

ojs.uv.es/index.php/qfilologia/index

Rebut: 18.07.2023. Accepatat: 15.09.2023

Per a citar aquest article: Renau, Irene. 2023. "A corpus-based study of semantic neology of the Covid-19 pandemic". *Quaderns de Filologia: Estudis Lingüístics* XXVIII: 55-76.

doi: 10.7203/QF.28.26663



## A corpus-based study of semantic neology of the Covid-19 pandemic

Estudio basado en corpus de la neología semántica de la pandemia del COVID-19

IRENE RENAU

Pontificia Universidad Católica de Valparaíso

irene.renau@pucv.cl

**Resumen:** Este artículo presenta una serie de estudios basados en corpus para la detección semiautomatizada de neología semántica en los sustantivos del español en 2022, año en el que todavía existía la pandemia del COVID-19, pero también otros eventos globales importantes, como la guerra de Ucrania o la inflación económica. Comparamos un corpus de textos de 2022 con un corpus previo a la pandemia, considerando cuatro tipos de cálculo de frecuencia y palabras clave y teniendo en cuenta la combinatoria léxica de los candidatos a neologismo. Los resultados indican que la combinación de palabras clave con textos agrupados según la temática da los mejores resultados para detectar neología semántica.

**Palabras clave:** colocaciones; corpus; lexicografía; metáfora; metonimia; neología semántica.

**Abstract:** This article presents a series of corpus-based studies to detect semantic neology in Spanish nouns in 2022, a year in which the Covid-19 pandemic still existed, but when another important global events were taking place, such as the Ukraine war, or economic inflation. We compared a corpus of texts from 2022 with a pre-pandemic corpus, considering four types of frequency calculation and keywords and taking into account collocations of the neologism candidates. The results indicate that the combination of keywords with texts grouped according to topic gives the best results for detecting semantic neology.

**Keywords:** collocations; corpus; lexicography; metaphor; metonymy; semantic neology.

### 1. Introduction

In this article, we present a corpus-based study of semantic neology that emerged during the Covid-19 pandemic. We combine quantitative and qualitative analysis, using different parameters to detect new meanings in pre-existing Spanish nouns. Specifically, we are interested in observing the relation between frequency and semantic neology, as well as keywords and semantic

neology. Additionally, we want to observe if grouping texts according to a common topic leads to better results than not filtering by topic. We use a corpus of news of 2022 and compare it with a corpus of 2017, 2018, and 2019. Our goal is to observe which criteria and corpus data are more useful for semantic neology discovery, specifically considering the needs of lexicographic teams. We consider this piece of research a preliminary attempt to conduct large-scale studies in the future, and for that reason we use a small dataset and a semi-automatic procedure. The results may offer clues about which strategies are worth considering with larger corpora and a fully automatic technique.

New words and meanings emerge everywhere and every time. Collective important events are linked to the surge of new words and meanings, e. g., economic crisis (Galanes-Santos, 2019; Renau, Nazar & Lecaros, 2020), environment (Sanmartín Sáez, 2016; Guslyakova, Valeeva & Vatkova, 2020), and, of course, the Covid-19 pandemic –cf. Klosa-Kückelhaus & Kernerman (2022) for a compilation of studies of neology during that period. In these events, “speakers are confronted not only with an unprecedented situation in life, but also with the need to understand many terms previously unknown to them and a large number of new lexical items emerging in a short time span” (Klosa-Kückelhaus, 2022: 115). In particular, the Covid-19 crisis was one of the most significant pandemics in human history, and it was also the first global event in which information could be shared on a large scale extremely quickly and efficiently thanks to the Internet (Harari, 2020). These factors, indeed, have a deep impact in the speed, volume, and ways new words and meanings are born in languages. However, so far, neology studies in different languages related in one way or another with the Covid-19 pandemic are mainly focused on formal neology, while there are almost no studies devoted to semantic neology during the same period. It is unclear, hence, which was the role of semantic neology in the acquisition of new meanings during the pandemic, in terms of frequency, typology or other aspects.

Dealing with semantic neology has a strong connection with the problem of dictionary updating (Cook & Hirst, 2011; Renouf, 2014), a transversal, recursive activity which affects all tasks of the lexicographic process (as it is described in Klosa-Kückelhaus, 2013). The sub-process of dictionary updating is a never-ending stage, which is broken into more detailed tasks such as correcting typos, updating orthography, adding supplementary materials, etc. Incorporating new words and meanings is one of the most relevant tasks of this sub-process. Today, a dictionary that does not include the word *Covid* in the lemma list is abruptly outdated, in the same way that the dictionaries

which, in the nineties, did not incorporate the word *Internet*. Online dictionaries, which are the prototypical dictionaries of our time, can –and must– be constantly updated to fulfil user needs, with whom they have an important social responsibility: “Practical lexicography must constantly be aware of its social responsibility and must strive for a comprehensive, pluralistic description of linguistic and factual realities” (*Vila Vigoni Theses*, 2018: thesis 3). One of the conditions for a socially responsible lexicography is its sustainability, that is, the ability of dictionaries to be maintained and constantly updated, at –apparently– great financial cost (Colman, 2018). This means that methods which can provide a faster and more efficient updating in dictionary projects are among the key aspects of practical lexicography today.

In the following pages, we present the state of the art in semantic neology, synthesizing the main theoretical and methodological issues (section 2). We then explain our materials and methods for our analysis (section 3). We present our results and discussion (section 4) and end with some concluding remarks and future work (section 5).

## 2. State of the art

### 2.1 *Conceptualisation and categorisation of semantic neology*

Semantic neology can be defined as the emergence –or creation– of a new meaning in an existing lexical item. In contrast, formal neology is the emergence –or creation– of a new form (lexical item, morpheme) (Sablayrolles, 2012; Renouf, 2014; Adelstein, 2017; Barque, Haas & Huyghe, 2018; Klossa-Kükenhaus & Kernermann, 2020). For example, *Covid* is formal neology and *bubble* (as in ‘social bubble’ or ‘safe sanitary social group’) is semantic neology. The concept of semantic neology inherits the complexities of defining what meaning is in general, or rather lexical meaning, which poses an extremely difficult problem (Yallop, 2004; Gérard & Kabatek, 2012).

Types of semantic neology are the same that are usually described for semantic change, especially metaphor, metonymy, specialisation, and generalisation (Geeraerts, 2010: 26-31) (table 1).

<i>Type of semantic neology</i>	<i>Description</i>	<i>Example</i>
Metaphor	The literal meaning is linked to the source domain of the metaphor, and the new meaning to the target domain (cf. Barcelona, 2003).	<i>burbuja</i> ‘bubble’ (Spanish). From a ‘sphere of gas contained by a substance’ (source domain), we obtain the metaphor ‘excessive increase of an asset, caused by speculation’ (target domain) (Renau, Nazar & Lecaros, 2020: 244).
Metonymy	The new meaning is linked to the source meaning by a relation of spatial, temporal, or logical contiguity.	<i>paternalisme</i> ‘paternalism’ (French). From the ‘attitude of treating adults as children’, we metonymically obtain the ‘action of treating adults as children’ (Lombard, Huyghe & Barque, 2023: 11).
Generalisation	The new meaning has a relation of superordination to the source meaning.	<i>corralito</i> ‘government restriction of the free disposition of cash from bank accounts’ (Spanish). This word was coined in 2001 in Argentina (Adelstein & Kugel, 2008: 76) and was later generalised to any similar restriction in any crisis or country (Renau, Nazar & Lecaros, 2020: 244, 248).
Specialisation	The new meaning has a relation of subordination to the source meaning.	<i>ajuste</i> ‘adjustment’ (Spanish). From the ‘action and result of adjusting something’ we obtain a new, very concrete meaning in the field of economics (Renau, Nazar & Lecaros, 2020: 244).

Table 1. Types of semantic neology

## 2.2 *Methods for the automatic detection of new meanings*

Addressing novel sense identification for lexicographic purposes requires covering large segments of vocabulary in a fast and systematic way. As explained in section 1, this is necessary for the dictionary to respect its social role in helping the user to learn about new words and realities, and so requires speed and effectiveness –the latter also in economic terms. Thus, even when there are experimental studies for neology detection (Lombard, Huyghe & Gyax, 2021; Lombard, Huyghe, Barque & Gras, 2023), they cannot be considered as systematically incorporated into the dictionary making process. We must also discard the “introspection by the linguist” (Sinclair, 1991: 39), that is, when the lexicographer adds new meanings because she/he remembers them from having read or heard them in a random source and intuitively label them as neologisms. This practice, even when it does not seem appropriate, is still

sometimes used today in the absence of better alternatives, which can be complex and expensive. Sinclair (1991: 39) clarifies that it is appropriate that linguists use their personal judgement to *interpret* corpus evidence, not to *create* the evidence.

The standard technique in semantic neology detection, which can be applied to dictionary projects, is using corpora, and the aim is to identify those parameters that provide the best candidates for semantic neologisms. It is much more difficult to detect semantic neology than lexical neology, because in semantic neology the change is not formal, thus, it must be detected in alternative, more complex ways, and the error rate can be high (Renouf, 2014; Torres-Rivera & Torres-Moreno, 2020; Sajous, 2022).

There is not yet a standard procedure for automatic novel sense identification using corpora. However, attempts made until now use a diachronic approach in which two or more corpora are compared using different metrics and strategies. The most common strategy consists on comparing differences in the surface of the text and, particularly, changes in the collocational patterns of the target words (Cook & Hirst, 2011; Cabré & Nazar, 2012; Renouf, 2014; Torres-Moreno et al., 2020; Bochkarev, Khristoforov, Shevlyakova & Solovyev, 2022; López-Hidalgo, 2022, among others). Thus, the new sense can be operationalised as a new cluster of collocations that is present in the more recent corpus and absent in the oldest, e. g., *navegar* ‘to navigate’ typically co-occurs with vocabulary related to the sea, while in its more recent meaning, ‘to browse or search for information on the internet’, it co-occurs with words related to this medium, information, technology, etc. The goal then is to detect these changes and establish a threshold for good precision and recall. A clue to narrow the candidates are those experimenting a substantial increasing of the frequency of the word in the most recent corpus with respect to a reference corpus (Renouf, 2014; López-Hidalgo, 2022). Another clue can be certain changes in syntactic behaviour, as the fact that a noun becomes more frequent as the subject of the verb, when before it was typically working as an object, or vice versa (López-Hidalgo, 2022). For example, the word *ratón* ‘mouse’ is prone to be the subject of the verb more frequently when it refers to the ‘mammal’ than when it refers to the ‘device’. Finally, semantic neology can correlate with the topic of the text (Renouf, 2014; Torres-Rivera et al., 2022), e. g., it is plausible that a word which is usually found in specialised texts and then found in general texts experimented a semantic change, such as *idiota* ‘idiot’ when it escaped from the medical field and started to be found in many other types of texts.

None of these criteria is by itself an indicator of the emergence of a new meaning. The goal, thus, is to search for appropriate combinations of different criteria that can be automatically applied to a set of words and confirmed by human expertise. This is the approach taken by most of the cited studies.

### 3. Materials and methods

As stated in the introduction, our goal is to find semantic neology of the pandemic period, specifically of 2022. As source of the data, we used a corpus from the pre-pandemic period and a corpus of 2022. We compared the collocations in both corpora, following the most common method of previous studies. We also considered frequency and topic as a way of narrowing the list of candidates.

We used Sketch Engine (Kilgarriff et al., 2014) as an interface to test some of the tasks in evaluation and leave for future work the possibility of implementing a more complex system for automatic semantic neology detection. Sketch Engine provides a work interface in which many tasks can be performed semi-automatically to test their validity, and in a second phase we can independently implement them in a fully automatic way.

#### 3.1 *Materials*

We used the Leipzig Corpora Collection (Goldhahn, Eckart & Quasthoff, 2012) as dataset. This collection is built upon texts from news websites, especially using RSS feeds on a daily basis. Each year contains one million sentences. Our corpus contained two sub-corpora: the PreCovid sub-corpus, consisting of news from the three years immediately before the pandemic (2017, 2018 and 2019), and the Covid2022 sub-corpus, with data from the year 2022 (we wanted to have a time span between the two samples) (table 2).

<i>Sub-corpora</i>	<i>Content description and reference</i>	<i>Year</i>	<i>File</i>	<i>Tokens</i>	
Precovid sub-corpus	Leipzig Corpora Collection (2017, 2018, 2019). Spanish newspaper corpus based on material from 2017. Dataset.	2017	spa_news_2017_1M.tar.gz	~22,343,341	~66,965,543
		2018	spa_news_2018_1M.tar.gz	~22,293,686	
		2019	spa_news_2019_1M.tar.gz	~22,328,516	
Covid2022 sub-corpus		2022	spa_news_2022_1M.tar.gz	~22,219,215	~22,219,215
Total					~89,184,758

Table 2. Description of the corpus

### 3.2 Methods

The corpus was uploaded in Sketch Engine with the same structure shown in table 2, that is, a corpus with two sub-corpora. We lemmatised the corpus and obtained the list of nouns. We worked only with the intersection of both sub-corpora, as our target nouns are those which appear in both datasets, and discarded all nouns with less than 100 occurrences, since very few occurrences led to impractical comparisons of collocations. The said intersection, with only those nouns with 100 or more occurrences, is a list of 4,141 nouns. In principle, all nouns in this list are candidates for semantic neology, but it is evident that most of them will not be, since the lexical system does not change that much. The challenge, thus, is how to obtain clues for narrowing the candidates to a more accurate list, as explained in section 2.2.

We compared the collocations of both sections of the corpus using Word Sketch Difference, a Sketch Engine tool that allows to compare collocations of two words of the same corpus, or the same word in different corpora. Word Sketch has different parameters for collocation analysis, but for this study we focused on the combinations of a) the target noun as the subject of a verb, b) the object of a verb (typically the direct object), and c) the adjective collocating with the noun. This is because these are the three parameters that were present virtually always in all the analysed nouns, and this way we had comparable





tion, only when one or more of the new collocates could not be used in the PreCovid section, because they are not equivalents to words in this section, or do not hold an IS-A relation with them. For instance, in the previous example, we find *flexibilizar*, *decretar*, *acabar*, *soportar*... ‘to make flexible, enact, finish, bear...’ as verbs collocating with *confinamiento*. These verbs were not found in the PreCovid sub-corpus, but could have appeared, since they refer to the exact meaning of the noun in this section. This collocation analysis was performed manually and it is an obstacle for full automatization of the procedure. However, software helped us systematise the analysis and make it more objective.

We conducted four different studies of collocations with Word Sketch Difference, trying different ways of targeting the candidates for the analysis:

- *A study of 100 random selected nouns of the intersection list.* We made a simple random selection of nouns for the collocation analysis. We consider this as the simplest baseline we could apply.
- *A study of the 100 most frequent nouns in the Covid2022 sub-corpus.* We analysed the nouns at the top of the frequency list of this sub-corpus because we wanted to test if high frequency correlates with new meanings. According to Hamilton, Leksovec and Jurafsky (2018), high-frequency words tend to be more polysemous, but also more semantically stable. Thus, according to this, we should not find semantic neologisms in high frequency nouns.
- *A study of the 100 nouns in the top of a ranking elaborated considering the ‘salience coefficient’  $s(t)$ ,* which measures the frequency difference between the lemma list of both corpora:

$$s(t) = (\text{Rel. freq. of } t \text{ in the PreCovid corpus}) / (\text{Rel. freq. of } t \text{ in the Covid2022 corpus})$$

This study was conducted to test if a considerable increase of the frequency of the word in the target corpus *in comparison with the reference corpus* is linked to the emergence of one or more meanings. This comparison provides keywords of the target corpus that can be linked to the topic of the text and can be considered for semantic neology detection, as advanced in section 2.2.

- *A study of the keywords of the Covid2022 sub-corpus, extracted using the Key Word tool in Sketch Engine.* According to the Sketch Engine guidelines, “keywords are words (single-token items), that appear more frequently in the focus corpus than in the reference corpus” (<https://www.sketch-engine.eu/guide/keywords-and-term-extraction/>). We assume then that

keyword extraction works in a similar way as it was described in (c), but we do not know the specific coefficient.

In study (d), we added a second round of analysis. We manually assigned a topic to the 500 nouns with the highest score in this ranking and used them as seed words to compile different topic-specific sub-corpora of the Covid2022 section. With Sketch Engine, one can search concordances and save them as a sub-corpus. In our case, we compiled four sub-corpora, each one of them created by searching nouns of the previous list related to the most salient topics: the pandemic, the Ukraine war, the new technologies, and economics. For example, we assigned the topic *Ukraine war* to words such as *Ucrania*, *Zelenski*, *Kiev*, and *Putin*. This way, the new keywords that were found in this second round were even more topic-related than the previous ones.

## 4. Results and discussion

### 4.1 *Synthesis of results and comparison of the studies*

Results of the studies described in section 3.2 are synthesised in table 3.

We analysed 604 nouns in total (excluding duplicates). Numbers in the sub-corpora studies in (d) are not round because we analysed the common nouns that were present in the top 500 words in the keywords list, which contains all parts of speech and a significant number of proper nouns. The lists of keywords extracted from texts related to the Ukraine war and technologies of communication contain plenty of proper nouns, which makes common nouns less frequent and less significant.

Considering the four strategies combined, we detected seven semantic neologisms. The significance of these results in terms of quantity is difficult to establish because we lack comparable studies in Spanish. Spanish formal neologisms of the same period seem to be much more frequent, e. g., Bueno and Freixa (2022) report 209 formal neologisms during 2020 and the first half of 2021 –we do not know the size of the corpora that was used for this study, though. Meanings of words can be infrequent even when they are well-established –for Spanish, see lexical databases such as SenSem (Vázquez & Fernández Montraveta, 2008) or Verbario (Renau, Nazar, Castro et al., 2019), which provide the frequency of each meaning of the verbs. For that reason, we should not be surprised to see semantic innovations of words kept

under the radar, which would confirm that, for semantic neology, it is especially relevant working with extensive datasets.

Type of study	N	Examples of nouns in the sample	Confirmed semantic neologisms
a) Simple random sample	100	(Random 10) <i>protección, clip, cumpleaños, munición, audio, detección, regalo, periodo, ser</i>	
b) Most frequent nouns in Covid2022	100	(Top 10) <i>año, día, persona, país, parte, vez, caso, momento, millón, tiempo</i>	<ul style="list-style-type: none"> <li>• <i>pandemia</i> 'pandemic'</li> </ul>
c) Top nouns in the 'salience coefficient' ranking	100	(Top 10) <i>pandemia, contagio, viruela, mascarilla, invasión, vacunación, vacuna, dosis, ucraniano, variante</i>	<ul style="list-style-type: none"> <li>• <i>pandemia</i> 'pandemic'</li> <li>• <i>positividad</i>, 'positivity'</li> <li>• <i>refuerzo</i>, 'boost'</li> <li>• <i>variante</i> 'variant'</li> </ul>
d) Top nouns in the keywords ranking	100	(Top 10 common nouns) <i>Covid-19, pandemia, Covid, coronavirus, ómicron, ucraniano, viruela, contagio, ucranio, SARS-CoV-2</i>	<ul style="list-style-type: none"> <li>• <i>pandemia</i> 'pandemic'</li> <li>• <i>positividad</i>, 'positivity'</li> <li>• <i>refuerzo</i>, 'boost'</li> </ul>
Sub-corpus: pandemic	181	(Top 10 common nouns) <i>pandemia, Covid-19, coronavirus, Covid, ómicron, mascarilla, contagio, confinamiento, cubrebocas, positividad</i>	<ul style="list-style-type: none"> <li>• <i>burbuja</i> 'bubble'</li> <li>• <i>pauta</i> 'scheme'</li> <li>• <i>positividad</i> 'positivity'</li> </ul>
Sub-corpus: Ukraine war	63	(Top 10 common nouns) <i>ucraniano, ucranio, invasión, misil, rublo, contraofensiva, tiroteo, tropa, bombardeo, oligarca</i>	
Sub-corpus: economy	121	(Top 10 common nouns) <i>encarecimiento, inflación, cripto, pandemia, Covid-19, Covid, coronavirus, estanflación, recesión, aceleración</i>	<ul style="list-style-type: none"> <li>• <i>galope</i> 'gallop'</li> </ul>
Sub-corpus: technologies of communication	42	(Top 10 common nouns) <i>presencialidad, teletrabajo, criptomonedas, posteo, videollamada, tuit, pandemia, Covid-19, coronavirus</i>	
Total	604		7

Table 3. Results of the analysis per type of study

As observed in table 3, the most accurate study at providing good candidates for semantic neologisms was the 'salience coefficient' (c). Random sam-

ple of nouns (a) does not provide results, and analysing high frequent nouns (b) does not lead to good results either, as in our study it only provides a neological metaphor of *pandemia* ‘pandemic’ (see section 4.2) that we also found with strategy (c). Poor results of study (b) are consistent with Hamilton et al. (2018) (see section 3.2), that is, high-frequency words tend to remain stable. The most common case in (a) and (b) studies is that nouns share most of the collocates in both sub-corpora, and the collocates that are exclusive of the Covid2022 sub-corpus are scarce and semantically similar to the previous ones.

Study (c) shows that using a coefficient of the difference between relative frequency in both corpora was the best option in the present investigation. This means that nouns that have experienced an increase in relative frequency have the potential to acquire new meanings, regardless of their absolute frequency figures. Figure 2 shows the nouns with a coefficient  $>2$ , or those 22 words with the highest coefficient. Out of the 7 neologisms found 4 are present in this top 22.

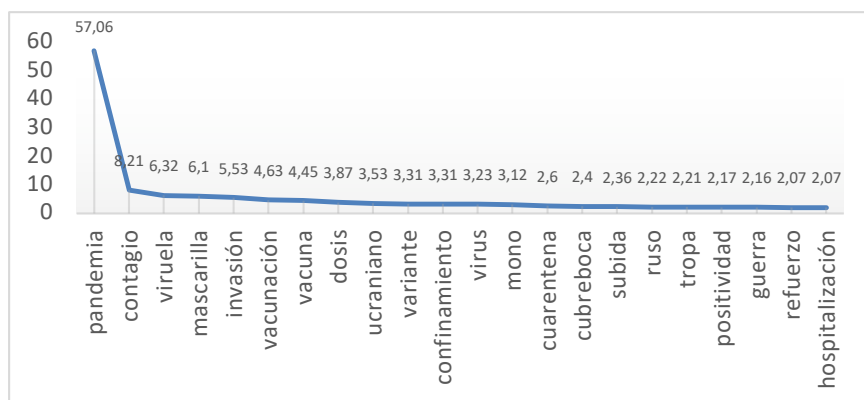


Figure 2. Nouns with a salience coefficient  $>2$ . Translation: *pandemic, contagion, pox, mask, invasion, vaccination, vaccine, dose, Ukrainian, variant, confinement, virus, monkey, quarantine, mask, rise, Russian, troop, positivity, war, booster, hospitalisation*

We also analysed the top 100 key nouns in the list provided by Sketch Engine (d). As said, this list contains all parts of speech and a considerable number of proper names. We did not find any semantic neologisms among these 100 nouns that were absent from (c). However, study (d) shows that it is worth analysing corpus data focusing on topics, as we detected 3 of the 7 new meanings by creating keywords lists with topic-related sub-corpora: *pauta* (as ‘vaccination schedule’), *burbuja* (as ‘sanitary bubble’), and *galope* (as ‘inflation

gallop'). If we observe all 7 new meanings, all of them are related to a specific topic: pandemic ( $n = 6$ ) or economics ( $n = 1$ ). Results of (c) and sub-corpora of study (d) are consistent with studies focusing on topic-specific areas for neology detection, such as López-Hidalgo (2022), who searched on news related to pandemic only and found 4 semantic neologisms in 24 studied nouns.

#### 4.2 Analysis of semantic neologisms

A description of the semantic neologisms found is shown in table 4.

López-Hidalgo (2022) also finds *refuerzo* and *positividad* as confirmed candidates in a corpus of Chilean press, also by using collocation analysis. He does not find *burbuja* because it was not used in Chilean Spanish, and, as an opposite case, he finds *pase* 'pass', as 'a document that authorised moving outside home', because it was a Chilean use only. As for the rest of results, we lack studies to compare our findings.

A description of each found semantic neologism is offered below.

*Burbuja* 'bubble'. In our study, this new meaning is related to collocations that clearly differentiate from the ones in the PreCovid data, e. g., *burbuja olímpica* 'Olympic bubble', *burbuja social* 'social bubble', *grupo burbuja* 'bubble group', etc. It has the same meaning as in English (Saladrigas et al., 2021, s. v. *social bubble*), which indicates that some new meanings are loans and, thus, it makes us consider the possibility to explore contrastive approaches to semantic neology. This new meaning was already incorporated in the *Diccionario de la lengua española*, DLE (RAE & ASALE, online), which includes also the compound *burbuja social* as sub-lemma.

Metaphorical use of *burbuja* is not new, and is extensively used in economy to describe a 'situation where the price of an asset rises to an exaggerated level' (Renau et al., 2020). It is also used in Spanish to refer to a 'situation where a person lives isolated and disconnected from reality, especially from the problems of other people' (DLE, meaning 2). The new meaning of the noun creates a new connection between the source domain, 'bubble', and the target domain, 'social group', that is, a new metaphor.

*Galope* 'gallop'. The new, metaphorical use of *galope* is used only in one context (see table 4) in our corpus. If we make a Google search restricted to year 2022, we find many similar contexts in different media, such as *el galope de los precios* 'the price gallop', *el galope inflacionario* 'the inflationary gallop', etc. Indeed, a new meaning can be infrequent even in large datasets (Cook &

Noun	New meaning	Examples of usage	Type of meaning change
<i>burbuja</i> 'bubble'	'A small group of people that are not necessarily cohabitants but maintain a social relation on a regular basis'	"...el uso de doble mascarilla y mantener su <i>burbuja</i> social" '...the use of double mask and maintain their social <i>bubble</i> '	metaphor
<i>galope</i> 'gallop'	'A rapid, uncontrolled development of an economic event'	"Se empalidecen cuando se corrobora el <i>galope</i> inflacionario" 'They turn pale when the inflationary <i>gallop</i> is corroborated'	metaphor
<i>pandemia</i> 'pandemic'	'A very harmful situation that affects a large number of people equally'	"La gran <i>pandemia</i> es la salud mental de todos y la incertidumbre por la crisis en los precios" 'The great pandemic is everyone's mental health and the uncertainty due to the price crisis'	metaphor
<i>pauta</i> 'scheme'	'A plan for administering a series of vaccine doses in an established period of time as a preventive treatment'  'The series of vaccine doses themselves'	"Ha invitado a la población a completar su <i>pauta</i> de vacunación" 'She/he invited the population to complete their vaccination <i>scheme</i> '. "La monarca recibió la <i>pauta</i> completa de la vacuna contra el coronavirus más una dosis de refuerzo" 'The monarch received the full coronavirus vaccination <i>scheme</i> , plus a booster dose'	specialisation and metonymy (respectively)
<i>positividad</i> 'positivity'	'Rate or percentage of positive cases of all tests performed' (see Drury, 2021, in Saladrigas et al., 2021, s. v. <i>positivity rate</i> )	"En las últimas 24 horas se registró una positividad a nivel nacional de 16,2 %" 'In the last 24 hours, a positivity rate of 16.2 percent was registered nationwide'	metonymy
<i>refuerzo</i> 'boost'	'Vaccine booster dose'	"Se puede administrar un segundo <i>refuerzo</i> a adultos de 80 años o más" 'A second booster may be given to adults 80 years and older'	metonymy
<i>variante</i> 'variant'	"A virus whose phenotype differs from the original wild-type strain but the genetic basis for the difference is not known" (from Cann, 2005, in Saladrigas, Navarro, Gómez Polledo & Munoa, 2021, s. v. <i>variante</i> )	"...PCR para casos sospechosos, capaces de detectar la nueva <i>variante</i> " '...PCR for suspected cases, able to detect the new <i>variant</i> '	specialisation

Table 4. Synthesis of the description of new meanings

Hirst, 2011), thus, it is expected to find this timid appearance in the corpus when working with semantic neologisms. The meaning of *galope* is new, but the underlying metaphor seems to be well established, as it is present also in the adjective *galopante* ‘galloping, rapidly increasing’ (meaning 2 in *DLE*). This makes us wonder how different parts of speech interact with each other in semantic neology when sharing a common metaphor.

*Pandemia* ‘pandemic’. A small group of contexts show the new metaphorical meaning of this noun, highlighted by collocates such as *macrista* ‘Macrist’ or *económica* ‘economic’. These collocates clearly differ from the literal meaning of the noun. In Spanish and other languages, *pandemia* (or equivalents) have been the target domain of many metaphors, due to the need of processing and understanding the new reality that, as stated in the introduction, was drastically new for human kind (Brugman, Droog, Reijnierse et al., 2022). Our data shows how, at the same time, the concept can also be the source domain for other harmful, catastrophic situations. We have an already stabilised precedent very similar to *pandemia: epidemia* ‘epidemic’, with a metaphorical meaning denoting an ‘evil or damage that spreads intensely and indiscriminately’ (*DLE*, meaning 2 of the entry), thus, even when literal meanings of both nouns are not synonyms, metaphorical meanings seem to be. It is soon to know which will be the final destiny of the new metaphor present in *pandemia*, but we could be witnessing a ‘cognitive competition’ between *pandemia* and *epidemia*.

*Pauta* ‘scheme’. The semantic neology of this noun is twofold. On the one hand, it was used during the pandemic to refer to the ‘*plan* for giving a series of vaccine doses’, that is, to describe an EVENT, and on the other hand, to refer to the ‘vaccine *doses*’ themselves, that is, the ARTEFACT. Both meanings may have been used before, but we do not find them in our data nor in the *DLE*. *Pauta* is found in the PreCovid sub-corpus with a general meaning to refer to a ‘series of instructions that must be followed in a task’ (meanings 3 and 4 in *DLE*). For that reason, we consider the ‘*plan*’ new meaning a specialisation, and the ‘*artefact*’ new meaning as a metonymy linked to it.

This is, thus, a case of EVENT/ARTEFACT regular polysemy (Apresjan, 1974), which is typical of many nouns and languages, when the artefact is indissoluble from the event, as in this case (Renau et al., 2022: 159-167). The EVENT reading can be found in contexts such as *seguir/modificar/completar... la pauta de vacunación* ‘to follow/modify/complete... the vaccination scheme’. The ARTEFACT reading can be found in contexts such as *tener/contar con/recibir la pauta completa (de vacunación)* ‘to have/count with/receive the complete

(vaccination) scheme'. We found here, thus, that a pair of meanings linked to each other by regular polysemy emerge at the same time. We consider this a relevant result for the theory of regular polysemy, because it could proof the systematicity and closeness of these two senses appearing together, as part of the same complex semantic type (Pustejovsky & Batiukova, 2019: 175-177). It is also relevant for the theory of semantic neology, as it shows that regular polysemy plays an important role on it (cf. Lombard, Huyghe & Gygax, 2021).

*Positividad* 'positivity'. In the PreCovid sub-corpus, *positividad* is used as the 'positive attitude towards a situation', as in "abordó la importancia de un liderazgo basado en la *positividad* y en el mindfulness" 'he/she addressed the importance of leadership based on *positivity* and mindfulness'. In the Covid2022 sub-corpus, where there is a substantial increment in the frequency of this noun, *positividad* is a 'rate' and collocates with many words that typically co-occur with numbers, e. g., *bajar*, *incrementar*, *descender*, *acumular* 'decrease, increase, descend, accumulate' (verbs), *alta*, *baja*, *semanal*, *diaria* 'high, low, weekly, diary' (adjectives). We could consider this a case of metonymy QUALITY / NUMERICAL VALUE where the 'quality' meaning is very general or intensional. Thus, regular polysemy is involved in this case as well. We can add here the role or the suffix *-(i)dad*, which in Spanish carries the same regular polysemy pattern as in other nouns, such as *velocidad* 'speed' or *profundidad* 'depth'.

Figure 3 shows nouns collocating with *positividad*. In contrast with figure 1, one can observe that there is much more semantic difference between nouns in PreCovid data and in Covid2022 data in this case than in *confinamiento*.

*Refuerzo* 'boost'. This noun is used in the PreCovid sub-corpora with its general meaning of the 'event, result, or agent that boosts, strengthen, or reinforces something (concrete or abstract)', e. g.:

- EVENT: "...concretar el *refuerzo* de las fuerzas federales" ('...concretise the *reinforcement* of the federal forces')
- RESULT: "Se exigió que se aclare el destino del *refuerzo* presupuestario por \$20 millones" ('It was demanded that the destination of the budget *reinforcement* for \$20 million was clarified')
- AGENT: "Fue necesario enviar *refuerzos* policiales para controlar la situación" ('It was necessary to send police *reinforcements* to control the situation').





Figure 3. Nouns collocating with *positividad* 'positivity' in both corpora.  
Visualisation: Sketch Engine

This is, then, a good example of a triadic regular polysemous structure, which, with different combinations, is found in many languages and nouns (Renau et al., 2022: 160-166). In the Covid2022 sub-corpus, we find a new meaning linked to this regular polysemy pattern: the ARTEFACT that is used to boost or strengthen something, specifically, a 'vaccine dose that is applied to boost the previous one(s)'. We find here, thus, as in *pauta* and *positividad*, that regular polysemy is involved in the emergence of the new meaning. Additionally, as with *pauta*, it is possible that this meaning already existed in Spanish, but we do not find it in our data (nor in the *DLE*), and it was certainly not as popular as during the pandemic, which was indeed used in non-specialised contexts.

*Variante* 'variant'. This case is similar to the previous one, as *variante* indeed already existed as a specialised term, but it is not present in the Covid2022 data, probably because it was not common in the general language. In addition, *variante*, as in English, was highly semantically codified during pandemic, with very specific, regulated compounds such as *variante preocupante* 'variant of concern', *variante de interés* 'variant of interest', *variante de gran consecuencia* 'variant of high consequence', and with specific names for variants of the virus, such as *alfa*, *beta*, *ómicron*, etc. (Saladrigas et al., 2021, s. v. *variant*). Considering the general meaning of the noun as 'a thing that is a variation of another pre-existing thing', we consider this a case of specialisation. In the Covid2022 analysis, we find many collocations for this semantic neologism that are very specific to the semantic field of 'virus' or 'disease',

such as *aparecer*, *detectar*, *surgir*, *predominar* ‘to appear, detect, emerge, predominate’ (verbs) or *sudafricana*, *dominante*, *circulante*, *recombinante*, *alfa*, *contagiosa* ‘South African, dominant, circulating, recombinant, alpha, contagious’ (adjectives), etc.

## 5. Conclusions and future work

In this paper, we showed different strategies for semantic neology detection, using a well-established corpus-based method to compare two datasets of different time periods. We considered different frequency measures to improve the list of possible candidates and used collocations as the main orientation to discriminate a new meaning from existing ones. The task was partially automatized and allowed us to access thousands of contexts through collocations statistics. Many tasks of the process were, however, manual, as the automatic system does not differentiate between changes in collocations in general and those due to semantic neology. This is a key task that we leave for future work. Even when results are preliminary, they are relevant, since we did not have virtually any similar attempt in Spanish, as far as we know.

The study showed that working with keywords and focusing on topics (as in studies c and d) gives the best results. Even with limitations and manual tasks, this is a feasible activity to incorporate to the process of dictionary revision and updating. In the same line, results of this study make us think that it would be positive to expand the perspective and focus on *revision* and *updating* of the dictionary in a more general way: results show potential for improving examples and definitions, as the context where words are used may change even when there are no new meanings to add to the entry. The case of *confinamiento* ‘confinement’ is one of them, with context related to distant historical events in the reference corpus and with the pandemic in the target corpus (see section 3.2).

Finally, another important finding of our study is that, even with few cases analysed, we observe that regular polysemy plays a central role in semantic neology. We can hypothesise that nouns holding a regular polysemy pattern are more prone to develop a new meaning than others. This would place regular polysemy in a privileged position in semantic neology. We leave this study for future research.

## Acknowledgments

This study was funded by the Fondecyt Regular project nr. 1231594, ANID, Chile government, and by project ESMAS-ES+ (PID2022-137170OB-I00) funded by MCIN/AEI//FEDER “Una manera de hacer Europa”. Many thanks to Rogelio Nazar for his help in data analysis.

## References

- Adelstein, A. 2017. Neología semántica y voces especializadas en la conformación del léxico común del español de la Argentina. In *Los argots hispánicos. Convergencias y divergencias en las hablas populares hispanoamericanas*. Ediciones de la UNLa, 103-122.
- Adelstein, A. & Kuguel, I. 2008. *De salarizado a corralito, de carapintada a blog: Nuevas palabras en veinticinco años de democracia*. Buenos Aires: Biblioteca Nacional / Universidad Nacional de General Sarmiento.
- Álvarez-Bolado Sánchez, C. M. & Álvarez de Mon Rego, I. 2013. Semantic neology in the domain of videogames in Spanish. *Ibérica* 25: 63-84.
- Apresjan, J. D. 1974. Regular polysemy. *Linguistics* 12(142).
- Barcelona, A. 2003. Introduction: Cognitive theory of metaphor and metonymy. In Barcelona, A. (Ed.) *Metaphor and Metonymy at the Crossroads: A Cognitive Perspective*. Berlin / New York: Mouton / DeGruyter, 1-28.
- Barque, L.; Haas, P., & Huyghe, R. 2018. Polysémie régulière et néologie sémantique: Constitution d'une ressource pour l'étude des sens nouveaux. *Neologica. Revue Internationale de Néologie* 12: 91-108.
- Bochkarev, V. V.; Khristoforov, S. V.; Shevlyakova, A. V., & Valery D. Solovyev. 2022. Neural network algorithm for detection of new word meanings denoting named entities. *IEEE Access* 10: 68499-68512. doi: <https://doi.org/10.1109/access.2022.3186681>
- Brugman, B. C.; Droog, E.; Gudrun Reijnierse, W.; Leymann, S.; Frezza, G., & Renardel de Lavalette, K. I. 2022. Audience perceptions of COVID-19 metaphors: The role of source domain and country context. *Metaphor and Symbol* 37(2): 101-113.
- Bueno, P. J. & Freixa, J. 2022. Lexicographic detection and representation of Spanish neologisms in the COVID-19 pandemic. In *Lexicography of coronavirus-related neologisms*. De Gruyter, 73-92. doi: <https://doi.org/10.1515/9783110798081-005>
- Cabré, M. T. & Nazar, R. 2012. Towards a new approach to the study of neology. *Neologica* 6. *Revue Internationale de Néologie*. doi: <https://doi.org/10.15122/is-bn.978-2-8124-4232-2.p.0063>
- Colman, L. 2016. Sustainable lexicography: Where to go from here with the ANW (*Algemeen Nederlands Woordenboek*, An Online General Language Dictionary of

- Contemporary Dutch)? *International Journal of Lexicography* 29(2): 139-155. doi: <https://doi.org/10.1093/ijl/ecw008>
- Cook, P. & Hirst, G. 2011. Automatic identification of words with novel but infrequent senses. *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*, 265-274. <https://aclanthology.org/Y11-1028>
- Galanes-Santos, I. 2019. El redactor periodístico ante el neologismo. Estrategias comunicativas para reexpresar los conceptos de la crisis económica mundial. *MonTI. Monografías de Traducción e Interpretación* 5: 121-145. doi: <https://doi.org/10.6035/MonTI.2019.ne5.5>
- Geeraerts, D. 2010. *Theories of Lexical Semantics*. Oxford University Press.
- Gérard, C. & Kabatek, J. 2012. Introduction: La néologie sémantique en questions. *Cahiers de Lexicologie* 100: 11-36.
- Goldhahn, D.; Eckart, T., & Quasthoff, U. 2012. Building large monolingual dictionaries at the leipzig corpora collection: from 100 to 200 languages. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation*. Istanbul: ELRA, 759-765.
- Guslyakova, A.; Valeeva, N., & Vatkova, O. 2020. “Green neology” and its role in environmental education (a media discourse approach). *E3S Web of Conferences* 169, 05004. doi: <https://doi.org/10.1051/e3sconf/202016905004>
- Hamilton, W. L.; Leskovec, J., & Jurafsky, D. 2016. Diachronic word embeddings reveal statistical laws of semantic change. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 1489-1501. doi: <https://doi.org/10.18653/v1/P16-1141>
- Harari, Y. N. 2020. Yuval Noah Harari: “Toda crisis es también una oportunidad”. *Correo de la Unesco* 3. <https://es.unesco.org/courier/2020-3/yuval-noah-harari-toda-crisis-es-tambien-oportunidad>
- López-Hidalgo, B. 2022. *Neologismos semánticos nominales de la pandemia por covid-19: Análisis contrastivo de patrones de uso en un corpus diacrónico de prensa chilena* [MA thesis]. Pontificia Universidad Católica de Valparaíso.
- Kilgarriff, A.; Baisa, V.; Bušta, J.; Jakubiček, M.; Kovář, V.; Michelfeit, J.; Rychlý, P., & Suchomel, V. 2014. The Sketch Engine: Ten years on. *Lexicography* 1(1): 7-36. doi: <https://doi.org/10.1007/s40607-014-0009-9>
- Klosa-Kückelhaus, A. 2013. The lexicographical process (with special focus on online dictionaries). In *Supplementary Volume Dictionaries. An International Encyclopedia of Lexicography*. De Gruyter Mouton, 517-524. doi: [https://doi.org/10.1515/9783110238136\\_517](https://doi.org/10.1515/9783110238136_517)
- Klosa-Kückelhaus, A. 2022. Lexicography for society and with society – COVID-19 and dictionaries. *Dictionaries and Society. Proceedings of the XX EURALEX International Congress, 12-16 July 2022, Mannheim, Germany, 113-128*. <https://euralex.org/publications/lexicography-for-society-and-with-society-covid-19-and-dictionaries/>
- Klosa-Kückelhaus, A. & Kernerman, I. 2020. Global viewpoints on lexicography and neologisms: An introduction. *Dictionaries: Journal of the Dictionary Society of North America* 41(1): 1-9. doi: <https://doi.org/10.1353/dic.2020.0001>

- Klosa-Kückelhaus, A. & Kernerman, I. (eds.). 2022. *Lexicography of Coronavirus-related Neologisms*. De Gruyter. doi: <https://doi.org/10.1515/9783110798081>
- Lombard, A.; Huyghe, R., & Gygax, P. 2021. Neological intuition in French: A study of formal novelty and lexical regularity as predictors. *Lingua* 254, 103055. doi: <https://doi.org/10.1016/j.lingua.2021.103055>
- Lombard, A.; Huyghe, R.; Barque, L., & Gras, D. 2023. Regular polysemy and novel word-sense identification. *The Mental Lexicon* 8(1). doi: <https://doi.org/10.1075/ml.21002.lom>
- Pustejovsky, J. & Batiukova, O. 2019. *The Lexicon*. Cambridge University Press.
- Ramírez-Henríquez, M. 2020. *La polisemia regular acción/efecto en los neologismos por derivación en el periodo 2003-2020* [BA thesis]. Pontificia Universidad Católica de Valparaíso.
- Real Academia Española & Asociación de Academias de la Lengua Española. Online. *Diccionario de la lengua española*. <https://dle.rae.es>.
- Renau, I.; Castro, A.; López-Hidalgo, B.; Martínez, F.; Obreque, J.; Ramírez-Henríquez, M.; Soto, E., & Terry, A. 2022. Tipos de polisemia regular: Revisión sistemática y análisis de corpus. En Lima Moreira, G.; Alves, L., & Alves, I. M. (eds.) *Pesquisas em Lexicologia, Lexicografia e Terminologia*. [https://www.ponteseditores.com.br/loja/index.php?route=product/product&product\\_id=1787](https://www.ponteseditores.com.br/loja/index.php?route=product/product&product_id=1787)
- Renau, I.; Nazar, R.; Castro, A.; López, B., & Obreque, J. 2019. Verbo y contexto de uso: Un análisis basado en corpus con métodos cualitativos y cuantitativos. *Revista Signos* 52(101): 878-901. doi: <https://doi.org/10.4067/S0718-09342019000300878>
- Renau, I.; Nazar, R., & Lecaros, V. 2020. La evolución de las marcas ortográficas y tipográficas en los procesos de lexicalización de neologismos: Un estudio en el vocabulario de la crisis económica en prensa española. *Revista Española de Lingüística Aplicada* 33(1): 227-277. doi: <https://doi.org/10.1075/resla.17028.ren>
- Renouf, A. 2014. Semantic Neology: The challenges for automatic identification. *Neologica. Revue Internationale de Néologie* 8. <https://doi.org/10.15122>
- Sajous, F. 2022. Using Wiktionary revision history to uncover lexical innovations related to topical events: Application to Covid-19 neologisms. In Klosa-Kückelhaus, A. & Kernerman, I. (eds.) *Lexicography of Coronavirus-related Neologisms*. De Gruyter, 275-306. doi: <https://doi.org/10.1515/9783110798081-014>
- Saladrigas, M. V.; Navarro, F. A.; Gómez Polledo, P., & Munoa, L. 2021. *Diccionario de covid-19 (EN-ES)*. Tremédica-Cosnautas. <https://www.tremedica.org/tremediteca/glosarios/glosario-de-covid-19-en-es>.
- Sanmartín Sáez, J. 2016. Sobre neología y contextos de uso: Análisis pragmlingüístico de lo ecológico y de lo sostenible en normativas y páginas web de promoción turística. *Ibérica* 31, art. 31. <https://revistaiberica.org/index.php/iberica/article/view/194>
- Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Torres-Rivera, A. & Torres-Moreno, J.-M. 2020. Detecting new word meanings: a comparison of word embedding models in Spanish. *Conference En Recherche*

- d'Informations et Applications* (CORIA 2019, France), 1-16. <https://arxiv.org/pdf/2001.05285.pdf>
- Vázquez, G., A. & Fernández Montraveta, A. 2009. Ampliación del Banco de Datos de Verbos del español SenSem. In Cantos, P. & Sánchez, A. (eds.) *A Survey on Corpus-based Research. Panorama de investigaciones basadas en corpus*. Murcia: Universidad de Murcia, 957-969.
- Vila Vigoni Theses. 2018. <https://www.emlex.phil.fau.eu/about-us/publications/other/>
- Yallop, Colin. 2004. Words and meaning. In Halliday, M. A. K.; Cermakova, A.; Teubert, W., & Yallop, C. (eds.) *Lexicology and Corpus Linguistics*. London: Continuum, 23-72.