false impression of how common a word is. Ultimately, corpus frequency alone cannot be taken as a hard-and-fast indicator of a word's use; the lexicographer must use his/her judgement in analyzing the lists." (O'Donovan and O'Neil 2008, 577-578)

observed by other authors[9], in a unit's total frequency of use different parameters

converge which, separately, may indicate a higher or lower degree of

institutionalization. We therefore support splitting frequency into three different axes:

time, social, and geolectal, because these three axes allow us to measure three different

features related to frequency[10].

On the time axis, the parameter *stability in use* must be analyzed, since

continued frequency over the years indicates a degree of institutionalization greater than

irregular or rising or interrupted frequency (Metcalf 2002, Adelstein and Freixa 2013,

Freixa 2016). On the social axis we analyze the *diversity of communicative situations* in

which a neologism is documented: the degree of institutionalization will be lower if

frequency is concentrated in informal oral productions, or in specialized written texts or

in general journalistic texts, and it will be higher if its use is observed in different types

of texts (Metcalf 2002, Ishikawa 2006, Kerremans 2011). On the geolectal axis, a word

will be considered more institutionalized the greater its *geographical extension*

(Adelstein and Freixa 2013).

---

[9] Metcalf (2002), for example, in his proposed FUDGE (Frequency, Unobtrusiveness,

Diversity of uses and situations, Generation of meanings and forms, Endurance of the

concept) rule for the selection of neologisms, distinguishes total frequency from other

aspects related to frequency. And in Ishikawa (2006) frequency of use, breadth of use,

track record of a minimum of three years and longitudinal changing patterns in frequency

are distinguished.

[10] For a more detailed bibliographic review of the frequency criteria, see Bernal et al.

(2018).

Another way to approach the frequency parameter when studying the institutionalization of words is to examine their histogram, in which changes in frequency on the time axis can be observed. According to Ishikawa (2006), a longitudinal change in frequency indicates that an existing word in use has become a necessary word, but in fact models of curves have not yet been correlated with types or degrees of institutionalization. Grieve and Nini (2016), who have studied *lexical emergence*, have used the *Spearman rank correlation coefficient* to identify *patterns of growth* types and to calculate *lexical emergence* using mathematical modalizations previously applied to population growth studies. Also in our study, as we shall see below, histograms will be analyzed to assess the institutionalization of a number of Spanish neologisms.

Data Analysis

The neologisms in our analysis come from BOBNEO[11], the database of the Observatori de Neologia of Pompeu Fabra University in Barcelona, which stores lexicographical neologisms from 1989 onwards. Our definition of *neologism* is a word which is found in the mainstream press but is not included in our reference dictionaries (*Diccionario de la Lengua Española* [DRAE] and *Diccionario de uso del español de América y España* [VOXUSO]). For our analysis we focus on a set of 132 BOBNEO neologisms selected for their diversity in terms of frequency, date of incorporation and formation type, which were the object of a perception study within the NEÓMETRO project[12]. The data we analyze is found in the 100 answers to the question "Do you think this word should
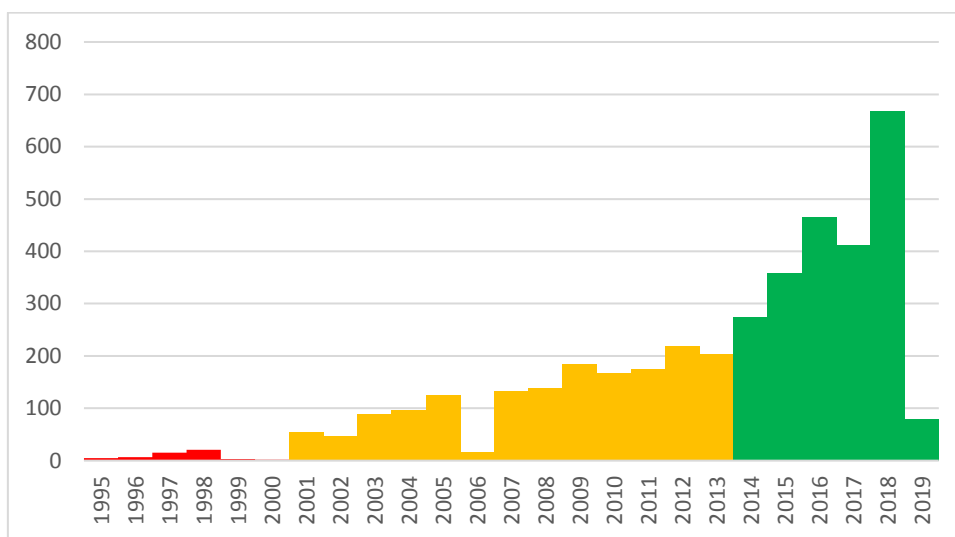
---

[11] http://obneo.iula.upf.edu/bobneo/index.php

[12] For a methodological description of the surveys and a first analysis of the results see Bernal et al. (2018).

appear in a general dictionary?" and are contrasted with information on real frequency from Factiva, the largest corpus of press texts in the world.

Thus, our analysis allows us to contrast objective data of frequency in corpora with the perception that speakers have on the dictionarization of new words. Perception is a datum that does not allow direct interpretation, insofar as speakers do not know the principles that regulate lexicography and therefore lack the technical criteria to make rigorous decisions about the inclusion of words in dictionaries. However, it is indicative of two significant factors in this study. On the one hand, speakers have intuitive, albeit partial, knowledge about the frequency with which a word is used; therefore, a high number of positive evaluations by many speakers is evidence of the institutionalization of a word. On the other hand, as users of dictionaries, speakers have certain expectations about the type of information found in lexicographic works and they know that certain words do not enter dictionaries even though they are highly institutionalized.

Corpus frequency, on the other hand, makes it possible to measure a neologism's degree of institutionalization. In this sense, it should be noted that the use of Factiva as a textual corpus allows us to observe, using histograms, how use evolves in time. Prototypically, the process of institutionalization of a word follows a curve similar to the one we see in the histogram of *rebautizar* (*rename*) in Figure 1.

Figure 1

HISTOGRAM OF *REBAUTIZAR*

As can be seen, there is an initial period (in red) in which the word is used occasionally; this is the creation phase. Then comes a period of gradual consolidation (in yellow), in which there is a tendency to grow, but slowly and with some backward steps. Finally, there is a moment (in green) in which the use of the word is clearly consolidated. It has been institutionalized[13].

In our analysis, then, we contrast the data of objective and subjective frequency, that is, the real presence in a corpus of neologisms and the perception that speakers have of their dictionarization. In most cases, frequency data and subjective data are congruent, and speakers consider that neologisms with histograms showing them to be institutionalized are dictionarizable. However, some discordances occur, either because

---

[13] The sharp decline in 2019 is due to data being taken at the beginning of the year, but it is to be expected that the graph will follow a similar evolution this year. On the other hand, when interpreting histograms, we must bear in mind that the Factiva corpus has been increasing, so that the growth of a neologism must be separated from the global growth experienced by all the units of the corpus.

speakers consider words that do not follow the pattern of prototypical temporal development to be dictionarizable, or because they consider clearly institutionalized words not to be dictionarizable. In these cases, it is necessary to determine what additional factors intervene in speakers' evaluations.

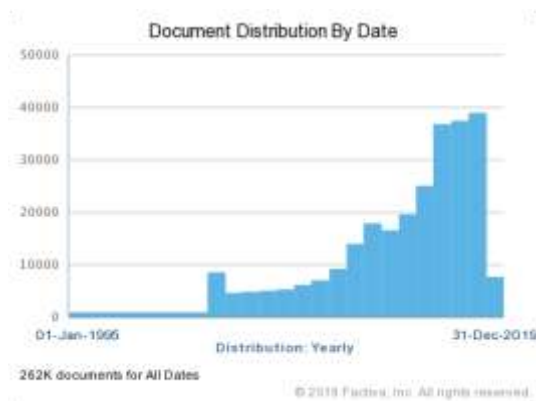Table 1 shows the 24 words of the sample with the highest frequency of appearance in the corpus.

Table 1

Neologism	CG	1st doc Factiva	Frequency in Factiva	% of speakers who acknowledge the neologism as dictionarizable
hoja de ruta	f	1995	261768	80.6
low cost	loc	1999	96259	38.7
carril bici	m	1995	67841	88.17
línea roja	f	1995	54138	72
big data	m	2007	52154	49.4
selfie	f	2013	41126	68.8
internet de las cosas	m	2004	36318	31.1
gap	m	1994	34817	15
intifada	f	1994	30863	74.19
fetiche	adj	1995	27065	81.7
mensaje de texto	m	1997	21310	75.27
biopic	m	1995	19364	43
feeling	m	1995	19228	44.1
economía colaborativa	f	2003	16584	67.7
hackeo	m	1997	16093	71
precuela	f	1998	14626	76.3

relectura	f	1995	14535	63.4
provisionar	v	1994	13916	59.4
top manta	m	2001	13570	68.8
cláusula suelo	f	2009	13480	60.2
techo de cristal	m	1995	13344	76.3
agencia calificadora	f	1995	13086	35.5
café para todos	m	1994	11916	34.4
puntocom	m	1996	11295	30.1

MOST FREQUENT WORDS IN THE CORPUS

In all cases, these words' histograms follow prototypical patterns of institutionalization. This can be seen, for example, in Figure 2, which shows the histogram of *hoja de ruta* 'road map'.
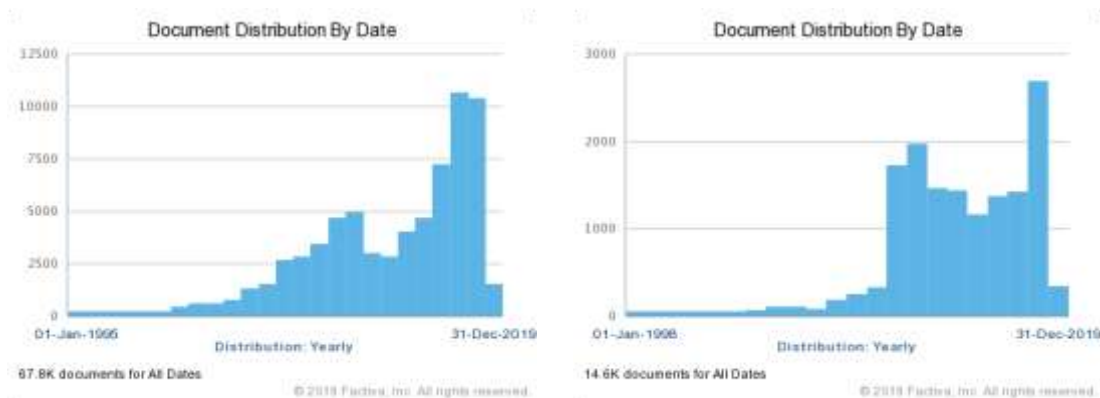
Figure 2

HISTOGRAM OF *HOJA DE RUTA*

In other cases slight variations with respect to the prototypical curve can be observed (see Figure 3): either periods of regression (*carril bici*), or sudden rises at the beginning of the consolidation phase (*precuela*). But consolidation in the use of the word can always be seen at the end of the period with a high frequency.

Figure 3

HISTOGRAMS OF *CARRIL BICI* AND *PRECUELA*

Given the behavior of these 24 words in the corpus, one would expect most speakers to evaluate them as highly dictionarizable words. However, some of these words have been evaluated as not very dictionarizable. Specifically, in 9 cases — shaded in Table 1 — less than 50% of those answering the survey considered that they should be included in the dictionary.

Similar mismatches can be observed if we go down the list following the order of frequency according to Factiva. This is the case with the neologisms listed in Table 2 — in which a column has been added indicating the place of the word in the order of 132 neologisms according to their absolute frequency in Factiva.

Table 2

Neologism	CG	1st doc Factiva	Frequency in Factiva	Order in Factiva	% of speakers who acknowledge the neologism as dictionarizable
superestrella	f	1994	10881	26	49.4
frame	m	1995	7012	31	12.9
enfant terrible	m	1994	6475	32	41
macrobotellón	m	2001	5852	34	30.1
pricing	m	1995	2473	45	15

WORDS WITH HIGH FREQUENCY AND LOW DICTIONARIZABILITY

14

We observe that at the top of the frequency list there are some mismatches between frequency data and speakers' perceptions, while there is also a large number of cases in which both types of data are concordant. A first provisional conclusion, then, is that frequency seems to act on speakers' perceptions as a necessary but not sufficient condition (and also that sometimes speakers are not really aware of the real frequency).

This provisional conclusion is confirmed if we look at the data at the bottom of the frequency list, where we find words that are objectively not very dictionarizable, since their low frequency and their histogram show that they have passed neither the creation phase nor the consolidation phase. This is shown in Table 3.

Table 3

Neologism	CG	1st doc Factiva	Frequency in Factiva	% of speakers who acknowledge the neologism as dictionarizable
proleta	m	1999	103	16.1
noviar	v	1999	97	30.1
infoxicar	v	2011	95	42
storyline	f	1998	92	26.8
democradura	f	1995	85	16.1
polimili	m	1996	84	12.9
governissimo	m	2006	75	6.4
antisectario -ria	adj	1998	72	47.3
desdiabolizar	v	2010	67	26.8
apoyapiés	m	1996	54	72
bolso baguette	m	1999	45	6.4
desarbolamiento	m	1997	43	43
extraocular	adj	1998	41	46.2

culerada	f	2000	37	24.7
hiperbarato -ta	adj	2005	37	7.5
miniensayo	m	1998	36	18.2
abretarros	m	2007	7	44.1
zakouski	m	2003	7	7.5
torturismo	m	2009	6	15
hipermachismo	m	2010	5	25.8
nintenditis	f	2009	5	8.6
turismática	f	2001	3	39.8
maradonear	v	2001	2	7.5
pantafalda	m	2016	1	15

LEAST FREQUENT WORDS IN THE CORPUS

Coherently, in all cases speakers perceive these words as not very dictionarizable, since in no case does the number of affirmative responses in the survey exceed 50%, with the exception of *apoyapiés* (*footrest*), considered dictionarizable by 72% of those surveyed.

Finally, in the intermediate area of the frequency list there are other inconsistencies between objective and subjective frequency in cases in which speakers perceive words with low frequency as dictionarizable, as we see in Table 4.
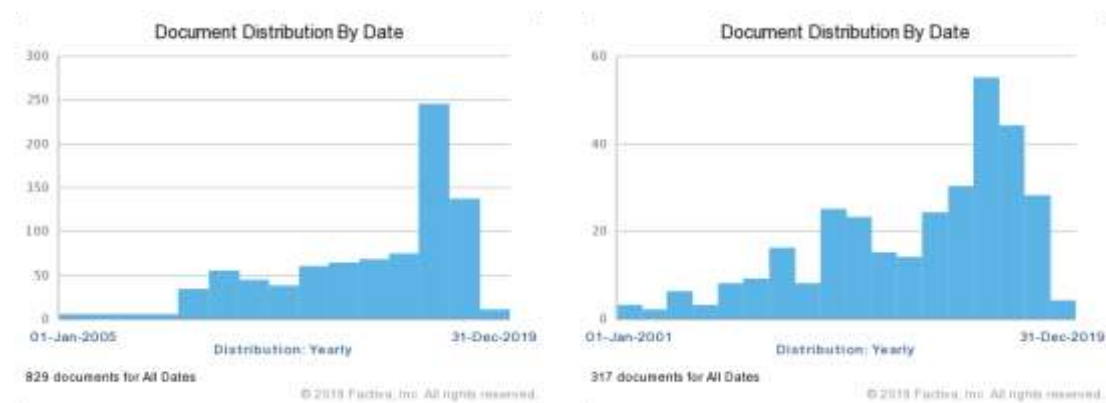
Table 4

Neologism	CG	1st doc Factiva	Frequency in Factiva	Order in Factiva	% of speakers who acknowledge the neologism as dictionarizable
vermutería	f	2006	906	65	74.19
pixelar	v	2005	829	66	89.2
hipersexualización	f	2004	733	67	55.9
psicoestimulación	f	1999	725	68	54.8

infoxicación	f	2002	461	75	51.6
isquiotibial	m	2001	317	78	80.6

WORDS WITH LOW FREQUENCY AND HIGH DICTIONARIZABILITY

All these cases have a relatively low frequency of appearance (less than 1000

occurrences), and they are generally more recent words. The histogram of these

neologisms, on the other hand, does not allow us to determine their institutionalization

precisely because of their low frequency range and the short period of use in which to

observe tendencies. Thus, for example, the histograms of *pixelar* and *isquiotibial* in

Figure 4 do not show a clear trend of longitudinal growth in use, even though there are
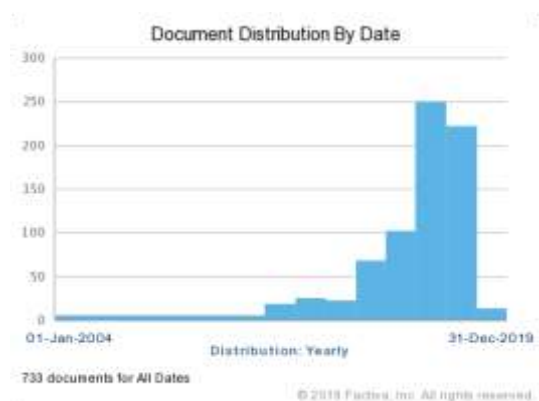
higher frequency peaks.

Figure 4

HISTOGRAMS OF *PIXELAR* AND *ISQUIOTIBIAL*

Similarly, the histogram of *hipersexualización* in Figure 5 shows a remarkable increase

in frequency of use in 2017, but a longer-term perspective would be needed to

determine whether this change will consolidate over time.

Figure 5

Document Distribution By Date

300
250
200
150
100
50
0

01-Jan-2004   Distribution: Yearly   31-Dec-2019

733 documents for All Dates

© 2019 Factiva, Inc. All rights reserved.

HISTOGRAM OF *HIPERSEXUALIZACIÓN*

Nevertheless, all of these neologisms were perceived by speakers to be words that

should be included in dictionaries. Finally, Table 5 presents two cases of words with

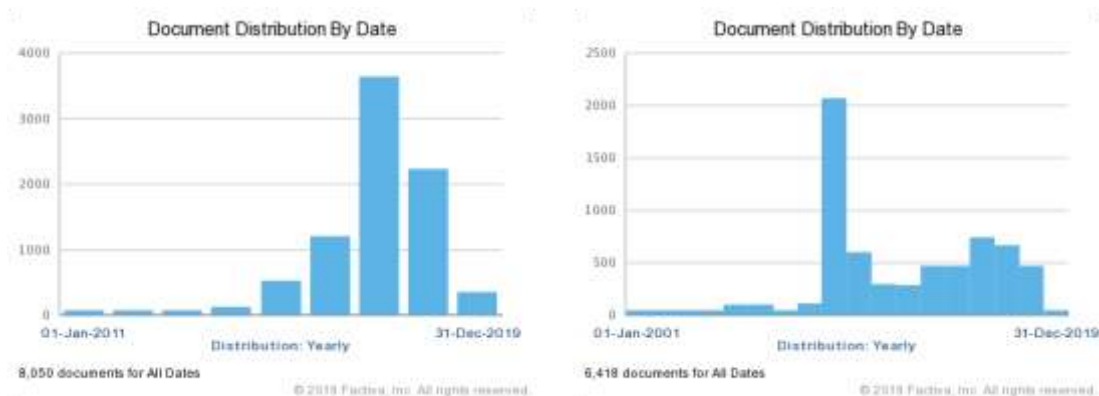irregular frequency behavior but with a high evaluation by speakers.

Table 5

Neologism	CG	1st doc Factiva	Frequency in Factiva	Order in Factiva	% of speakers who acknowledge the neologism as dictionarizable
emoji	m	2011	8050	29	59.1
niqab	m	2001	6418	33	80.5

DATA FOR *EMOJI* AND *NIQAB*

As can be seen in Figure 6, in neither case does the frequency behavior over time follow

a regular pattern.

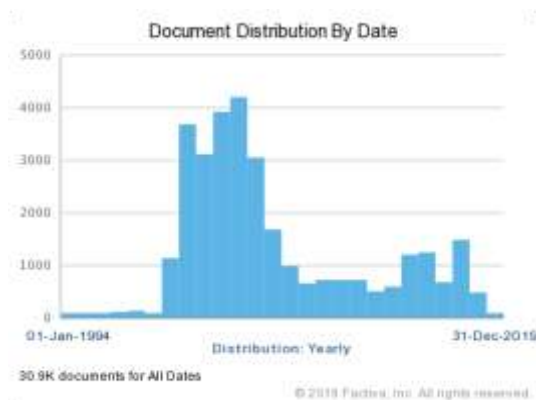**Figure 6**

HISTOGRAMS OF *EMOJI* AND *NIQAB*

All this data shows that in speakers' perceptions, high frequency of use in a neologism is not a sufficient motive for its dictionarization, since not all frequent words are perceived as dictionarizable, nor are all words that speakers consider dictionarizable frequent. It is therefore necessary to look beyond overall frequency in order to understand what criteria have led to such assessments by speakers.

In this sense, it seems that speakers' assessments also take into account how the neologisms are formed and, to a lesser extent, their pragmatic conditions of use. Two groups of examples can be distinguished: in the first place, inconsistency between frequency and perception occurs in loanwords. As a general rule, crude loans are considered by speakers as not very dictionarizable. This is the case of *big data, biopic, enfant terrible, feeling, frame, gap, low cost, pricing*. In addition, some of these words display other features that make them not very dictionarizable for speakers, for example the fact that they are mainly used orally (*feeling*), that they are portmanteau words in the original language (*biopic*) or that there is a Spanish variant for this concept (*bajo precio* for *low cost*). In fact, in the analysis sample, speakers only considered three loans to be dictionarizable: *selfie*, *emoji*, *nikab*. Perhaps the difference between these and the ones mentioned above is not their frequency of use or a denominative need, but above all the

fact that they are internationalisms. If so, it is to be expected that *big data*, still a relatively new term, will end up being perceived by speakers as a necessary word in dictionaries.

This assessment of internationalisms would explain the high perception of dictionarization for *intifada*, which, although it has a very high frequency of use, displays periodic decreases in frequency in the corpus, but with intermediate changes corresponding to the periods of the different intifadas[14], as shown in Figure 7.

Figure 7

HISTOGRAM OF *INTIFADA*

In contrast to this, speakers have no problem considering derivatives formed with a borrowed base as dictionarizable, as long as they have a high frequency of use, such as *hackeo* and *pixelar*.

The second noteworthy group is that of derivatives with a predictable meaning. This includes *superestrella* and *macrobotellón*, very frequent words but with a low valuation by speakers. This case, however, requires a more nuanced reflection, because speakers do not consistently reject words formed with the language's own resources: they consider classical compounds, such as *hipersexualización, psicoestimulación,*

---

[14] The increase in the word after 2000, when the Second Intifada began, and its decrease after 2005, when it officially ended, can be observed very clearly.

infoxicación and *isquiotibial* to be dictionarizable. However, in all of these cases we are dealing with terminological units — albeit sufficiently widespread to be understood by the average speaker — responding to a strong denomination need and used primarily in formal registers. In contrast, *macrobotellón* and *superestrella* are formed with elements of positive intensification, which gives them an expressive rather than denominative character. Therefore, in the process of formation by means of the language's own resources, it seems that speakers value denominative need over frequency, while they tend to value colloquial words negatively[15].

On the other hand, speakers also seem to agree in their assessment of words formed by syntagmation, which they perceive as dictionarizable as long as they have a high frequency of use (*hoja de ruta, carril bici, línea roja, mensaje de texto, economía colaborativa, cláusula suelo*). These neologisms are only rejected when they constitute terms that are little known to the average speaker (*internet de las cosas*), or when they tend to be used in a colloquial register (*café para todos*), regardless of their frequency of use.

Conclusions

The analysis of the histograms of a number of Spanish neologisms shows that frequency of use plays an important role in determining their degree of institutionalization, and this provides, although not directly, information on the need to dictionarize them. The importance of frequency has also been observed in perception surveys, in which the

---

[15] This tendency would explain the high valuation of *vermutería*, a word that fills a denominative gap and is therefore perceived as necessary. The case of *puntocom* is different: speakers value it negatively despite its high frequency, but it does not follow regular word-formation rules.

correlation with objective frequency has become evident. Surveys, moreover, reinforce our understanding of the relationship between dictionarization and institutionalization precisely in cases of discordance, since they show which other factors, in addition to institutionalization in use, are involved in dictionarization. In other words, they allow us to offer an answer to the question of what there is *beyond frequency*.

It should first be noted that frequency of use is a necessary but not a determining condition, as already stated in the literature, and that the perception of a denominative need seems to be a determining factor for dictionarization. This can be observed especially in terminological units (especially classical compounds or units formed by prototypical syntagmatic patterns), which speakers consider highly dictionarizable because, although they may not be familiar with the concept, they notice its existence by necessarily presupposing a hypostatized concept under a terminological form. Similarly, the following kinds of units are not considered highly dictionarizable by speakers: units which are difficult to lexicalize due to the fact that they are formed by means of somewhat irregular mechanisms, and neologisms that are not absolutely necessary in general language (though necessary in spoken language); loans for which a word already exists in the language and units with pragmatic or expressive rather than denominative value are clear examples of this judgment on the need for dictionarization.

In this research we have not been able to observe the specific weight of the different dimensions of frequency, since histogram analysis has only given us access to overall frequency and temporal stability. Geolectal extension and textual diversity have been beyond our scope, and both must be significant and would surely explain some of the divergences between objective and subjective frequency, since the perception of the speakers surveyed refers to the total number of texts where a neologism may have appeared, whereas the corpus analyzed contains only press texts.

Finally, we would like to stress that the results observed are very consistent with what is foreseen in the literature: between the *nonce*, the first appearance, and its definitive establishment as a unit in the linguistic system, a process takes place at a structural, social and cognitive level, and it is only when these three dimensions advance together that the dictionarization of the unit can take place.

References

Base de dades de l'Observatori de Neologia. 1989–. Barcelona: Institut de Lingüística Aplicada – Universitat Pompeu Fabra. http://obneo.iula.upf.edu/bobneo/index.php (accessed March, 10, 2019). [BOBNEO].

Diccionario de la Lengua Española. 2015. 23a. ed. Madrid: Espasa Calpe. [DRAE].

Dow Jones. 1989–. *Factiva*. Nueva York: Down Jones & Company. https://global.factiva.com (accessed March, 10, 2019).

Diccionario de uso del español de América y España. 2003. Barcelona: Spes Editorial SL. [VOXUSO].

Adelstein, Andreina, and Judit Freixa. 2013. Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. Unpublished presentation, Congreso Internacional El Diccionario: neología, lenguaje de especialidad, computación, Ciudad de México (Mexico), 28-30th October 2013. Available at: https://repositori.upf.edu/handle/10230/34891 (accessed March, 10, 2019).

Algeo, John. 1993. Desuetude among new words. *International Journal of Lexicography* 6.4: 281-293.

Auger, Pierre. 2010. Neologicité et extraction néologique automatisée. In *Actas del Congreso Internacional de Neología en las lenguas románicas (CINEO)* edited by Maria Teresa Cabré, Ona Domènech, Rosa Estopà, Judit Freixa and Mercè Lorente, 117-121. Barcelona: Institut de Lingüística Aplicada, Universitat Pompeu Fabra.

Barnhart, David K. 1985. Prizes and Pitfalls of Computerized Searching for New Words for Dictionaries. *Dictionaries* 7: 253-260.

Ben Hariz, Soundous. 2009. Diminuer les fluctuations du sentiment néologique. *Neologica* 3: 37-51.

Bernal, Elisenda, Judit Freixa, and Sergi Torner. 2018. Frecuencia, estabilidad... ¿y después qué? Criterios para la diccionarización de neologismos. Unpublished presentation, VII Congreso Internacional de Lexicografía Hispánica, Valencia (Spain), 28th June 2018.

Bernal, Elisenda, Judit Freixa and Sergi Torner. In press. Néologicité et dictionnairiabilité: deux conditions inverses? *Proceedings of 4ème Congrès international de néologie des langues romanes,* Lyon (France), 4th July 2018.

Bouzidi, Boubakeur. 2010. Néologicité et temporalité dans le processus néologique. *Algérie* 9: 27-36

Cabré, Maria Teresa, Ona Domènech, Rosa Estopà, Judit Freixa, and Elisabet Solé. (2004). La lexicografia i la identificació automatitzada de neologia lèxica. In *De Lexicografia. Actes del I Symposium Internacional de Lexicografia, 16-18 maig de 2002* edited by Paz Battaner and Janet DeCesaris, 287-294. Barcelona: Institut Universitari de Lingüística Aplicada - Universitat Pompeu Fabra.

Cook, C. Paul. 2010. *Exploiting Linguistic Knowledge to Infer Properties of Neologisms* [PhD. dissertation]. Toronto: University of Toronto.

Estopà, Rosa. 2015. Sobre neologismos y neologicidad: reflexiones teóricas con repercusiones metodológicas. In *Neologia das Línguas Românicas* edited by Alves, Ieda Maria and Eliane Simões Pereira. São Paulo: CAPES; Humanitas: 111-150.

Fischer, Roswitha. 2014. Lexical creativity reconsidered. In *The Evolution of Englishes: The Dynamic Model and Beyond* edited by Sarah Buschfeld, Thomas Hoffmann, Magnus Huber, Alexander Kautzsch, 448-469. Amsterdam: John Benjamins.

Freixa, Judit. 2010. La neologicidad en las unidades formadas por prefijación. *Puente* 12: 28-29.

Freixa, Judit. In press. Criteris per a la diccionarització de neologisme. In *Canvi lingüístic, estandardització i identitat en català* edited by Benjamein Meisnitzer, and Josep Martines. Leipzig: Leipziger Universitätsverlag.

Freixa, Judit. 2016. Reagrupamiento lexicográfico de neologismos frecuentes. In *Relaciones morfológicas y diccionario* edited by Elisenda Bernal and Sergi Torner, 95-113. A Coruña: Universidade da Coruña.

Freixa, Judit, Elisabet Solé, and Maria Teresa Cabré. 2003. Neologismos para la actualización de gramáticas y diccionarios. In *Neoloxía e lingua galega: teoría e práctica* edited by Xosé María Gómez and Alexandre Rodríguez, 179-190. Vigo: Universidade de Vigo.

Gardin, Bernard, G. Lefèvre, Christiane Marcellesi and Marie-Françoise Mortureux. 1974. À propos du sentiment néologique. *Langages* 36: 45-52.

Grieve, Jack and Andrea Nini. 2016. Analyzing lexical emergence in Modern American English online. *English Language and Linguistics* 21.1: 99-127.

Hohenhaus, Peter. 2005. Lexicalization and institutionalization. In *Handbook of Word-Formation* edited by Pavol Štekauer and Rochelle Lieber, 353-373. Dordrecht: Springer.

Ishikawa, Shin'schiro. 2006. When a word enters the dictionary: A data-based analysis of neologism. In *English Lexicography in Japan* edited by JACET Society of English Lexicography, 39-52. Bunkyo-ku: Taishukan.

Kerremans, Daphné, Susanne Stegmayr, and Hans Jörg Schmid. 2011. The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. In *Current Methods in Historical Semantics* edited by Kathryn Allan and Justyna Robinson, 59-96. Berlin: Mouton de Gruyter.

Lipka, Leonhard, Susanne Handl, and Wolfgang Falkner. 1994. Lexicalization and institutionalization. In *The Encyclopedia of Language and Linguistics* edited by Ron E Asher, 2164-2167. Oxford: Pergamon.

Metcalf, Allan A. 2002 *Predicting new words: The secrets of their success*. Boston, Mass: Houghton Mifflin Company.

O'Donovan, Ruth and Mary O'Neil. 2008. A systematic approach to the selection of neologisms for inclusion in a large monolingual dictionary. In *Proceedings of the 13th Euralex International Congress*, 571-579. Barcelona: Institut Universitari de Lingüística Aplicada. Universitat Pompeu Fabra; Documenta Universitaria.

Sablayrolles, Jean-François. 2003. Le sentiment néologique. In *L'innovation lexicale* edited by Jean-François Sablayrolles, 279-295. Paris: Honoré Champion.

Sánchez Manzanares, Carmen. 2013. Valor neológico y criterios lexicográficos para la sanción y censura de neologismos en el diccionario general. *Sintagma* 25: 111-125.

Schmid, Hans Jörg. 2008. New words in the mind: Concept-formation and entrenchment of neologisms. *Anglia,* 126.1: 1-36.

Štekauer, Pavol. 2002. On the theory of neologisms and nonce-formations. *Australian Journal of Linguistics* 22: 97-112.