

---

Article

# BEYOND TIMELINES: THE CHALLENGES OF COMBINING THEORETICAL PREMISES AND SPEAKERS' INSIGHTS ABOUT THE ASSESSMENT, VALIDATION AND INCLUSION OF SPANISH NEOLOGISMS IN DICTIONARIES

Miguel Sánchez Ibáñez<sup>1</sup>, Nava Maroto <sup>2</sup>

<sup>1</sup>AteNeo-NeoUsal and Universidad Politécnica de Madrid (miguel.sanchezi@upm.es)

<sup>2</sup>AteNeo-NeoUsal and Universidad Politécnica de Madrid (mariadelanava.maroto@upm.es)

## Abstract

Lexicographic exclusion filtering as a strategy to determine the suitability of neologisms to be included in dictionaries has proved to be insufficient. Therefore, a methodology to rank neologisms in a more objective and quantitative way, based on their most distinctive features as new words is put forward here. In addition, the results of applying this methodology to a sample of more than 3,000 units compiled from Spanish newspapers over a specific time span (2009–2016) are contrasted along with the insights provided by a group of native speakers collected by means of a crowdsourcing strategy. By combining theoretical approaches and speakers' insights, the validity of the suggested methodology is partially corroborated. Moreover, insights given by the speakers involved concerning their prescriptive conception of neology and its place within dictionaries are outlined.

**Key words:** Spanish neologisms, citizen science, neological perception, lexicographical filter

---

## 1. Background and objectives

In this section the theoretical background that supports our study is presented in detail. Section 1.1 is devoted to the departing point of our research: the concept of neologism and how neologisms are identified. Then, section 1.2 explains why the lexicographic filter is considered to be insufficient for neology detection. Finally, section 1.3 focuses on the

relationship of our contribution to the use of citizen science as a methodological approach to involve speakers in linguistic research.

### 1.1. Neologisms and their detection: definition and departing point

Neologisms can be defined as linguistic units which have recently emerged or been imported from other codes, subjected to different and specific circumstances, with morphological, semantic and pragmatic elements closely related to the time they start being used. Some definitions of these units (Guilbert 1973, Guilbert 1975, Rey 1976, and Cabré 1999) point out certain characteristics that, when possessed by a lexical unit, may reveal its neological nature. These units include formal and semantic instability, absence from dictionaries, and the perception users have about them as new or uncommon in some way. Sablayrolles (2000: 149) identifies three types of features shared by all neologisms which help define them more accurately: they are functional (they hold the same distributional potential in the language as any non-neological unit), they increase the stability of a new link between a concept and its designation, and they can be memorized by speakers to the same extent as non-neological units. Beyond these features, Boulanger (2010: 63) focuses on their role as first-stage realizations of the lifecycle of any word. According to him, neologisms are nothing but the initial stage of a lexical unit, just a useful label to track words throughout the chronological axis of their existence. Thus, neologisms are units marked by their instability and novelty, which crystallize in different ways and whose position in a language depends to a great extent on their linguistic, social and political situation, which explains why they are usually bound to raise questions of purism (Ten Hacken & Koliopoulou 2020: 131).

In view of the previous characterizations of neologisms, it seems clear that our challenge is to determine more operative, objective and systematic ways to assess and measure out those apparently vague and arguable features. These strategies have evolved noticeably over the past decades; linguistic and statistical premises have been combined in order not to depend so much on external referents to bring relevant results and to optimize the processes (Drouin 2002: 94). In addition, in the last few years, diachrony has also become a relevant parameter to explore, detect and classify new units, particularly in specialized domains (Temmerman 2000a, Temmerman 2000b, Dury & Picton 2009, Picton 2009). Other methodologies combining elements such as frequency and potential syntactic combinations to detect lexical units (Nazar 2008, Nazar 2011, Nazar 2012, Nazar & Renau, 2012) have also smoothed the path for new methodologies to compile and organize neologisms. These approaches have honed the way in which all those elements must be combined in order to obtain more specific and intuitive results thanks to the design of quantitative language-independent algorithms that, for example, allow units to be selected from a corpus by calculating their distribution and frequency during a time span (Nazar 2011).

### 1.2. Advantages and disadvantages of using a lexicographic exclusion filter

If besides detecting and classifying new units we also want to organize and distinguish them according to their acceptability and relevance for speakers, other elements which move away from the above-mentioned strategies must be considered and we need to focus on neological features, such as formal stability, novelty, or rates of use during a specific time span (Cartier 2017, Falk et al. 2018). In this sense, a criterion that has proved to be remarkably efficient as a complement to statistical approaches is lexicographic exclusion filtering

(‘criterio de exclusión lexicográfica’). [Cabré and Estopà \(2009: 20\)](#) define it as the criterion which allows us to consider a neologism as any word not appearing in a previously defined lexicographic exclusion corpus, that is, a group of dictionaries, glossaries, and other lexicographic tools relevant and related to the set of units to be tested in terms of language, register, field of knowledge, etc.

If we have a list of candidates to be considered as neologisms, the lexicographic exclusion filter will help to cast aside a considerable proportion of non-neological words just by dismissing the units already appearing in that corpus of previously defined lexicographic works. The advantages of this easily automatable filter are substantial, since 75% of common neologisms are lexical, formal and monolexematic, hence the efficiency of this filtering to detect them ([Freixa & Solé 2006: 146](#)).

Nevertheless, there are several factors that limit the efficiency of this strategy. First is the definition of the lexicographic filter to be used, which conditions the quantity and kind of neologisms obtained. In this sense, if the chosen dictionaries are normative, the number of units getting through the filter will be much higher than if those dictionaries have a more descriptive orientation. Secondly, we should not overlook the criteria followed throughout history to determine the exact amount of time that has to go by for a word to become eligible for inclusion in a dictionary or underestimate the great difficulty determining the moment when a new unit is coined ([Sánchez Manzanares 2013: 111](#)). In his introduction to the Grand Larousse dictionary (1971), Guilbert states that the minimal time lapse is ten years as do the guidelines from the Neology Observatory OBNEO (2004) used in our research, whereas [Estornell \(2009\)](#) reduces that period to five years. Moreover, Guerrero Ramos and Pérez Lagos (2003: 347) remind us that the inclusion of a word in a dictionary does not make it lose its neological nature immediately, whereas [Alvar Ezquerro \(1994: 6\)](#) claims that thinking that a unit is neological just because it is not included in a dictionary is simply naïve. To sum up, the presence of neologisms in dictionaries is far from homogeneous ([Castillo Carballo 2003: 97](#)). In the third place, many of the candidates selected by means of this filter differ significantly from what a language user perceives as new words ([Nazar & Vidal 2008: 1](#)). In other words, the lexicographic filter is unable to dismiss a significant amount of noise, which, according to [Freixa and Solé \(2006: 137\)](#) can be of different kinds when applied to journalistic texts. After applying the lexicographic filter, the neology extraction tool:

- Suggests words from other languages included in textual quotes as candidates.
- Selects words with some spelling, typing or spacing mistakes.
- Allows some proper nouns situated at the beginning of sentences to get through, even if the system excludes most of them thanks to their initial capital letter.
- Does not allow for the identification of multiword or semantic neologisms.

In summary, and drawing on our experience, the analysis and classification of neologisms selected by applying only the lexicographic exclusion filtering strategy can be considered a useful starting point to detect neologisms eligible for inclusion in a dictionary, but it soon proves to be insufficient to do so consistently and may lead to scarcely motivated and biased selections of neologisms ([Vega Moreno & Llopert Saumell 2017: 1418](#)).

### 1.3. Neology and citizen science

Besides neological detection, another theoretical framework relevant for this study is that of citizen science (CS) applied to Linguistics, since part of the experiment conducted is

based on this methodological approach. Citizen science is an ever-increasing field of research where lay people play an active and crucial role in research, which may range from short-term data collection to more complex and time-consuming activities such as regularly observing the environment. In Linguistics, these participatory experiments are not new, but the digital era has made citizen collaboration easier to monitor and store. In Lexicography in particular, there is a long well-documented tradition of crowdsourcing, that is, the active collaboration of citizens in the collection of data. A paradigmatic example was the compilation of the *Oxford English Dictionary*, which resorted to volunteers who collected cards that recorded and documented the use of the words to be entered in the dictionary as early as the mid-19th century (Gilliver 2016: 44). In more recent years, another excellent example is the case of Slovenia, where crowdsourcing activities have been seamlessly introduced in the workflow of a lexicographic project for compiling a general dictionary of Slovene (Cibej et al. 2015).

Another field of Linguistics in which crowdsourcing is long established is that of dialect geography, and in particular in the field of dialectometry (Wieling & Nerbonne 2015). Some very interesting projects that have inspired our approach are the SNF-AGORA project carried out at the Zurich Center for Linguistics<sup>1</sup>, whose aim is to set up an online platform where users can identify dialect samples by locating them on a map, or the Milmots project<sup>2</sup>, in which citizens are encouraged to locate Catalan words on a map (Grimaldo et al. 2017).

In spite of this long practical tradition of citizen participation, the truth is that Linguistics, and Lexicography in particular, do not seem to feel comfortable with the label “citizen science”, probably due to the eternal opposition between science and humanities. Nevertheless, citizen collaboration is often labelled as “crowdsourcing” in humanities. This activity is defined by Brabham (2013: xix) as “an online, distributed problem-solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals”. Crowdsourcing is considered by some authors (Bonney et al. 2009, Haklay 2013) as the most basic level of participation in CS, where citizens limit their collaboration to acting as “sensors” that detect and collect data, but without any further intellectual processing. We claim that Linguistics and Lexicography should profit from the experiences in the life sciences in order to develop participatory research and technological collaborative work platforms (Fazio et al. 2016). In our view, citizens may engage at higher levels in lexicographic projects, such as distributed intelligence projects, where participants may not only collect but also interpret the data collected.

Some authors like Rundell (2016) support the potential and opportunity of crowdsourcing in the collection and analysis of linguistic data for their incorporation in dictionaries. He identifies three possible approaches: user-generated content, the wiki model, and crowdsourcing itself. Rundell (2016) suggests that in the field of Lexicography, lexicographers, computers and lay people should contribute with their strengths and weaknesses, but quality control mechanisms and ways of motivating potential contributors need to be implemented in order to ensure the quality of the collaboration.

We believe that citizen science is beneficial, as it strengthens the exchange of information and provides increased access to knowledge (Pettibone & Vohland 2016). Some of the benefits for researchers in Linguistics are that the corpora and datasets used can be larger and that this approach allows for the verification of the practical relevance and applicability of results. From the citizens’ point of view, individuals are invited to actively participate

in the development of useful tools such as dictionaries. However, the fact that citizens participate in the research process implies that scientific standards need to be met. As the Green Paper on Citizen Science for Europe (European Commission 2014) suggests, transparency with regard to data collection methodology and the open discussion of the results are key to guarantee the value of the data collected.

With this background in mind, we have pursued three main objectives in our research.

First of all, to describe a detection filter, complementary to the lexicographic one, that out of a list of candidates, helps determine which neologisms in Spanish are more suitable for inclusion in a dictionary, and to implement them with a meaningful sample of units. In order to do so, some of the most significant neologicity markers, such as formal variation, novelty, usage, and motivation, as well as unstable meaning for speakers, have been taken into account. Secondly, to partially contrast the validity of that new filter by contrasting the results of its application with (a sample of) speakers' insights. Finally, to draw some conclusions regarding the speakers' perception of neology and its role, based on demographic facts such as their age or profession.

The remainder of the article is structured as follows. First, the methodology of the study is presented (section 2), focusing on the scoring scale (2.1) and the online survey carried out (2.2). Section 3 is devoted to the presentation and discussion of the results of the survey; finally, the conclusions of the study and our future steps are outlined (section 4).

## 2. Methodology

This section details the different steps followed in our study. First, the focuses is placed on the description of the scoring scale put forward and applied for the assessment of neologisms in order to decide how suitable and relevant they are for inclusion in a dictionary (section 2.1). Then, in 2.2, the online survey carried out to partially validate the scores obtained, in view of the input and insights given by speakers is outlined.

### 2.1. Design and implementation of a neological scoring scale

The study stems from the work carried out at the observatory of neology of Castile and Leon, a region in Spain<sup>3</sup>. A corpus was established, dating from 2009 to 2016, and 3,209 neologisms appearing in local newspapers<sup>4</sup> during that time span were detected, extracted and classified. The whole process was based on the use of a lexicographic exclusion filter<sup>5</sup>.

However, we came up against the flaws listed by Freixa and Solé (2006) concerning the application of lexicographic exclusion filtering. On the one hand, a variety of non-neological units not included in dictionaries for a wide range of causes were nonetheless identified as candidate neologisms, as they did not meet the lexicographic filtering criteria. On the other hand, and besides the causes listed by Freixa and Solé, at least four other kinds of non-neological units were identified in our research:

- Donyms and geographical terms: in our case, this is the largest group of non-neological units getting through the lexicographic filter. These kinds of words refer to the geographical origins of people from the region where the newspapers analysed are published (*benaventano* –*na* ‘from Benavente’, *alisteño* –*ña* ‘from Aliste’, *terracampino* –*na* ‘from Tierra de Campos’), or to specific geographical entities not commonly used (*valenciano-balear* ‘Valencian-balearic’, *boucense* ‘from Bouzas’), which despite of not being neological, do not appear in the dictionaries used in our research. These words do

not follow a consistent pattern of inclusion and comprise some but not all demonyms and geographical terms.

- Units formed through highly predictable or recurrent procedures: these units, even if novel, were created thanks to the addition of morphological elements whose formal and semantic features are very transparent for speakers, which makes the meaning of the resulting neologism extremely obvious, and in consequence, somehow irrelevant for inclusion in a dictionary. This happens, for instance, with words ending in *-nte* (*sonriente* ‘smiling’, *naciente* ‘nascent’) or starting with *ex-* (*expresidente* ‘former president’, *exmarido* ‘ex-husband’)
- Hyphenated, one-off compounds which rarely persist in language, such as *histórico-biológico* ‘historical-biological’ or *económico-hospitalario* ‘economic and hospital-related’.
- Neologisms whose use declined during the time span defined for the detection process. These also generate some noise during the filtering process because their rate of use is very low and not likely to increase in the future, which questions their eligibility to be included in a dictionary. Examples of these include *sorpaso* ‘overtaking’ or *copago* ‘co-payment’<sup>6</sup>.

To select relevant neologisms to be included in a dictionary, a complementary filter was designed, applicable as a scoring scale which rewarded the neologisms that conform to the main characteristics of prototypical neologisms and penalized those that do not. The three aspects of neologicity rated in the scale were the following:

*Formal consistency:* Given that all our neologisms appeared in written texts, it was useful to analyse their formal variations and features in detail. The instability they often present is part of their process of stabilization and adaptation to the language (Loubier 2011: 10). For instance, regarding neological loanwords, Sánchez Manzaneres (2013: 118) points out that adapted variants of loanwords usually alternate with their original form for some time until the use of one prevails over the other and becomes established in the language. In our case, as Table 1 illustrates, the following factors were taken into account to design this part of the scoring scale: frequency of formal variants (the higher, the better), number of formal variants (the fewer, the better) and difference in frequency of formal variants (the higher, the better for the most used variant). Regarding formal consistency, points were given to each of the 3,209 neologisms as shown in Table 1:

*Motivation:* In this case, the aim was to assess systematically and quantitatively to what extent there were reasons justifying the existence of the neologisms scored. Casado Velarde (2015: 22) states that the appearance of a neologism is rooted in the need to name a new reality, implying that some referential drive is needed. However, he also mentions the “expressive” needs influenced by fashion, culture, or ideologies. In this connection, he refers to the traditional division between giving a name to a new reality which is necessary, and doing so just for stylistic, subjective reasons, which has also been explained by Guilbert (1975) and Sablayrolles (2000, 2003, 2016). This view is especially relevant because neologisms usually coexist with other equally valid units or strategies to refer to the same reality, which might mean that the latest ones could be considered a stylistic resource rather than a referential necessity. This assumption, together with factors such as acceptability and adaptation to the rule, could lead to the conclusion that a neologism is not necessary if its use creates some redundancy, even when there are “stylistic” reasons which explain its emergence.

**Table 1.** Formal consistency scoring system.

Formal consistency	
Criterion	Points
The neologism appears only once in the corpus	0
The neologism appears more than once in the corpus	
With no formal variation	
More than ten times	+4
Between 5 and 9 times	+3
Between 2 and 4 times	+2
With formal variation	
Two variants	
If it is the variant with the highest frequency	
+100% of difference in frequency with the least frequent variant	+2
−100% of difference in frequency with the least frequent variant	+1
If it is the variant with the lowest frequency	
−100% of difference in frequency with the most frequent variant	+1
+100% of difference in frequency with the most frequent variant	0
Three variants	
If it is the variant with the highest frequency	
+100% of difference in frequency with the second variant	+2
−100% of difference in frequency with the second variant	+1
If it is the variant with the second highest frequency	
+100% of difference in frequency with the least frequent variant	+1
−100% of difference in frequency with the least frequent variant	0
If it is the variant with the lowest frequency	0
Four variants	
If it is the variant with the highest frequency	
+100% of difference in frequency with the second variant	+2
−100% of difference in frequency with the second variant	+1
If it is the variant with the second highest frequency	
+100% of difference in frequency with the third variant	+1
−100% of difference in frequency with the third variant	0
If it is the variant with the third highest frequency	
+100% of difference in frequency with the least frequent variant	+1
−100% of difference in frequency with the least frequent variant	0
If it is the variant with the lowest frequency	0

However, this binary classification has been thoroughly contested, as reviewed in Llopart-Saumell and Freixa (2014) and Llopart-Saumell (2016, 2019) because there are no conclusive reasons to think that those two categories are exclusive, and also because its application has turned out to be inconsistent in many cases. For example, some authors like Estornell (2009) or Loubier (2011), state that stylistic neologisms tend to be poorly implanted in the language, something that can be easily refuted: particularly, the cases in which the use of new and unstable words outnumbers the normative variants that can be very easily found in Spanish (Sánchez Ibáñez 2018). Regarding this issue, Llopart-Saumell (2019: 666) concludes that every neologism has a referential motivation, but only some hold additional functions.

To overcome this division, Llopart-Saumell and Freixa (2014) suggested a set of parameters to define the type of motivation underlying a group of neologisms they had collected following the same methodology used in our study: formal neologisms detected in written press by means of the lexicographic filter. Those parameters included speakers' general background, use, stability, novelty, discursive position and discursive context. Taking the insights of Llopart-Saumell (2019) into account, the scoring scale regarding motivation (Table 2) was organized as follows:

- Use: the stronger the presence of the units outside the local newspapers where they were detected, the better. In this case, we took two external Spanish resources as reference: CORPES XXI<sup>7</sup>, a corpus compiled by the Spanish Academy with a holistic, pan-Hispanic approach, and OBNEO<sup>8</sup>, a database set up by the Institut Universitari de Lingüística Aplicada (Pompeu Fabra University, Barcelona, Spain), which focuses on neological units. Thus, we aimed at testing our units against both a wider, referential corpus, and a descriptive database containing neologisms, in order to refine our results.
- Explanatory strategies: the stronger the presence of explanatory strategies in the contexts where the neologisms were detected, helping the readers understand their meanings, the more points neologisms got.

**Table 2.** Motivation scoring system.

Motivation	
Criterion	Points
Use	
If it appears in CORPES XXI	
More than ten times	+2
Less than ten times	+1
If it appears in OBNEO	
More than ten times	+2
Less than ten times	+1
Explanatory strategies	
Defining context	+2
Metalinguistic information	+2
Presence of equivalent expressions	+2
Presence of synonyms	+2
In case of loanwords, presence of equivalents in Spanish	+2



Taking these criteria into account, motivation points were given as explained in [Table 2](#):

*Diachrony*: This concept is at the root of the concept of neology itself, and, therefore, must be considered when defining neologisms ([Boulanger 2010](#): 69). To design this part of the scoring scale, three key concepts were considered: novelty, implantation and relation with current concepts and realities.

*Diachrony*: This concept is at the root of the concept of neology itself, and, therefore, must be considered when defining neologisms ([Boulanger 2010](#): 69). To design this part of the scoring scale, three key concepts were considered: novelty, implantation and relation with current concepts and realities.

- Novelty: we found it necessary to take the idea of neologisms as the first stages of any word ([Boulanger 2010](#): 63) as the basis for deciding how to score our neologisms. We turned again to OBNEO, in this case as the source to reference the dates of appearance for our neologisms. In this sense, the most recent first appearances of our neologisms in OBNEO were rewarded, whereas those which dated back more significantly were penalized.
- Relation with current concepts and realities: it was assumed that the neologisms appearing in the written press would be more relevant if they referred to current concepts even if they were not totally stabilized or were still unstable and with an uncertain survival rate yet to be proved. Consequently, we decided to favour those novel units appearing in our corpus only during the last year of the time span considered for the study, that is, 2016.
- Implantation: in this case, the aim was to measure the trends of use of the neologisms appearing in our corpus during the chosen time span (2009-2016). Our work draws on Quirion's conception of implantation as something "based upon actual vocabulary usage (rather than claimed usage), precisising, quantitative results as to the degree of implantation and yielding a quantitative measure of relative frequency of use (rather than absolute frequency)" (2003: 43). This approach was adopted because our corpus of neologisms was organized according to the years in which neologisms were detected (2009 to 2016), thus facilitating the measurement of their presence in our corpus progressively. We decided to reward those units whose frequency increased over that time span.

As a result, diachrony points were assigned to our 3,209 neologisms as explained in [Table 3](#):

For all three parameters (formal variation, motivation, and diachrony), our scoring scale was based on contrasts among the neologisms of the study themselves, between them and other corpora, or among the different features they may have. [Table 4](#) shows a brief sample of the application of the scoring scale, with the points obtained by 20 units among the 3,209 scored. This varied sample was later presented to the participants in our on-line survey in order to assess whether they would include these words in a dictionary. While some of the words were at the top of the resulting ranking, others scored low, even insignificantly, since they did not meet the neologicity markers assessed by the scale. Likewise, they epitomize a varied range of neological procedures (loanwords, prefixation, classical

**Table 3:** Diachrony scoring system

Diachrony	
Criterion	Points
Novelty	
First appearance in OBNEO	
1989	−2
1990-2000	−1
2001-2005	+1
2006-2010	+2
2011-2016	+3
Relation with current concepts and realities	
The unit only appears in 2016 in our corpus	+1
Implantation	
Decreasing trend (more frequent in 2001-2010)	−1
Increasing trend (more frequent in 2011-2016)	+1

compounding etc.). Approximate translations of the neologisms into English are provided in brackets for those words which are not borrowings from English themselves.

If our premises were correct, and the results of the scoring were totally relevant and valid, the higher a position a neologism attained in our ranking, the more suitable it would be for inclusion in a dictionary. However, this methodology did not take into account one key aspect in the detection and assessment of neologisms: the way speakers perceive them. This perception of neology or *sentiment néologique* (Gardin et al. 1974, Sablayrolles 2003), should be taken as a relevant principle to decide whether a lexical unit is new or not, either from a formal or a semantic point of view (Díaz Hormigo 2007: 41). That is why a step forward was taken by introducing the citizen science approach to the study.

## 2.2. Survey

In order to find out whether the scoring scale was workable, a survey was conducted. An open call was made through different social networks, professional linguistic associations, professional distribution lists, and university students, among others. The survey was designed using Google Forms and was open between 22 March and 8 June 2018. The number of complete answers returned was 946. The overall aim of the survey was to shed some light on the following questions closely related to the neologicity markers measured by the scoring scale:

- What is a neologism?
- What makes a neologism more neological than others?
- Which neological words should be included in a general dictionary?
- How do age and professional profile determine the perception of neologicity?

In order not to condition the participants, these questions were not asked directly. Instead, relevant insights were obtained by providing participants with a set of questions whose answers could be measured using Likert scales so as to find out their views as reliably as possible. The survey was divided into one preliminary section and three main parts.

Table 4: Final score of a sample of neologisms after the application of the scoring scale.

Neologism	Formal consistency	Motivation		Diachrony		Global score
		Explanatory strategies	Use	TOTAL	Novelty Implantation Relation with current concept <sup>a</sup>	TOTAL
like	3	2	4	6	3	4
coworking	2	4	2	6	3	13
benzodiacepina ('benzodiazepine')	2	2	3	5	3	4
wearable	2	2	3	5	3	11
mindfulness	2	4	1	5	3	4
geolocalización ('geo-location')	3	2	2	4	2	11
startup	1	2	4	6	2	3
ciberdelito ('cybercrime')		2	2	4	1	10
sicav ('open-ended collective investment scheme')			4	4	1	8
nomofobia ('nomophobia')	2	2	2	4	1	2
alcolock	2	2	1	3	2	8
start-up	1		4	4	-1	7
eurozona ('eurozone')	2		2	2	2	5
termoarcilla ('honeycumb clay')			1	1	-1	0
nanociencia ('nanoscience')			2	2	3	-2
sobrepesca ('overfishing')			4	4	1	3
micogastronomía ('mushroom culinary art')					-1	3
blended-learning		2		2	3	3
cardsharing		2		2	0	3
biodisponibilidad ('bio-availability')			2	2	0	2
					-1	2
						1

<sup>a</sup>None of the neologisms in this sample got any value for this section. The first neologism in our ranking which scored in this section was *chikunguña*, the name of a virus which spread significantly in 2016 especially in Africa and South America and, therefore, had an important relevance in the journalistic texts published during that year. It obtained a global score of 8 points.

First, a set of demographic details was asked for, such as gender, age, professional status, and the role Spanish plays in the respondents' daily work. After this first section came the three parts of the survey, each one with its specific aim.

*Identification of neologisms:* In the first part, respondents were asked to decide whether a set of words in a text about online music platforms published in the science supplement of the Spanish newspaper *El País* on 17 December 2017 could be considered as neologisms. The text was chosen by the researchers due to the high density of words that could be potentially considered as neologisms by the participants, because of their reference to new realities or their unusual forms due to their recent borrowing from English. Likewise, we thought that if all of them appeared in a consistent context, extracted from a real text which had just been published in a newspaper at the moment the survey was conducted, it would help the participants better understand and take part in this section, making their answers more meaningful and accurate. Thus, a sample of eight words was highlighted in the text: four of them are neologisms according to the theoretical parameters that support our scoring scale ('gigante', 'streaming', 'millennial', 'segmentación'), together with a few distractors, that is, words that are not neological according to theoretical parameters ('online', 'operación', 'desembarcar', 'plataforma'). Participants in the survey had to decide to what extent they considered each of these words to be a neologism. Figure 1 shows the text participants were asked to analyse, with the words they were asked about in bold. An English translation of the text is provided for reference.<sup>9</sup>

*Characterization of neologisms:* The second part of the survey was aimed at identifying the features most used by speakers to define what is a neologism. For that purpose, respondents were given several statements related to some of the most relevant neologicity markers that were also at the root of the scoring scale. Participants had to say to what extent they agreed with each statement, using a Likert scale of five items which ranged from "I totally disagree" to "I totally agree". Table 5 shows the statements included in the survey:

*Inclusion of neologisms in the dictionary:* The third part of the survey aimed to contrast the results directly obtained after having applied our scoring scale against the participants' intuition. For that purpose, the sample of neologisms included in Table 4 was presented to the participants randomly listed, and they were asked to state to what extent they believed it was justified to include each one in a dictionary of neologisms. To do so, a Likert scale was suggested with answers ranging from "I think it is totally necessary to include this word in a dictionary" to "I think this word should by no means be included in a dictionary".

"El **gigante** de la música **online**, Spotify, se ha aliado con su rival chino Tencent gracias a un acuerdo por el que cada compañía adquirirá una participación minoritaria de la otra. Además, la **operación** supondrá un impulso para los dos servicios de música en **streaming** en sus planes de expansión internacional. De este modo, la alianza servirá de punto de apoyo para los planes de Spotify de **desembarcar** en China. La música, hoy en día, tiene mucho que ver con Spotify. Esta **plataforma**, creada en 2008, es actualmente una referencia mundial, con 60 millones de suscriptores de pago el pasado mes de julio, y muchos millones más de modo gratuito. El 72 % de los usuarios son **millennials**, lo que permite una grandísima **segmentación** hacia este público."

**Figure 1.** Text where participants were asked to determine whether the highlighted words were neologisms.

**Table 5.** Neologicity markers and questions asked to assess respondents’ agreement.

Señala hasta qué punto estás de acuerdo con las siguientes informaciones sobre los neologismos. <i>To what extent do you agree with the following statements about neologisms?</i>
1. Son palabras cuyo significado es difícil de deducir sin un contexto que nos lo aclare. <i>They are words whose meaning is difficult to understand without some extra explanation or note</i>
2. Son palabras que ya conoces, pero que de pronto tienen un nuevo significado para ti. <i>They are words you already know but suddenly seem to have adopted a new meaning</i>
3. Son palabras que tienen un uso puntual, en un momento determinado, pero que terminan por desaparecer. <i>They are words used at a very specific period that end up disappearing</i>
4. Son palabras que cada vez se usan más hasta que se ven totalmente incorporadas a la lengua. <i>They are words that are used more and more until they get totally assimilated into the language</i>
5. Son palabras necesarias para denominar nuevas realidades. <i>They are words that are necessary for naming new realities</i>
6. Son palabras que aparecen con frecuencia en cursiva o entrecomilladas en los textos. <i>They are words that are often written in italics or between inverted commas</i>
7. Son palabras que no siempre ves escritas de la misma manera. <i>They are words that you do not see always written in the same way</i>
8. Son palabras estrechamente ligadas a la actualidad y a las modas y tendencias del momento. <i>They are words closely related to current affairs, fashion and trending topics</i>
9. Son palabras que has empezado a oír/leer hace poco tiempo. <i>They are words that you have read or heard for the first time recently</i>

3. Results and discussion

In this section the results of the study are presented and discussed. A rich and diverse set of data was obtained from the survey thanks to the multiple intersections of the demographic variables of the participants (age, gender, profession, etc.), as well as their answers and impressions. In the following sections the most relevant results for each part of the survey are described in detail. First, the demographic data of the participants are detailed in 3.1, followed by an analysis of the three parts of the survey in 3.2, that is: neologism identification (3.2.1), characterization (3.2.2), and, finally, inclusion in a dictionary (3.2.3).

3.1. Preliminary part. Demographic data

This section presents the two aspects of the demographic data collected that yielded the most interesting results, that is, age (Table 6) and professional field (Table 7). Other data, such as gender and the role the Spanish language plays in the respondents’ lives were also collected but did not seem to have a significant impact on their attitude towards neologisms.

Table 6 shows the number of participants in each different age group. About half the participants ranged between the ages of 20 and 40, but a significant number of answers from the other age groups were also collected so as to make up a representative sample. Table 7 shows the participants’ profession. Respondents were grouped according to generic professional fields (Table 7).

Table 7 reflects that a significant proportion of the participants were teachers and students (around 60%). However, other fields, such as translation and interpreting or business

**Table 6.** Number of participants according to age.

Age group	Number of participants
10-20	155
20-30	266
30-40	215
40-50	147
50-60	114
60-100	49

**Table 7.** Number of participants according to profession.

Professional field	Number of participants
Biology and Medicine	35
Science and Engineering	43
Communication	31
Teaching	238
Editing and Proofreading	15
Students	336
Retired and Unemployed	37
Translation and Interpreting	86
Business and Administration	73
Others	54

and administration had a noticeable representation as well. Fewer of them work in the health, scientific and technological sectors, although they are also present in the sample. These data give an idea of the varied range of participants that responded to the survey.

### 3.2. Results of the survey

#### 3.2.1. Neologism identification

Neologism identification. Participants had to decide whether they considered each of the units highlighted in the text to be a *neologism* or not. The key point here was to determine to what extent their views matched our parameters. For us, the units were considered neologisms if they did not appear in the online version of the Diccionario de la Lengua Española (DLE) by the Spanish Royal Academy<sup>10</sup> (with a definition *relating* to the meaning the term has in the suggested text), or if their first occurrence in the OBNEO database dated from 2000 or later. If they did not meet at least one of these two requirements, they were not considered neologisms. Table 8 shows to what extent our criteria match participants' opinions. The cases where opinions diverged are highlighted in grey:

Taken as a whole, the 946 participants' opinions point to an easily defined trend: even if they match our criteria in five out of eight neologisms, they tend to identify loanwords as neologisms even in cases in which the unit has been used in Spanish for several decades. This is the case of the distractor *online*, whose first register in OBNEO dates to 1989 and therefore is no longer considered to be neological according to the chronological criterion. Likewise, semantic neologisms, which have the same form as existing words but have acquired a new meaning, such as *gigante* ('giant' applied to a big company) or *segmentación* ('segmentation' applied to

**Table 8.** Agreement/disagreement in neologism identification.

Words		Neologism for participants?	
		Yes	No
YES	gigante	10%	90% NO
Neologisms according to our criteria	streaming	91% YES	9%
	millennial	95% YES	5%
	segmentación	20%	80% NO
NO	online	81% YES	19%
Distractors (non neological according to our criteria)	operación	5%	95% NO
	desembarcar	10%	90% NO
	plataforma	32%	68% NO

marketing), are overlooked by participants and not considered neologisms. The only case in which a unit of this kind is not considered neological both according to our criteria and to the participants’ opinion is *plataforma* (‘platform’ in the sense of a method of communication or entertainment, for example television, radio, or the internet), and its percentages are not as conclusive as in the other cases, where the unanimity of the results (participants agree or disagree with the consideration of the units as neologisms in proportions above 80% in every case except that one) is also remarkable. Nevertheless, we must not overlook the fact that, despite almost unanimous agreement among participants, the sample of words they were asked about was clearly small and insufficient to draw any definite conclusion and the trends [Table 8](#) shows should be further researched. The choice of those eight words was constrained to the fact that they belonged to one single, consistent and real text, published a few weeks before the survey was conducted, and its relevance derives mainly from its suitability to get participants acquainted with the topic the whole survey was going to be about: their views about neologisms and their inclusion in a dictionary.

3.2.2. *Characterization of neologisms*

Participants were confronted with several statements which define what a neologism might be and were asked to rate their degree of agreement. [Figure 2](#) shows the nine statements followed by a bar in which the five-option Likert-scale answers are displayed with a grayscale ranging from dark grey (‘I totally agree with the statement’) for total agreement to light grey for total disagreement (‘I totally disagree with the statement’). The five options the Likert scale consisted of were the following:

- I totally agree with the statement
- I agree with the statement
- I neither agree nor disagree
- I disagree with the statement
- I totally disagree with the statement

[Figure 2](#) shows the proportion of participants who chose each option from the Likert scale for each statement they were asked about. Statements are classified into categories which correspond to the neologicity markers pointed out in section 1.2, the same that were also taken into account to design the neological score explained in section 2.1, but they

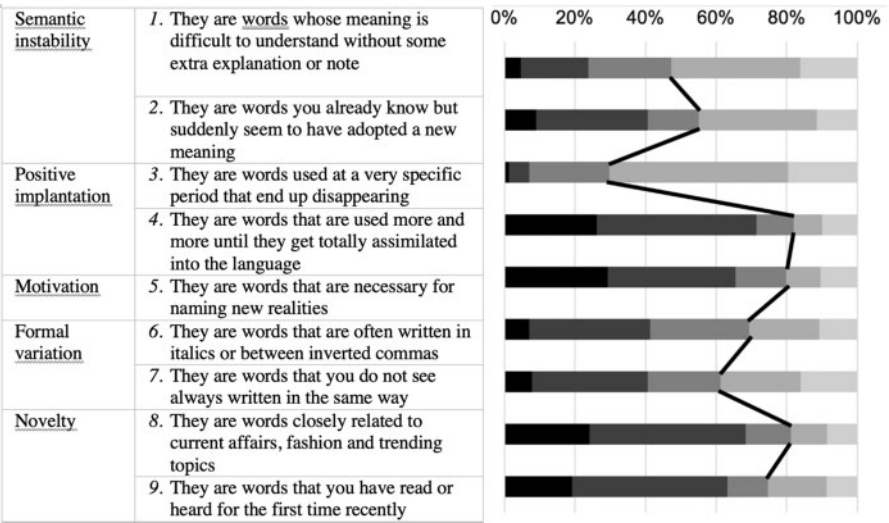


Figure 2: Characterization of neologisms: results.

could have been grouped in many different ways. For instance, statements 2 and 7 could have been classified into the same category, since both of them refer to linguistic changes, (formal and semantic, respectively) but we preferred to categorize them as Table 2 displays because in this way the result can be presented in accordance with our theoretical and methodological standpoints, where formal instability was a relevant factor itself for the creation of our neological score.

A black line divides the ensemble of grey scales into two parts: to the left of the line, we can see the proportion of participants who agree with the statements, whereas to its right we can see the proportion of participants disagreeing with them. Respondents generally agreed (or were indifferent) with our theoretical standpoints regarding positive implantation (question 4), motivation (question 5) and novelty (questions 8 and 9) and, to a lesser extent, semantic instability (questions 1 and 2) and formal variation (questions 6 and 7). The case of question 3 is special: the level of agreement ranks below 40%, but it is the only case in accordance with our standpoint, since it is consistent with what the answers to question 4 display. To sum up, the combination of the results of both questions 3 and 4 confirm that participants consider positive implantation as a defining neological trait, the same as we did in our neological score.

Apart from the information conveyed in Figure 2, several trends were noted when considering other factors that are worth mentioning: when age was considered, older respondents tend to regard neologisms as less necessary. Also, if we take the professional profile into account, quite a surprising result emerges: respondents working in the communication sector, who are generally thought to be natural creators and importers of neologisms, seem to be less prone to consider that neologisms are changing words and closely related to current issues.

3.2.3. Inclusion of neologisms in the dictionary

As for the final part of the survey, whose purpose was to discover which neologisms speakers would most likely include in a dictionary, the following results were drawn.



First, speakers seem more willing to accept or include in a dictionary words made up by composition using lexemes of Latin or Greek origin (such as *geolocalización*, *ciberdelito*, *eurozona*, *nanociencia*). However, loanwords which have not undergone a process of adaptation to Spanish spelling are not so easily accepted by our respondents (*blended-learning*, *wearable*, *cardsharing*, *alcolock*). Figure 3 shows this trend in a grayscale (from dark grey for total agreement to light grey for total disagreement).

As can be seen in Figure 3, the four neologisms which obtain the highest scores, situated on the left-hand side of the graph are either compounds made up by lexemes deriving from Greek or Latin (*geolocalización*, *ciberdelito*, and *nanociencia*) or by lexemes behaving like neoclassical forms (*eurozona*). Therefore, speakers would expect this kind of neologisms to be registered in lexicographic works. At the same time, the four neologisms that scored lower — situated on the right-hand side of the figure (*blended-learning*, *wearable*, *cardsharing*, and *alcolock*) — are all loanwords which have not been adapted to the Spanish spelling at all.

Age seems to be a determining factor when considering the inclusion of loanwords without spelling adaptation. Figure 4 shows the average proportion of participants agreeing that the four loanwords with the lowest scores in Figure 3 according to their age should be included in a dictionary. As in figures 2 and 3, the trend is displayed in a grayscale, (from dark grey for total agreement to light grey for total disagreement):

As can be seen, the younger the participants, the more likely they were to include non-adapted loanwords in a dictionary, whereas older respondents tended to consider their inclusion unnecessary.

If we combine the data from Figure 3 with the score the units obtained in Table 4, we can trace the way theoretical premises considered in our scoring system and participants' perception expressed in the survey interact per each neologism, as illustrated in Figure 5.

Figure 5 displays the same sample of neologisms that can be seen in Table 4, the ones participants were asked about. In this case they are included in a dispersion diagram based on two axes encompassing all the values obtained by these neologisms along our study, ranging from the minimal to the maximal ones they got in each phase: the vertical axis corresponds to the score they obtained in our scoring scale and the horizontal axis corresponds

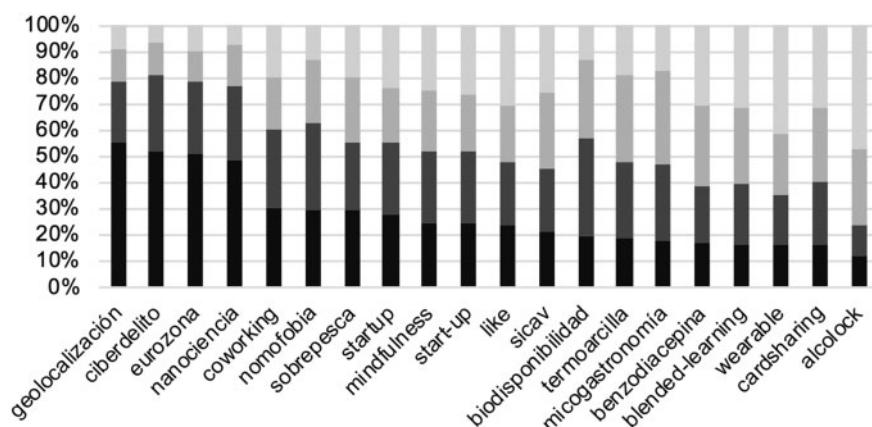
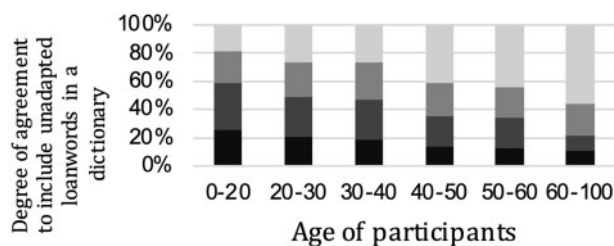


Figure 3. Participants' degree of agreement to include the suggested neologisms in a dictionary.



**Figure 4.** Participants' degree of agreement to include loanwords in the dictionary regarding their age.

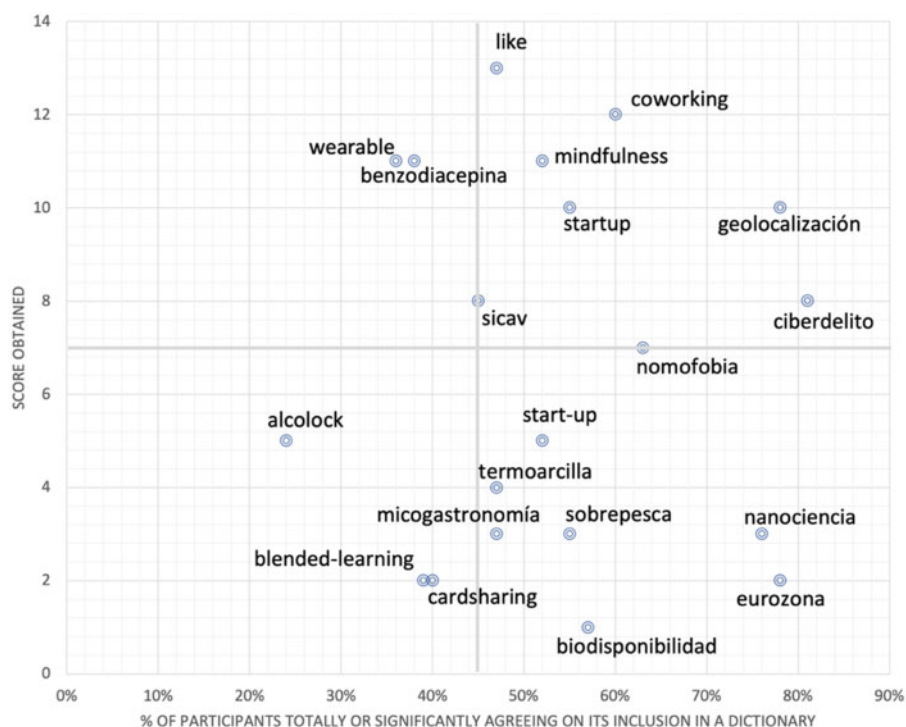
to their level of suitability to be included in a dictionary, according to the views of the participants in our survey. Both axes have been divided into two equal halves, which divides the figure into four quarters that help us better interpret the location of the neologisms in the intersection between both sets of values. In consequence, this division allows us to identify the neologisms placed in the upper-right quarter, the one where neologisms with higher values in both axes can be found, namely *coworking*, *like*, *mindfulness*, *ciberdelito*, *geolocalización*, and *startup*. The way these neologisms have been differentiated from the others thanks to the intersection of the results of the two methodological phases of our study could be a feasible starting point to systematize the integration of both theoretical premises and speakers' intuitions in the selection of Spanish neologisms to be included in dictionaries. The representation of the results with this diagram highlights the contrast among the data obtained, which allows us to consider the weight of both the participants' insights and the theoretical premises in the position of each neologism in the quarters. The case of *startup* and *start-up* -which are situated in different parts of the dispersion diagram also proves this methodology to be useful to rank formal variants and choose in consequence the most relevant one according to our premises.

## 4. Conclusions

The results obtained from this survey allow us to draw several conclusions regarding the participants' perception of neology and to what extent their perception agrees with the results of applying our own scoring system.

Our main objective was to contrast our scoring scale, based on theoretical premises, with speakers' insights; in this respect, and according to the results presented, our scale was mostly validated with respect to formal consistency, motivation, up-to-datedness quality, and implantation. However, the semantic aspects seem not to have struck a chord with the respondents. This can be confirmed firstly by the difficulty participants had in identifying semantic neologisms such as *segmentación* or *gigante*, because 80% and 90% of the participants, respectively, did not recognize those units as neological and showed a tendency to identify loanwords as neologisms even if they have been present in Spanish for a few decades. In this sense, the fact that answers were quite unanimous (i.e., no matter whether they agreed or disagreed with our criteria, participants generally agreed with one another in their answers) should not be overlooked.

In addition, it seems clear that participants are more willing to accept and include in a dictionary those neologisms formed by compounding Latin and Greek lexemes and neoclassical forms, whereas non-adapted loanwords tend to be left out, regardless of their position



**Figure 5.** Dispersion diagram with the score obtained by each neologism and the participants' perceptions on their suitability for inclusion in a dictionary.

in our scoring scale. Compounds such as *geolocalización*, *ciberdelito*, and *nanociencia*, formed with classical roots such as *geo-*, *ciber-*, or *nano-*, and *eurozone*, in which the neo-classical form *euro-* appears, gained a favourable perception from over 80% of the participants concerning their hypothetical inclusion in dictionaries. Moreover, less than 40% of the participants considered *blended-learning*, *wearable*, *cardsharing*, and *alcolock* as eligible for inclusion in a dictionary. It is also interesting to point out that these four units are all loanwords imported from English, with a very low degree of formal and orthographical adaptation to Spanish. Their low rates contrast with the increasing weight English loanwords have in the updating of Spanish (Gerding et al. 2014), something which connects with the global figures of our study: 704 out of the 3,209 units analysed and scored were loanwords (22%), and, of them, 470 (67%) had an English origin. The remaining 33% correspond to more than ten other languages.

The combination of these impressions with the scores obtained by those units and included in Table 4 allows us to find a compromise between theoretical, preliminary premises, and participants' theoretical conception about neologisms, and participants' impressions about those same neologisms. This gives us a balanced final disposition of units regarding both lexicographic suitability and neological relevance of the units. Furthermore, the survey reveals that the participants' perception of the motivation of neologisms and the suitability of the units to be included in a dictionary is clearly conditioned by their age: 60% of participants under the age of 20 agree to include loanwords in

dictionaries, in contrast with a mere 20% of participants over 60 with the same opinion. In other words, older participants are less inclined to consider neologisms necessary or deserving of inclusion in a dictionary. It has also been noticed that what participants consider to be a neologism in theory does not correspond with their perception of specific neologisms in practice. As just mentioned in the case of *online*, participants tend to identify non-adapted loanwords as neologisms even when they are long established in Spanish. However, when asked in general terms, 60% to 70% agree when it comes to defining neologisms as units related to current times and novelty as shown in the answers to questions 4, 5, 8, and 9 in [Figure 2](#).

All these conclusions lead us to a final revealing insight: according to the results culled from our experiment, we can venture to say that participants adopt a prescriptive position towards dictionaries. As [Ten Hacken and Koliopoulou \(2020: 128\)](#) point out, whereas lexicographers tend to take dictionaries as inventories of word meanings and facts about language, the general public considers them as an authority on what constitutes acceptable or correct language. In other words, they consider lexicographic resources as lexical validators or sanctioning authorities that confirm which units have sufficient legitimation to be used, rather than tools to learn about the meanings and uses of new words. Units created within Spanish and/or based on classical roots are given much more prestige than those borrowed from other codes. In consequence, novelty tends to be identified with otherness ([Irimias 2011](#)) and, ultimately, with unacceptability. The fact that participants agree with the characterization of neologisms in theoretical terms but fail to apply this characterization to specific units could imply the existence of reasons for this rejection that go beyond language itself and generate a prescriptive consideration of the links between neology and lexicography.

As for the fact that we rely on speakers as a valuable source of data, the favourable reception of our survey among different groups of participants encourages us to consider that citizen science could be a valuable method for working in the field of neology detection and validation. Different applications could be envisaged to allow for the participation of speakers in the collection and evaluation of neologisms, taking advantage of the digitalization of society and the undeniable interest shown by speakers in matters related to the appearance of new words in a language.

Our research has also allowed us to identify some limitations that need to be addressed in our future work. For example, we realized that it is vital to establish and maintain transparent communication channels between researchers and collaborators. In order to improve our communication with the respondents, a website is being developed that will constitute an open channel to disseminate our results and attract volunteers to participate in our CS initiatives. Another aspect where speakers' opinions could and should be heard and considered is that of dictionary design. A further survey to determine the expectations of speakers on what kind of information a dictionary of neologisms should contain will be set up so as to bear in mind speakers' needs in the design of lexicographic products.

In an era in which the compilation of dictionaries has evolved dramatically due to the advent of technological applications, these should be developed in a way that speakers are involved both in collecting and validating linguistic data.

## Notes

1. Zurich Center for Linguistics. 2020. *SNF-AGORA Citizen Linguistics*. <https://www.linguistik.uzh.ch/en/forschung/agora.html>. Accessed on 10 February 2020.
2. Grimaldo, F. et al. 2015. *Milmots. Un món de paraulas*. <http://milmots.eu/>. Accessed on 10 February 2020.
3. This research has been carried out in the frame of the funded project *Neogenes: el préstamo léxico como mecanismo actualizador del español general y especializado* (Junta de Castilla y León, REF.: SA059G19).
4. El Norte de Castilla, El Día de Valladolid, El Diario Palentino, Diario de León. Diario de Burgos, Adelantado de Segovia, La Opinión de Zamora, La Gaceta de Salamanca, Diario de Ávila, El Mundo (Castilla y León Edition).
5. The lexicographic filter was composed of the indexes of entries of two dictionaries: a prescriptive one (Real Academia Española 2014) and a descriptive one (Battaner et al. 2001).
6. ‘Sorpaso’ is a loanword from Italian which was quite frequent a few years ago in newspapers to refer to new political parties overtaking traditional ones, whereas ‘copago’ referred to copayment of medical services, and was also very popular at a certain moment. Both words had a surge for some time, but then ceased to appear in newspapers.
7. Real Academia Española. 2020. *Corpus del Español del Siglo XXI (CORPES)*. <http://web.frl.es/CORPES/view/inicioExterno.view;jsessionid=BDCBF6B732F76A71A8E8FBE01AAE69D9>. Accessed on 23 April 2021.
8. IULA. 2020. *Obneo*. <https://www.upf.edu/es/web/obneo/recursos>. Accessed on 23 April 2021.
9. *The online music giant, Spotify, has allied itself with Chinese rival Tencent through an agreement whereby each company will acquire a minority stake in the other. The deal will also give both streaming music services a boost in their international expansion plans. The alliance will thus serve as a foothold for Spotify's plans to land in China. Music today has a lot to do with Spotify. The platform, created in 2008, is now a global benchmark, with 60 million paid subscribers last July, and many millions more for free. 72% of the users are millennials, which allows for a very large segmentation towards this public.*
10. Real Academia Española. 2019. *Real Academia Española*. [www.rae.es](http://www.rae.es). Accessed on 23 April 2021.

## References

### A. Dictionaries

- Alvar Ezquerro, M. 1994. *Diccionario de voces de uso actual*. Madrid: Arco.
- Battaner Arias, P. (dir.). 2001. *Lema. Diccionario de Lengua Española VOX*. Barcelona: Spes.
- Real Academia Española. 2014. *Diccionario de la Lengua Española (23rd edition)*. Madrid: Espasa. (DLE)
- Stevenson, A. (ed.). 2010. *Oxford Dictionary of English*. Oxford: Oxford University Press.
- Larousse, P. 1973. *Grand Larousse Encyclopédique*. Paris: Larousse.

## B. Other literature

- Bonney, R., C.B. Cooper, J. Dickinson, S. Kelling, T. Phillips, K.V. Rosenberg, and J. Shirk. 2009. 'Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy.' *BioScience* 59.11: 977–984. <https://doi.org/10.1525/bio.2009.59.11.9>
- Boulanger, J.-C. 2010. 'Sur l'existence des concepts de "néologie" et de "néologisme" Propos sur un paradoxe lexical et historique.' In Cabré, M.T., O. Domènech, R. Estopà, J. Freixa, and M. Lorente (coords), *Actas del I Congreso Internacional de Neología en las lenguas románicas*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, 31–74.
- Brabham, D. C. 2013. *Crowdsourcing*. Cambridge: MIT press.
- Cabré, M. T. 1999. *La terminología. Representación y comunicación. Una teoría de base comunicativa y otros artículos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Cabré, M. T. and R. Estopà (eds). 2009. *Les paraules noves: Criteris per detectar i mesurar els neologismes*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Cartier, E. 2017. 'Neoville, a web platform for neologism tracking.' In Martins, A. and A. Peñas (eds), *Proceedings of the Software Demonstrations of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Valencia: Association for Computational Linguistics, 95–98.
- Casado Velarde, M. 2015. *La innovación léxica en el español actual*. Madrid: Síntesis.
- Čibej, J., D. Fišer, and I. Kosem. 2015. 'The role of crowdsourcing in Lexicography.' In Kosem, I., M. Jakubiček, J. Kallas, and S. Krek (eds), *Proceedings of the eLex 2015 conference 'Electronic Lexicography in the 21st century: Linking lexical data in the digital age'*. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd., 70–83.
- Díaz Hormigo, T. M. 2007. 'Aproximación lingüística a la neología léxica.' In In Martín Camacho, J.C. and M.I. Rodríguez (eds). *Morfología: Investigación, docencia, aplicaciones. Actas del II Encuentro de Morfología: Investigación y Docencia*. Cáceres: Universidad de Extremadura, 33–54.
- Drouin, P. 2002. *Acquisition automatique des termes : l'utilisation des pivots lexicaux spécialisés*. Montréal: Presses Universitaires de Montréal.
- Dury, P. and A. Picton. 2009. 'Terminologie et diachronie : vers une réconciliation théorique et méthodologique ?' *Revue Française de Linguistique Appliquée* 14.2: 31- 41.
- Estornell, M. 2009. El reconocimiento de neologismos y su caracterización en un corpus de prensa escrita 2004-2007. Doctoral thesis. Valencia: Universidad de Valencia.
- European Commission. 2014. Green Paper on Citizen Science. Towards a Society of Empowered Citizens and Enhanced Research. <https://ec.europa.eu/digital-single-market/en/news/green-paper-citizen-science-europe-towards-society-empowered-citizens-and-enhanced-research>. Accessed on 23 April 2021.
- Falk, I., D. Bernhard, and C. Gérard. 2018. 'The Logoscope: A Semi-Automatic Tool for Detecting and Documenting French New Words.' *arXiv preprint arXiv:1810.10797*. Accessed on 10 February 2020. <https://arxiv.org/pdf/1810.10797.pdf>. Accessed on 23 April 2021.
- Fazio, M., A. Fresa, and B. Justrell. 2016. Civic Epistemologies. The Roadmap. Accessed on 10 February 2020. . [https://www.civic-epistemologies.eu/wp-content/uploads/2014/08/CE\\_Roadmap-Handbook.pdf](https://www.civic-epistemologies.eu/wp-content/uploads/2014/08/CE_Roadmap-Handbook.pdf). Accessed on 23 April 2021.
- Freixa, J. and E. Solé. 2006. 'Análisis lingüístico de la detección automática de neologismos léxicos.' *Sendebat* 17: 135–147.
- Gardin, B, T. Michel, and M. F. Mortureux. 1974. 'A propos du "sentiment néologique".' *Langages* 8.36: 45–52.

- Gerding-Salas, C., M. Fuentes, L. Gómez, and G. Kotz. 2014. 'Anglicism: An active word-formation mechanism in Spanish.' *Colombian Applied Linguistics Journal* 16.1:40–54.
- Gilliver, P. 2016. *The making of the Oxford English Dictionary*. Oxford: OUP.
- Grimaldo, F., C. Sánchez Curto, E. López Iñesta, J.E. Vicente i Girona, and E. Querol Puig. 2017. 'Ciència ciutadana contra prejudicis lingüístics: el projecte Milmots.eu.' *Llengua, societat i Comunicació* 15: 48–53.
- Guilbert, L. 1973. 'Théorie du néologisme'. *Cahiers de l'Association internationale des études françaises* 25: 9–29.
- Guilbert, L. 1975. *La créativité lexicale*. Paris: Larousse.
- Irimias, G. 2011. 'Breaking the Linguistic Norm in Mass Media. The limits of Neology.' *Acta Universitatis Sapientiae, Philologica* 3.2: 175–183.
- Llopart-Saumell, E. 2019. 'Neologisms from a functional perspective: Correlation between perception and empirical data.' *Signos Signos* 52.100: 665–687.
- Llopart-Saumell, E. 2016. La funció dels neologismes: revisió de la dicotomia *neologisme denominatiu* i *neologisme estilístic*. Doctoral Thesis. Barcelona: Universitat Pompeu Fabra.
- Llopart-Saumell, E. and J. Freixa. 2014. 'La función de los neologismos: revisión de la dicotomía *neología denominativa* y *neología estilística*.' *Neologica* 8: 135–15.
- Loubier, C. 2011. *De l'usage de l'emprunt linguistique*, Montreal: Office Québécoise de la Langue Francaise.
- Medina Guerra, A. (ed.). 2003. *Lexicografía española*. Barcelona: Ariel.
- Nazar, R. 2008. 'Bilingual Terminology Acquisition from Unrelated Corpora.' In Bernal, E. and J. deCesaris (eds), *Proceedings of the XIII EURALEX International Congress* (Barcelona, 15-19 July 2008). Barcelona: Universitat Pompeu Fabra, 1023–1030.
- Nazar, R. 2011. 'Estudio diacrónico de la terminología especializada utilizando métodos cuantitativos: Ejemplos de aplicación a un corpus de artículos de lingüística aplicada.' *Signos* 44.75: 48–77.
- Nazar, R. 2012. 'Bifid: un alineador de corpus paralelo a nivel de documento, oración y vocabulario.' *Linguamática* 4.2: 45–56.
- Nazar, R. and I. Renau. 2012. 'Agrupación semántica de sustantivos basada en similitud distribucional. Implicaciones lexicográficas.' In *Actas del V Congreso Internacional de Lexicografía Hispánica*. Madrid: Universidad Carlos III.
- Nerbonne, J. (2015). 'Advances in dialectometry.' *Annual Review of Linguistics* 1: 243–264.
- OBNEO (Observatori de Neologia). 2004. *Metodología del trabajo en neología: criterios, materiales y procesos*. Barcelona: Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra.
- Pettibone, L. and K. Vohland. 2016. *Citizen Science for All: A Guide for Citizen Science Practitioners*. Berlin: Burger Schaffen Wissen (GEWISS).
- Picton, A. 2009. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus spécialisés. Un exemple appliqué au domaine spatial. Doctoral Thesis. Toulouse: Université Toulouse-Le-Mirail.
- Quirion, J. 2003. 'Methodology for the Design of a Standard Research Protocol for Measuring Terminology Usage.' *Terminology* 9.1: 29–49.
- Rey, A. 1976. 'Néologisme: un pseudo-concept?' *Cahiers de Lexicologie: Revue Internationale de Lexicologie et Lexicographie* 28: 3–17.
- Rundell, M. 2016. 'Dictionaries and crowdsourcing, wikis and user-generated content.' In Hanks, P. and G.M. de Schryver (eds), *International Handbook of Modern Lexis and Lexicography*, Berlin: Springer-Verlag, 1–16.
- Sablayrolles, J. F. 2000. *La néologie en français contemporain. Examen du concept et analyse des productions néologiques récentes*. Paris: Honoré Champion.
- Sablayrolles, J. F. 2003. *L'innovation lexicale*. Paris: Honoré Champion.

- Sablayrolles, J. F. 2016. 'Néologismes et idées reçues'. In García Palacios, J., G. De Sterck, D. Linder, N. Maroto, M. Sánchez Ibáñez, and J. Torres del Rey (eds). *La neología en las lenguas románicas*, Frankfurt: Peter Lang, 43–56.
- Sánchez Ibáñez, M. 2018. 'Definiendo en positivo los neologismos formales: Hacia un análisis cuantitativo de la correlación entre sus características' *Pragmalingüística* 26: 349–372.
- Sánchez Manzanares, C. 2013. 'Valor neológico y criterios lexicográficos para la sanción y censura de neologismos en el diccionario general.' *Sintagma: revista de lingüística* 25: 111–125.
- Temmerman, R. 2000a. *Towards New Ways of Terminology Description: the Sociocognitive-Approach*. Amsterdam, Philadelphia: John Benjamins Publishing Company.
- Temmerman, R. 2000b. 'Une théorie réaliste de la terminologie : le sociocognitivism'. *Terminologies nouvelles* 21: 58–64.
- Ten Hacken, P. and M. Koliopoulou. 2020. 'Dictionaries, neologisms and linguistic purism.' *International Journal of Lexicography* 33.2: 127–134.
- Vega Moreno, É. and E. Llopart-Saumell, 2017. 'Delimitación de los conceptos de novedad y neologicidad'. *Rilce. Revista de Filología Hispánica* 33.3: 1416–1451.
- Wieling, M. and J. Nerbonne, 2015. 'Advances in dialectometry', *Annual Review of Linguistics* 1: 243–264.