

Pedro J. Bueno, Judit Freixa

Lexicographic detection and representation of Spanish neologisms in the COVID-19 pandemic

1 Introduction

The syntagma *gel hidroalcohólico* ‘hydroalcoholic gel’ or the noun *hidroalcohol* ‘hydroalcohol’ cannot be found in *Diccionario de la lengua española* (DLE) of the Real Academia Española (‘Royal Spanish Academy’) or other general reference dictionaries of the Spanish language. This is so despite the fact that, for well over a year and to this very day, we have not been able to do anything without first sanitising our hands with this product. It is one of the many neologisms that the COVID-19 pandemic has brought us, and these have become commonly used words that dictionaries should consider as candidates for future updates.

By looking at the dictionarisability of these neologisms, in this work we try to set their boundaries on the continuum along which they fall. “Dictionarisability” means, in our context, the greater or lesser interest of these unities regarding the updating of general language dictionaries. At both ends of this continuum, there are surprising nonce words, as well as neologisms that have recently lost their status as such because they have now been incorporated into the dictionary. To identify different groups on the continuum of pandemic neologisms, we take into account the criteria proposed in the current literature and, by so doing, we are able to assess the extent to which they are discriminatory. This will allow us to address the neological process and to reflect on the various stages of it, from the time a neologism is born until the moment it ceases to be one because it has been dictionarised.

Before that, however, we present the framework of our study and refer to the mechanisms available for detecting neologisms in general and pandemic neologisms in particular.

Note: This article was prepared with the help of the LEXICAL project “Neología y diccionario: análisis para la actualización lexicográfica del español” of the Ministry of Economy and Competitiveness (ref. PID2020-118954RB-I00), funded by the State Research Agency (AEI) and the European Regional Development Fund (ERDF).

Pedro J. Bueno, Universitat Pompeu Fabra, Barcelona, Spain, email: pedrojavier.bueno@upf.edu
Judit Freixa, Universitat Pompeu Fabra, Barcelona, Spain, email: judit.freixa@upf.edu

2 Study framework

2.1 Defining what we understand by *pandemic neologism*

In most of the neology literature in Spanish, a neologism is considered to be a new word, either formally or semantically, or taken from another language.¹ However, we will use the following definition of a neologism: “A recent word that is in the process of becoming established in a language”² (Freixa, in press), so not all recent words are neologisms unless first signs of their common use by speakers can be noted in corpus data.

In this case, the words that we consider ‘recent’ are the following: a) all the terms related to the COVID-19 pandemic that first appeared between January 2020 and June 2021, and b) those that had appeared earlier but have experienced a big increase in use during the pandemic.

2.2 Detection of *pandemic neologisms*

Novelty is the main characteristic of neologisms and, since novelty is a perceptively subjective quality, a methodological criterion must be established to obtain data objectively. This criterion will necessarily be separate from the theoretical understanding of the concept of neologism. Moreover, it will always be an unsatisfactory one because we will be trying to square the circle. Assuming these limitations, the most reliable criterion for the detection of neologisms will be the comparison of analysis texts (necessarily current, these texts are the ones from which neologisms are expected to be extracted) with an exclusion corpus that must be capable of being deemed representative of the language. Ideally, this corpus should be a balanced body of texts in terms of discursive genres, themes and linguistic varieties, and it should include historical and current language. Thus, all the lexical units documented in the current texts that do not appear in the corpus deemed representative of the language may be considered new.

However, most neology observatories around the world do not have such an ideal corpus or the equipment to exploit it, so an exclusion corpus usually employs a lexicographic corpus composed of one or more dictionaries deemed representative of the language on which work is being done. In the case of Spanish neology

¹ The origin of this definition can be found in the early authors of French lexicology, who faced the challenge of defining such an undefinable concept, such as Matoré (1952), Guilbert (1975) and Rey (1976).

² Our definition is clearly inspired by Hohenhaus (2007: 18) who argues that neologisms are “words that are ‘young’, diachronically speaking, but which nevertheless have already entered the language as more or less institutionalised vocabulary items”.

observatories, this method is used, and every unit from the analysed text not found in the exclusion corpus formed by DLE or other general reference dictionaries of the Spanish language are regarded as neologisms.

When the criterion for the detection of neologisms is determined in this way, it is called the *lexicographic criterion (for the detection of neologisms)*. Criticism can easily be levelled at it (and it is widely criticised, indeed) because it does not discriminate neologisms from other words not found in the dictionaries for other reasons. As discussed in Bueno/Freixa (2020), by using the lexicographic criterion, what we actually get are *lexicographic neologisms*, some of which are true *neologisms* while others are *pseudoneologisms*. The following are considered as pseudoneologisms: a) morphologically regular and semantically transparent non-new words, whose meanings can be deduced from words and/or elements already found in dictionaries (this is the reason why dictionaries are reluctant to accept them); b) specialised lexical units (terms) that are already in the corresponding terminology dictionaries, whose novelty is simply the fact that they have entered general use; c) colloquialisms, non-recent units that dictionaries do not systematically include; d) old and new, general and specialised, frequent and occasional loanwords that, due to language policy criteria, dictionaries restrictively select for their lists of words; e) words bearing witness to an era and a place that are generally not likely to have a long course to run in society; f) localisms and dialectalisms that, again, dictionaries do not systematically include because of their lack of general use; g) nonce words, which appear for reasons that are more expressive than denominative, have a strong playful component and not necessarily with the object that they become part of general language; and h) variants, errors and other non-new units that are not found in dictionaries for various reasons and, by applying the lexicographic criterion, also become *pseudoneologisms*.

However, neology observatories are led by linguists who are well aware of these shortcomings and therefore filter neologisms by the type of research that is intended to be carried out. To do this, all lexicographic neologisms are accompanied by different pieces of information relating to linguistics (type of neologism, grammatical category, etc.), use (type of text, context, linguistic markers, frequency, etc.) and documents (relationship to words already documented, presence in other dictionaries, etc.).

Currently, the detection of neologisms is carried out using information technology tools designed for this purpose. In the case of the Barcelona Neology Observatory³ the tool is called Buscaneo,⁴ which was developed by the group itself in 2004 and is now used by all the Spanish neology observatories. Buscaneo scans the press and searches for all the words in the computerised dictionary. To those it cannot

3 <https://www.upf.edu/web/obneo> (last access: 10 June 2022).

4 <http://obneo.iula.upf.edu/buscaneo/> (last access: 10 June 2022).

find, it applies filters to reject proper nouns and other uninteresting units. For the remaining ones, Buscaneo provides an interface allowing users to complete an entry form, adding data or information to fields that the program cannot automatically complete.

Buscaneo (like other automatic neology detectors), which is currently used to extract words from different types of written text (newspapers, magazines, Twitter), makes the task of detecting and recording neologisms considerably less onerous and offers a high degree of reliability. However, it has two limitations that, to date, can only be overcome by performing an additional manual extraction: first, such programs cannot detect semantic or syntactic neologisms (because, formally, they are already in the dictionary) or compound units (because the search strategy is monolexical); and second, they are not yet ready to work with oral-based texts, which are crucial to the study of lexical innovation because they are texts with a more spontaneous style.

2.3 The neological process

Beyond the discussion about which words are neological and which are not, we believe that, from a lexicographic perspective, it is more interesting to try to explain the neological process; a process that begins when a word is born and then becomes a unit that is sufficiently well-established in social use to be included in a general dictionary (although such formalisation may not occur for reasons specific to a particular dictionary), because neologisms at a more advanced stage of the neological process should be the first to be recorded in dictionaries.

This dynamic and complex vision of a neologism is based on the debate initiated by Bauer (1983), with the distinction of three moments in the establishment of a new word: the first occurrence, called a *nonce word*, followed by *institutionalisation* in use, and lastly by *lexicalisation*. That vision reached its culminating point with the work by Schmid (2008), who offered a much more comprehensive evolutionary process that split the evolution of a new word – from its first appearance to the end of its journey – into three stages, which he called *creation*, *consolidation* and *establishment*. At each stage, three processes take place simultaneously until the end of the road: firstly, at the structural level, lexicalisation occurs, which is the formal process from the creation of the word to its fixation; secondly, at the socio-pragmatic level, a neologism spreads among speakers and is potentially institutionalised; and thirdly, at the cognitive level, the concept is hypostatised, and speakers incorporate the lexicalised unit into their mental lexicon.

Based on Schmid's (2008) approach and Kerremans' (2015) review, Freixa (in press) tries to identify different neological behaviours. Of course, a *nonce word* comes first because it is the one that starts the process off. If it stops at that first

occurrence, it will remain as such and not be a neologism proper, precisely because it meets just a momentary expressive need.

Ephemeral neologisms come second. These are units that manage to acquire a certain frequency of use and, in accordance with Schmid (2008), also start the process at the cognitive and structural levels with hypostatisation of the concept and lexicalisation of the form. However, the process then stops because the neologism soon falls into disuse for some reason (but, ultimately, because the concept or form ceases to hold any interest for speakers).

If they do not stop at *nonce words* and are not characterised as being ephemeral, neologisms can follow the stabilisation process in different ways. Renouf (2013) referred to the evolution of neologisms as their *life-cycle*, based on the observation of their frequency. She identified several stages: birth, increase in frequency and occurrence, establishment, death and revival (2013: 182):

The diachronic approach to the study of neologisms in text allows us to observe the existence of a measurable 'life-cycle' for each word. According to this metaphor, used by analogy with a human life-span, the life-cycle of a word is conceived as consisting of some or all of the following major stages: birth, or perhaps just first occurrence in text; possible increase in frequency and occurrence; productivity, creativity, settling down, assimilation and establishment in the language, obsolescence, possible death – and possible revival.

Similarly, in Freixa (in press) the histograms of a set of Spanish neologisms were studied and the following behaviours were identified: first, the *ideal neologism*, characterised by a sustained rise, which necessarily shows that the process has not concluded; second, the *logical neologism*, characterised by a rise and followed by stabilisation; and third, the *realistic neologism*, which rises, falls and then stabilises; and lastly, the *variable neologism*, which fluctuates between more or less pronounced rises and falls.

In this paper, we intend to show how much progress the different Peninsular Spanish pandemic neologisms detected by the Barcelona Neology Observatory have made in the neological process, and whether the behaviours observed in Freixa (in press) can be confirmed. We will also offer some examples of the lexicographic representation that some neologisms already dictionarised have received.

3 Corpus and methodology

The corpus neologisms that we analyse were obtained by manual and automatic extraction from oral texts (radio) and written texts (high circulation newspapers, magazines and Twitter accounts) using the lexicographic criterion mentioned above.

The corpus comprises 209 COVID-19-related neologisms that either appeared for the first time in 2020 and the first half of 2021, or had appeared earlier but experienced a striking increase over this period. The data were extracted from the

BOBNEO database,⁵ but data relating to frequency were supplemented by consulting Factiva,⁶ the world's biggest press database. In the corpus we observed how the frequency of some words was negligible or even non-existent till the beginning of the pandemic as in the case of *nueva normalidad* 'new normality' which numbered 910 occurrences in the year 2015, as a non-lexicalised placement, reaching 162,843 in 2020. We also noticed the extraordinary rise of *covid* and *coronavirus*, making up to more than two and a half million occurrences in just a year, and the emergence of some words exclusively related to the pandemic, such as *anticovid*, not presenting a real evolution and starting to be used in 2020 with a high frequency.

Based on these results, for the analysis we divided the neologisms into different groups, which form a continuum, by taking into account their frequency over the past twenty years (the chart shows the last three years only). We obtained the six groups in Table 1, following a progression in base 10. The table also shows the frequency results from the BOBNEO neologism database over the past thirty years to supplement the previous ones. As can be seen, the neologisms are fairly evenly distributed except in groups 4 and 5, where a greater concentration of cases occurs.

Table 1: Pandemic neologisms in frequency groups.

Group	F Factiva	F OBNEO	Number
1	0–9	1	21
2	10–99	1–4	26
3	100–999	1–9	27
4	1,000–9,999	1–20	67
5	10,000–99,999	1–50	40
6	+100,000	+50	28
			209

For the analysis, information on the horizontal axis of Factiva (age) was also taken into account, and neologisms were labelled according to whether they were first documented in 2020 or whether they already existed, in which case, their distribution was calculated over the years.

As we can see in the last row of Table 2, the neologisms that appeared in 2020 represent one third of the total, but the table shows how they are distributed according to their frequency of appearance: in the more frequent groups of neologisms, the percentage of new ones is 14–15%, whereas in the less frequent groups of neologisms, the percentage of new ones is higher than 80%. This correlation between age and frequency is quite logical.

⁵ <http://obneo.iula.upf.edu/bobneo/index.php> (last access: 10 June 2022).

⁶ <https://global.factiva.com> (last access: 10 June 2022).

Table 2: Age of pandemic neologisms by frequency group.

Group	Earlier	New	Number
1	1	20	21
2	6	20	26
3	20	7	27
4	53	14	67
5	34	6	40
6	24	4	28
	138	71	209

4 Analysis

For the analysis, we took our corpus of pandemic neologisms, organised into different groups by their frequency, and assumed that the more frequent they were, the more dictionarisable they would be. But, based on the most recent literature on updating of dictionaries (Metcalf 2002, Ishikawa 2006, O'Donovan/O'Neill 2008, Cook 2010, Adelstein/Freixa 2013, Freixa 2016, Nam et al. 2016, Freixa/Torner 2020, Klosa-Kückelhaus/Wolfer 2020, Bernal et al. 2020, among others), we also assumed that neologisms would have greater or lesser lexicographic interest depending on how long they had been in use (age), their denominative or stylistic function, their formation mechanism, and other aspects such as record of use.

To observe the extent to which trends in the units' dictionarisation and formation mechanism exist, we take into account the results shown in Table 3, where it is possible to see how the neologisms in each frequency group are distributed by the type of neologism in question. We do not, of course, intend to draw conclusions from a corpus of 200 examples and subgroups of such low numbers, but we do want to comment on the trends observed.

Little can be said about the first five types, since almost no examples were found, but Table 3 shows trends that are taken into account in the analysis, such as the concentration of neologisms formed by blending and neoclassical compounding in the groups where frequency is lower, the concentration of syntagmatic neologisms in the groups where frequency is higher, or the concentration of prefixed neologisms in the intermediate group.

In the analysis discussed below, we have put the six groups into three blocks due to the small corpus of examples. As we shall see, these three blocks have internal consistency: we can consider those in frequency groups 1–2 as *non-dictionarisable neologisms*, those found in groups 3–4 as *neologisms in the antechamber of dictionarisation* and, lastly, those in groups 5–6, where frequency is higher, as *dictionarisable neologisms*.

Table 3: Distribution by types of pandemic neologism, by frequency group.

Type	1	2	3	4	5	6	Total
abbreviation						1	1
conversion						1	1
initialism					1	1	2
borrowing from English		1			1	2	4
semantic change				3	1	1	5
compounding	1	3		4	1	1	10
blending	5	7	1	2	1		16
suffixation	2	3	3	6	3		17
neoclassical compounding	9	5	7	9	5	4	39
prefixation	2	3	9	19	7	7	47
syntagmatic compounding	2	4	7	24	20	10	67
TOTAL	21	26	27	67	40	28	209

4.1 Non-dictionarisable neologisms

In the main, the metaphor of war has been used to frame the discourse around the crisis caused by the COVID-19 pandemic. Most world leaders have done so, although some sectors, especially healthcare, have pointed out that this should not have been the mindset conveyed to the population. But it has been, and continues to be, because it has been observed that the general public reacts obediently to this approach (Sabucedo et al. 2020).

There are, however, different ways of dealing with a crisis, both socially and individually, and words that are heavily loaded with humour or criticism have also appeared in the vocabulary generated by the pandemic. Thus, rather than meeting a denominative need, some of the pandemic neologisms fulfil an expressive one that sometimes seeks to find the funny side of the situation to make it more bearable. These are nonce words.

In our corpus, nonce words account for almost a quarter of the total number of neologisms (47 out of 209). We found *pure nonce words* (group 1, 21 examples), i.e., those that have a really low frequency. But, by extending the concept of nonce word, we have also considered *disseminated nonce words* (group 2, 26 examples), i.e., those spread via social media, with a little higher frequency, although they are still occasional lexical events in the language.

More than half of these examples are formed by neoclassical compounding, a mechanism whose playfulness lies precisely in the seemingly serious and specialised result it yields (*teletrabajopatía* ‘compulsive teleworking’, *metacrisis*, *boeólogo -ga* ‘boeologist’), or by blending, a word formation mechanism to which the literature has attributed a transgressive character (Hohenhaus 2007, Renner 2015, Winter-Froemel/Zirker 2015). In this case, the most recursive blending occurs with the *corona* element (*coronapincho* ‘coronaspike’, *coronahambre* ‘coronahunger’, *coronamiedo* ‘coronafear’). Therefore, some authors refuse to consider them as neologisms (Gérard 2018, Klosa-Kückelhaus/Wolfer 2020, Bueno/Freixa 2020) while not seeking to take away their value; indeed, the study of these units allows us to find out about speakers’ resources and dynamics in terms of linguistic creativity.

4.2 Neologisms in the antechamber of dictionarisation

The block of pandemic neologisms that falls in the central or mean frequency space is the most numerous one and comprises 27 group 3 neologisms (up to 1,000 occurrences in Factiva) and 67 group 4 ones (up to 10,000 occurrences in Factiva). These are, therefore, neologisms that have clearly begun the neological process, but, as we shall see in the analysis, have not yet completed it.

Social institutionalisation is certainly underway, but most have not been around long enough, as only a quarter of these neologisms had been documented previously. In this case, they are non-neological units in specialised use, and the novelty lies in their spread to general use: *azitromicina* ‘azithromycin’ has been documented since 1997 and has a total of 3,670 occurrences, of which 2,451 were observed in 2020 (in previous years, there were no more than 240 a year); in a lower frequency range, *apoyo respiratorio* ‘breath support’ has a total of 651 occurrences since it was first documented in 1995, of which 440 were observed in 2020 (in previous years, there were no more than 34 a year). Other units like these are *test serológico* ‘serological test’, *pluripatología* ‘multipathology’, *presintomático -ca* ‘presymptomatic’, etc. These units will most likely not complete institutionalisation in general use, and will return to specialised use, although this will depend on what happens with the pandemic we are still experiencing.

The abandonment of the neological process that some units have initiated will also depend on how the pandemic develops: ephemeral neologisms are units that disappear from use when they are dependent on a passing social phenomenon (be it a technological discovery, a health crisis or perhaps something related to the fashion world). *Covidiota* ‘covidiot’, *balconero -ra* ‘balconer’, *telecolegio* ‘tele-school’, *coronabono* ‘coronabond’, *grupo burbuja* ‘bubble group’, to mention a few, may disappear from use before they become stable. But we must bear in mind that a characteristic feature of ephemeral neologisms is that their birth may occur more than once, i.e., a neologism that did not become institutionalised may have new

opportunities. *Coronavirus*, for example, has been sporadically documented in high circulation newspapers for more than 20 years, but it had an opportunity to become institutionalised in 2003, when the number of occurrences reached more than 1,000 due to the severe acute respiratory syndrome coronavirus (SARS-CoV) epidemic in Southeast Asia. However, the word's appearance became residual in just two years. A new attempt to become institutionalised occurred in 2015, with the Middle East respiratory syndrome coronavirus (MERS-CoV). Although its high frequency peak lasted only one year, *coronavirus* remained in use with about 100 occurrences per year until 2020, when it finally became institutionalised.

According to Schmid (2008), in the establishment of a word or what we call *the neological process*, besides institutionalisation in use, lexicalisation⁷ occurs at the structural level and hypostatisation takes place at the conceptual level. Lexicalisation is a process of linguistic fixation of a new word's formal and semantic aspects, and thus it acquires a more precise meaning and a less variable form. This process, which is initiated with the first occurrences of a neologism, does not appear to have been completed in some of the examples making up the block of neologisms under analysis. For example, in Table 4, we can see that the neologism *distancia social* 'social distance' coexists alongside a diverse range of forms that show different degrees of social institutionalisation. These variants display the most defining semantic features of the concept, and together show that there has not yet been any formal fixation that, to some extent, lexicalisation entails (although the number of occurrences does inform us of the preferred variants in use).

Table 4: The neologism *distancia social* 'social distance' and its variants.

neologism	occurrences in Factiva
distanciamiento social 'social distancing'	114,386
distancia de seguridad 'security distance'	101,773
distancia social 'social distance'	68,134
distanciamiento físico 'physical distancing'	19,154
distancia interpersonal (interpersonal distance)	17,084
distancia física 'physical distance'	17,018
distancia sanitaria 'sanitary distance'	375

And lastly, the concepts denominated by these neologisms cannot be deemed hypostatised by the majority of speakers. When a speaker is faced with a new word, he or she analyses its morphological constituents. The more transparent and less ambiguous the morphological structure of the word is, the faster the process of understanding

⁷ See Lipka et al. (2004) for a review of concepts of the concepts of institutionalisation and lexicalisation.

it will be. And, depending on its level of semantic transparency, the formation of the new concept will be easier or harder. Such semantic transparency is determined by the frequency of the constituents, the number of existing lexemes with those constituents, and the semantic relationship between them. In addition, the information provided by the co-text and the context influences the development of the new concept (Schmid 2008). Some of the neologisms in this block are at an advanced stage of hypostatisation (*mascarilla higiénica* ‘hygienic mask’, *posconfinamiento* ‘posconfinement’, *antimascarillas* ‘antimasks’), but others are not for a variety of reasons, such as the fact that they are highly specialised units (*gerontofobia* ‘gerontophobia’, *sobreinfección* ‘overinfection’, *dexametasona* ‘dexamethasone’).

We have therefore said that the neologisms in this block (frequency groups 3 and 4) are in the antechamber of dictionarisation because it is not yet time for them to enter it. The lexicographic interest that these units hold will depend on the course they take over the coming years, which in turn will depend on the evolution of the COVID-19 pandemic. Some of them, i.e., those bordering on the block of more frequent neologisms, are more institutionalised in use, are more lexicalised units, and a higher number of speakers has already hypostatized the concept, but the neological process has not yet been completed.

Those neologisms that succeed in completing this process will then face selection by a dictionary, in line with its internal criteria. In relation to DLE, Bernal et al. (2020) have noted that the internal balance of the dictionary ultimately determines the decision-making. So, for example, in the dictionary update, those neologisms forming a derivative series are good candidates. But, of course, the series cannot be unlimited: the words *infección* ‘infection’, *infectar* ‘to infect’ and *infeccioso* -sa ‘infectious’ are already in DLE. In pandemic use, however, the derivatives *sobreinfección* ‘overinfection’, *reinfección* ‘reinfection’, *reinfectar* ‘to reinfect’ and *reinfectado* -da ‘reinfectad’ are recurrent and, since all of them are predictable derivatives, the dictionary may not consider them necessary (Bernal, 2021). The same applies to the *pandemia* ‘pandemic’ family (*postpandémico* -ca ‘postpandemic’, *prepandémico* -ca ‘prepandemic’, *antipandémico* -ca ‘antipandemic’ and its variants, and the *cuarentena* ‘quarantine’ family (*precuarentena* ‘prequarantine’, *postcuarentena* ‘postquarantine’, *semicuantena* ‘semiquarantine’), among others.

These neologisms are not usually included in general dictionaries and, at most, can be found in dictionaries of neologisms, especially, in those produced in digital format. This is the case with Antenarío,⁸ a dictionary of neologisms monthly updated by the neology groups in the net of Antenas Neológicas,⁹ with unities from the different geolectal varieties of Spanish. In Antenarío, more than 50 neologisms have

⁸ Antenarío: <https://antenario.wordpress.com> (last access: 10 June 2022).

⁹ Antenas Neológicas: <https://www.upf.edu/web/antenas> (last access: 10 June 2022).

already been published under the thematic label of *Pandemia Covid-19* ‘COVID-19 pandemic’. One of them is shown in Figure 1:

reinfectarse v intr

Año de la primera documentación: 2020

Definición	Volver a infectarse con el virus o la bacteria que produce una enfermedad después de haberse recuperado de ella.
Contextos	<p>«Este caso muestra que es posible reinfectarse solo unos meses después de haberse curado de una primera infección», indicó en un comunicado el departamento de microbiología de la Universidad de Hong Kong (HKU). [<i>El Universal</i> (México), 24/08/2020]</p> <p>Ha advertido hoy de que han detectado que hay personas que han pasado la infección que no desarrollan la inmunidad y pueden reinfectarse. [<i>El Periódico</i> (España), 7/05/2020]</p>
Nodos	ARG CHL ESP MEX PER
Diccionarios	Alvar1 • Alvar2 • Clave • DAMER • DEA • DNEA • DUE4 • LAR
+ info	Martes Neológico • Neologismo del mes

Figure 1: Example of pandemic neologism published in Antenarior.

In Antenarior, the choice made is one of building a blog-format dictionary with thematic, linguistic and pragmatic tags users can send their comments to. As seen in Figure 1, neologisms are accompanied by the usual information in the micro-structure of a dictionary (lemma, grammatical category, definition and examples) and complementary information related to geolectal information as well as to the neologicity of the word (age and dictionaries in which they are already documented).

4.3 Dictionarisable neologisms

The 68 most frequent and, in principle, more dictionarisable neologisms can be found in this block. They are more dictionarisable because they are the most institutionalised ones in use and probably the most lexicalised and hypostatized ones too, because lexicalisation and hypostatisation come from use. This block, which includes 40 neologisms with a frequency between 10,000 and 99,999 occurrences and 28 neologisms with a frequency of at least 100,000 occurrences, also contains the highest percentage

of pre-existing neologisms (85.3% had already been documented prior to the pandemic). It is therefore a set of neologisms that meet two of the criteria that are often mentioned in the literature for the purpose of assessing their dictionarisation (Metcalf 2002, Ishikawa 2006, O'Donovan/O'Neill 2008, Cook 2010, Adelstein/Freixa 2013, Freixa 2016, Freixa/Torner 2020). Also mentioned in it are other criteria relating to use, which the pandemic neologisms in this group also fulfil, such as currency (they are current neologisms, although all the pandemic neologisms meet this criterion) and textual spread (they are used in texts of different types).

As for linguistic criteria, all the neologisms fulfil the criterion of correct formation and semantic necessity because, although most have predictable and compositional meaning (*semipresencial* 'semipresential', *gel hidroalcohólico* 'hydroalcoholic gel', *supercontagiador* 'superinfecter'), the speaker does not know its precise meaning. In fact, the most lexicalised syntagmatic neologisms are concentrated in this block; they are clearly denominative and, in this case, widespread in use: *crisis sanitaria* 'sanitary crisis', *presión hospitalaria* 'hospital preassure', *servicio esencial* 'essential services', among others. While general dictionaries have tended to restrict the incorporation of polylexematic units, DLE has gradually become more open to units like these, which become subentries of existing words.

The neologisms in this block also meet documentary criteria because most of them are listed in pandemic-themed dictionaries that have recently appeared, such as the *Diccionario de covid-19 (EN-ES)*¹⁰ by the International Association of Medical Translators and Writers and Related Sciences (Tremédica). Thus, they are neologisms that have completed the neological process and, in fact, some of them have recently been incorporated into DLE, as we shall see. *Close contact* and *social bubble* are two of the pandemic neologisms already collected in the terminological dictionary published by Tremédica, as seen in Figure 2.

As can be seen, the lexicographic representation is different in this case, as the most important information for translators has been prioritised, precisely because Tremédica is an international association of medicine and related sciences translators. This way, as well as the equivalents in English, we can also consider the synonyms in both languages.

5 Already dictionarised pandemic neologisms

Fourteen of the neologisms in our corpus have already ceased to be neologisms according to the lexicographic criterion because they have recently been incorporated into DLE. These words are shown in Table 5, and reference is made to the frequency

¹⁰ <https://www.tremedica.org/tremediteca/glosarios/diccionario-de-covid-19-en-es/> (last access: 10 June 2022).

close contact

contacto estrecho; contacto cercano

Sinonimia (en): *close-distance contact; high-risk contact.*

Sinonimia (es): contacto de alto riesgo.

Concepto: se refiere a estar dentro de un radio de 2 metros de la persona contagiada por tiempo prolongado y sin equipo de protección. Incluye a las personas que habitan bajo el mismo techo (→ *household contact*).

Ejemplo: *According to the Center for Disease Control and Prevention, close contact is defined as: a) Being within approximately 6 feet (2 meters) of a COVID-19 case for a prolonged period of time; close contact*

social bubble

burbuja social

Sinonimia (en): *bubble group; corona bubble; COVID bubble; COVID-19 bubble; germ bubble; quaranteam.*

Sinonimia (es): burbuja; grupo burbuja; grupo de convivencia estable [GCE].

Concepto: fórmula de aislamiento social intermedia entre relacionarse con todo el mundo y no hacerlo con nadie. Consiste en establecer un grupo cerrado de pequeño tamaño (por lo general, seis personas) con el que, sin ser convivientes, se mantiene contacto de manera recurrente.

Figure 2: Examples of pandemic neologisms collected by Tremédica.

group from our analysis for the purpose of seeing whether the dictionarised neologisms matched the more dictionarisable ones:

Table 5: Pandemic neologisms incorporated into RAE dictionary.

Covid	6a	medicalizar ‘to medicalize’	4
coronavírico -ca ‘coronaviral’	4	panémico –ca ‘pandemic’	5
Coronavirus	6	positivo ‘positive’	5
cuarentenear ‘to quarantine’	2a	telemedicina ‘telemedicine’	5
desconfinamento ‘de-confinement’	6	teletrabajador ‘to telework’	4
desconfinar ‘to de-confine’	4a	teletrabajo ‘teleworking’	6
desescalada ‘de-escalation’	6	videollamada ‘videocall’	6

Indeed, most of the incorporated neologisms are in the higher frequency range (groups 5 and 6) although, as we can see, some of them are in the middle range (groups 3 and 4) and one is in the lower frequency range (groups 1 and 2). We will first focus our attention on the latter, the neologisms which, because of their frequency, were not the best candidates for updating the dictionary. The first, and most exceptional one is *cuarentenear* ‘to quarantine’, a verb that occurs just three times in BOBNEO and 91 times in FACTIVA, so it seems to be a nonce word that has spread to some extent. Given that the verb *cuarentenar* ‘to quarantine’ already exists in DLE,

the introduction of the verb ending in *-ear* might be linked to a willingness to provide better representation of non-peninsular varieties of Spanish, since *cuarentenear* ‘to quarantine’ has mostly been documented in Latin American countries.

The adjective *coronavírico* *-ca* ‘coronaviral’, the noun *teletrabajador* *-ra* ‘teleworker’ and the verbs *medicalizar* ‘to medicalize’ and *desconfinar* ‘to de-confine’ have, in our opinion, been rightly dictionaryed for the reasons set out below. These, as Bernal et al. (2020) have already stated, are associated with DLE’s internal criteria. On the one hand, they all have a relatively high frequency (more than 1,000 occurrences in 2020) and, on the other, they all complete a derivative series of other words that were already present or have been recently incorporated into the dictionary: *teletrabajador* *-ra* ‘teleworker’ (lower frequency) is consistent with the incorporation of *teletrabajo* ‘teleworking’ (but clearly inconsistent with the absence of the verb *teletrabajar* ‘to telework’), and *coronavírico* *-ca* ‘coronaviral’ is relevant since *coronavirus* and certain derivatives thereof have also been incorporated. In some cases, the neologisms also meet the criterion of age: *teletrabajador* *-ra* ‘teleworker’ has been documented since 1995 and *medicalizar* ‘to medicalize’ since 1999, and the *cruciality* (Sheidlower 1995) of both is evident, since they are not products of a passing fad. All of them have a clear denominative function, had already been documented in specialised dictionaries, and refer to terms about which users may have some doubts regarding meaning or use (thus, for example, DLE gives two meanings for *medicalizar*: “dotar a algo, como un medio de transporte, de lo necesario para ofrecer asistencia médica” [to give something, such as a means of transport, what is needed to offer medical care] and “dar carácter médico a algo” [to give something a medical character]. Lastly, we should add that there is no characteristic in their formation that would render them unsuitable candidates for updating DLE.

The verb *desconfinar* ‘to de-confine’ deserves special attention. We would argue that its incorporation is justified in accordance with most of the criteria set out above, such as the completion of a derivative series: *confinar* ‘to confine’ and *confinamiento* ‘confinement’ were already in the dictionary, so the incorporation of reversible forms (*desconfinar* ‘to de-confine’ and *desconfinamiento* ‘de-confinement’) is as logical as the incorporation of other members of the same family having a similar frequency of use and cruciality would be, but which have nevertheless been left out: *preconfinamiento* ‘preconfinement’, *posconfinamiento* ‘postconfinement’, *reconfinamiento* ‘reconfinement’ and *autoconfinamiento* ‘autoconfinement’. However, as already mentioned in previous paragraphs, the criterion of completion of a derivative series is limited by the criterion of formal and semantic predictability, which is used to reject units.

The other neologisms incorporated into DLE (Table 5) are in the higher frequency groups in the consulted corpora; some appeared in 2020 while others had occurrences in previous years, yet the cruciality of all of them has been evident during the pandemic. In descending order, with four million occurrences in Factiva, is *covid* (slightly more than *coronavirus*) and, with much lower frequencies but still in the highest frequency group, are *desescalada* ‘de-escalation’ (263,000) and *teletrabajo*

‘teleworking’ (156,000). The other dictionarised neologisms from the frequency group ranging from 10,000 and 99,000 occurrences are *desconfinamiento* ‘de-confinement’ (56,000), *pandémico -ca* ‘pandemic’ (52,424), *videollamada* ‘videocall’ (36,800) and *telemedicina* ‘telemedicine’ (31,392).

In Figure 3 we can see three of the already collected pandemic neologisms in DLE:

coronavirus

Del ingl. *coronavirus*, de *corona* ‘corona solar’, por el aspecto del virus al microscopio, y este del lat. *corōna* ‘corona’, y *virus* ‘virus’, y este del lat. *virus* ‘veneno’, ‘ponzoña’.

1. *m. Med.* Virus que produce diversas enfermedades respiratorias en los seres humanos, desde el catarro a la neumonía o la COVID.

cuarentenear

Conjugar

1. *intr.* Pasar un período de **cuarentena** (|| aislamiento preventivo por razones sanitarias). *Es más llevadero cuarentenear con alguien.*

2. *tr. p. us.* Poner algo o a alguien en **cuarentena** (|| aislamiento preventivo por razones sanitarias). *Tendremos que cuarentenear el ganado. Las autoridades cuarentenearon el crucero.*

desconfinamiento

1. *m.* Levantamiento de las medidas impuestas en un confinamiento.

Figure 3: Three pandemic neologisms collected in DLE.

Figure 3 also shows how the lexicographic representation fits this kind of dictionary, in this case, a general Spanish language dictionary, also being an academic dictionary. This way, for *coronavirus*, a neologism that speakers could consider semantically unclear, the dictionary provides information about its origin and its usage in the medical area. For *cuarentenear* ‘to quarantine’ or *desconfinamiento* ‘de-confinement’, words formed following the word formation rules in Spanish, this information about origins is not provided but linguistic and usage information are.

DLE’s rapid incorporation of these words is certainly positive. They meet various dictionarisation criteria and their frequency is high. However, in line with these criteria, many others may get the opportunity to be accepted into the dictionary in future updates: examples such as *gel hidroalcohólico* ‘hydroalcoholic gel’, *ensayo clínico* ‘clinical trial’, *distancia social* ‘social distance’ (or *distanciamiento social* ‘social distancing’, or variants deemed preferential, precisely pointing to usage) are units that are clearly denominative, even in the form of subentries, because of their

syntagmatic nature. Equally necessary are other words formed by compounding, such as *infectólogo -ga* ‘infectologist’, *sociosanitario -ria* ‘sociosanitary’, *semipresencial* ‘semipresential’; by blending, such as *conspiranoico -ca*; or by initialism, such as *EPI* ‘PPE’ and *ERTE* ‘furlough’. Likewise, the fact that other high frequency neologisms have been left out is understandable because they are descriptive syntagmas, such as those that have ‘crisis’ as their base: *crisis del covid* ‘covid crisis’, *crisis sanitaria* ‘sanitary crisis’, *crisis social* ‘social crisis’, or those with different families of derivatives, especially with *pre-*, *post-* and *anti-* attached to *covid*, *coronavirus*, *pandemia* ‘pandemic’ and other pandemic-related terms.

That said, they are neologisms that have become stable in use, and their incorporation into the dictionary will depend on the criteria that the dictionary applies to the units, not as neologisms but as language units. According to Torner (in press), a study for the lexicographic sanctioning of neology should consider this dual dimension and observe neological forms from this two-fold perspective. The dictionarisation of neology is a dual property acting on a two-fold plane: that of consolidation in use on the one hand, and that of the criteria governing the elaboration of dictionaries on the other (Torner in press).

6 Conclusions

In her magnificent work published in 2015, Kerremans compared neologisms to casting show winners: some become stable or consolidated as singers, others get to have a hit, yet most fall into oblivion. The television industry provides a context within which they can gain huge popularity within a very short space of time, but as the focus of the industry’s interest shifts, the artists’ popularity may quickly fall. Some manage to keep going for a while, while others manage to break into the industry without even winning the contest, so there does not appear to be a recipe for guaranteed success (Kerremans 2015: 15).

Indeed, the most dictionarisable neologisms are those with certain characteristics, yet reality has shown us time and again that many of the neologisms that fulfil those seemingly essential characteristics may not become stable, while others that do not fulfil them may.

The pandemic has mobilised vocabulary in an unprecedented way, as noted by Pons (2020) and, just 20 days into the first lockdown, words that had not previously existed began to appear, words that had not been used for a long time were revived (*lexical resurrection*, according to Pons, such as the verb *desescalar* ‘de-escalate’), or a new sense or a more specific meaning was given to words already in use.

We do not know how many neologisms have been created since the start of the pandemic, but there are undoubtedly many more than the 209 analysed in this work, based on the Neology Observatory’s extraction of neologisms from oral and written

texts. Such extraction has been performed annually since 1989. It provides a snapshot of how the lexicon of the language has developed to adapt to the changes in society. However, that extraction is not systematic, and although the most frequent neologisms have been detected because of their recurrent appearance in the press, many of the more fleeting ones have not. Had they been detected, the latter would have considerably enlarged our corpus of pandemic neologisms. Nonetheless, with the corpus available to us, we have been able to see that new words did appear, others were reborn, and some of the already existing ones have taken a new path.

Looking at the corpus from a lexicographic perspective, we divided this new vocabulary into three blocks. In the first, we found good examples of speakers' creativity in terms of meeting their more expressive and less denominative needs with nonce words, which performed their function yet held no lexicographic interest. In the second, we analysed a set of neologisms midway along the neological process, which could not be deemed stable in use and, therefore, were in the antechamber of dictionarisation; the path that these might ultimately take is unknown. And, in the last block, we observed those neologisms that had already completed their journey; some have already been lexicographically sanctioned, and others may be in due course.

Bibliography

Research literature

- Adelstein, Andreína/Freixa, Judit (2013): Criterios para la actualización lexicográfica a partir de datos de observatorios de neología. Presentation at the International Congress *El diccionario: neología, lenguaje de especialidad, computación* (28–30 October 2013) in Ciudad de México.
- Bauer, Laurie (1983): *English Word-formation*. Cambridge: Cambridge University Press.
- Bernal, Elisenda (2021): El asedio lexicográfico a los derivados predictibles. In: *Revista Española de Lingüística* 51(1), 79–96.
- Bernal, Elisenda/Freixa, Judit/Torner, Sergi (2020): Criterios para la diccionarización de neologismos: de la teoría a la práctica. In: *Signos* 53(104), 592–618.
- Bueno, Pedro J./Freixa, Judit (2020): ¿Son neologismos los ocasionalismos? In: Guerrero, Gloria/Pérez Lagos, Fernando (coord.): *Terminología, Neología y Traducción*. Madrid: Comares, 1–20.
- Cook, Paul C. (2010): *Exploiting Linguistic Knowledge to Infer Properties of Neologisms*. Toronto: University of Toronto.
- Freixa, Judit (in press): El proceso neológico. In: Bernal, Elisenda/Freixa, Judit/Torner, Sergi (coord.): *La neología del español: del uso al diccionario*. Madrid: Iberoamericana.
- Freixa, Judit (2016): Reagrupación lexicográfica de neologismos frecuentes. In: Bernal, Elisenda/Torner, Sergi (eds.): *Relaciones morfológicas y diccionario*. A Coruña: Universidade da Coruña, 95–112.
- Freixa, Judit/Torner, Sergi (2020): On the Dictionarization of New Words in Spanish. In: *Dictionaries* 41(2), 131–153.

- Gérard, Christophe (2018): Variabilité du langage et productivité lexicale. Problèmes et propositions méthodologiques. In: *Neologica* 12, 23–45.
- Guilbert, Louis (1975): *La créativité lexicale*. Paris: Larousse.
- Hohenhaus, Peter (2007): How to do (even more) things with nonce words (other than naming). In: Munat, Judith (ed.): *Lexical Creativity, Texts and Contexts*. Nottingham: John Benjamins, 15–38.
- Ishikawa, Shin'ichiro (2006): When a word enters the dictionary: A data-based analysis of neologism. In: JACET Society of English Lexicography (ed.): *English Lexicography in Japan*. Bunkyo-ku: Taishukan, 39–52.
- Kerremans, Daphné (2015): *Web of New Words: A Corpus-Based Study of the Conventionalization Process of English Neologisms*. *English Corpus Linguistics* 15. Frankfurt am Main: Peter Lang.
- Klosa-Kückelhaus, Annette/Wolfer, Sascha (2020): Considerations on the acceptance of German neologisms from the 1990s. In: *International Journal of Lexicography* 33, 150–167.
- Lipka, Leonhard/Handl, Susanne/Falkner, Wolfgang (2004): Lexicalization & Institutionalization. The State of the Art in 2004. In: *SKASE Journal of Theoretical Linguistics* 1, 2–19.
- Matoré, Georges (1952): Le neologisme: naissance et difusión. In: *Le français moderne* 2, 87–92.
- Metcalf, Allan (2002): *Predicting New Words*. Boston: Houghton Mifflin.
- Nam, Kilim et al. (2016): The life and death of neologisms. On what basis shall we include neologisms in the dictionary? In: Margalitadze, Tinatin/Meladze, George (eds.): *Proceedings of the XVII EURALEX International Congress: Lexicography and Linguistic Diversity*. Tbilisi: Ivane Javakhishvili Tbilisi State University, 389–393.
- O'Donovan, Ruth/O'Neill, Mary (2008): A systematic approach to the selection of neologisms for inclusion in a large monolingual dictionary. In: Bernal, Elisenda/DeCesaris, Janet (eds.): *Proceedings of the XIII EURALEX International Congress (Barcelona, 15–19 July 2008)*. Barcelona: Institut Universitari de Lingüística Aplicada – Universitat Pompeu Fabra, 571–579.
- Pons, Lola (2020): 'Covidiota', 'balconazis', 'cuarenpena' . . . los neologismos que nos ha traído la pandemia. In: *El País*, 08/04/2020.
- Renner, Vincent (2015): Lexical blending as wordplay. In: Zirker, Angelika/Winter-Froemel, Esme (eds.): *Wordplay and Metalinguistic/Metadiscursive Reflection: Authors, Contexts, Techniques, and Meta-Reflection*. Berlin/Boston: De Gruyter, 119–133.
- Renouf, Antoinette (2013): *A finer definition of neology in English: The life-cycle of a word*. In: Hasselgård, Hilde/Ebeling, Jarle/Oksfjell Ebeling, Signe (eds.): *Corpus Perspectives on Patterns of Lexis*. Amsterdam: John Benjamins, 177–208.
- Rey, Alain (1976): Néologisme: un pseudo-concept? In: *Cahiers de lexicologie* 28(1), 2–17.
- Sablairolles, Jean-François/Humbley, John (2021): Terminologie de la néologie. In: *Neologica* 15, 63–96.
- Sabucedo, José M./Alzate, Mónica/Hur, Domenico (2020): Covid-19 y la metáfora de la guerra. In: *International Journal of Social Psychology* 35(3), 618–624.
- Schmid, Hans-Jörg (2008): New words in the mind: concept-formation and entrenchment of neologisms. In: *Anglia* 126(1), 1–36.
- Sheidlower, Jesse T. (1995): Principles for the inclusion of new words in college dictionaries. In: *Dictionaries* 16, 33–44.
- Torner, Sergi (in press): Selección del leuario en los diccionarios de lengua y diccionarización de la neología. In: Bernal, Elisenda/Freixa, Judit/Torner, Sergi (eds.): *La neología del español: del uso al diccionario*. Madrid: Iberoamericana.
- Winter-Froemel, Esme/Zirker, Angelika (2015): Wordplay and its interfaces in speaker hearer interaction: an introduction. In: Winter-Froemel, Esme (ed.): *The Dynamics of Wordplay*. Berlin/Boston: De Gruyter, 1–22.

Dictionaries

[BOBNEO] *Base de datos de los observatorios de neología*. Observatori de Neologia (Institut de Lingüística Aplicada, Universitat Pompeu Fabra). [<http://obneo.iula.upf.edu/bobneo/index.php>, accessed July 2021].

[DLE] *Diccionario de la lengua española*. Real Academia Española. [<https://dle.rae.es>, accessed July 2021].

[FACTIVA] Dow Jones. 1989–. *Factiva*. New York: Dow Jones & Company. [<https://global.factiva.com>, accessed July 2021].

[Tremédica] *Diccionario de covid-19 (EN-ES)*. International Association of Medical Translators and Writers and Related Sciences (Trémédica). [<https://www.tremedica.org/tremediteca/glosarios/diccionario-de-covid-19-en-es/>, accessed July 2021].