

**Treball de fi de màster *de Recerca***

**Neology in English and Spanish. 2005-2015.  
Lexicological characteristics and lexicographic representation.**

**Nom i Cognoms:** Yingfei Lu

**Màster:** Lingüística Teòrica i Aplicada

**Edició:** 2022-2023

**Director o directora:** Dr. Janet DeCesaris

**Any de Defensa:** 2023

**Col.lecció:** Treballs de fi de màster

**Departament de Traducció i Ciències del Llenguatge**

## Abstract

Our comparative study focuses on examining the word formation processes of English and Spanish neologisms collected from *Among The New Words in American Speech* (English data) and *El Banco de Neologismos* from *Centro Virtual Cervantes* (Spanish data) between 2005 and 2015. By analyzing lexicological characteristics and lexicographic representation, our objective is to uncover both similarities and differences between these two languages and these two approaches to neology.

To analyze English neologisms, we draw upon Bauer's (1983) influential work *English Word-formation* and the collaborative contribution of Bauer and Huddleston (2002). For Spanish neologisms, we incorporate the comprehensive reference *Nueva gramática de la lengua española* by *Real Academia Española* (2009), along with Cabré's (2006) theory on *La clasificación de neologismos: una tarea compleja*, which guides the *Observatori de Neologia* (OBNEO) project.

Our findings reveal distinct differences in the productivity of word formation processes between English and Spanish. For instance, English neologisms primarily manifest as compounds and phrases, while Spanish exhibits significant productivity in suffixation and neoclassical compounds. Additionally, borrowing words in Spanish demonstrate a prominent presence, as opposed to the low amount of loanwords in English. These disparities can be attributed to various factors, including the differing nature and purpose of the projects, the inherent characteristics of the languages, and their historical origins. Furthermore, the utilization of different classification criteria contributes to the observed results. To address the challenges encountered during our study, we suggest revising English compound definitions and conducting comprehensive re-evaluations of word formation process information in *Banco de Cervantes* to ensure improved accuracy.

Moreover, our study emphasizes the importance of exploring unconventional neologisms that challenge traditional morphological theories. It highlights the significant role of word elements in generating new words and calls for future research to further investigate their implications.

Keywords: English-Spanish neologisms, word formation process, comparative linguistics, lexicology, lexicography

## Acknowledgements

Every time I open a new academic reading, my greatest anticipation lies not in discovering new theories or findings, but rather in immersing myself in the acknowledgements section. To be honest, I am someone who always says *thank you* and *sorry*. But someone taught me that actions speak louder than words. Nevertheless, it is evident that I am not someone who merely pays lip service. I possess courage and determination, investing maximum effort to accomplish every task I set my sights on—just like running a marathon or scaling a mountain for five consecutive hours, or even completing this very thesis.

There are so many people I am grateful for. But my greatest and most sincere thanks go first to my supervisor, Dr. Janet DeCesaris. I am deeply grateful to have been your student during your final teaching year. I will miss every meeting in your office where we talked as both teachers and friends, discussing the weather in Spain, China, and even Norway, the Netherlands, and the United States. Amidst laughter, we would also engage in free collisions of academic ideas, and you always gave me a lot of inspiration. Whether it is neologism research, morphology or lexicography, you are my most important guide, patiently answering all my questions and giving me the greatest support, encouragement and help in both academics and life without reservation. Without Janet, I could not make such progress in my academic field and even finish this paper.

Furthermore, I would like to express my heartfelt appreciation to Dr. Elisenda Bernal and Dr. Alba Milà-García for accepting me as their PhD student so that I can continue to pursue my path in lexicography. My gratitude also goes to Dr. Carmen Pérez-Vidal for her encouragement and guidance in second language acquisition research. And to my undergraduate supervisor Dr. Fan Wu and Dr. Yan He from Guangdong University of Foreign Studies, who first introduced me to lexicology research.

During this nearly one year of study in Spain, I have gained growth and friendship. My heartfelt thanks go to Zhao Zeng for always being there for me, and to my roommate Yunke He, who has been with me since the day I left China. I also want to extend my gratitude to Yanji Chen who always tells me that I can do it when I feel like I can't. And to my undergraduate classmate Yuqian Peng, who is also studying in Barcelona, we witnessed the tenderness of tulips in Keukenhof Gardens; as well as to my American roommates Lizzy Daly and Ellie Hawkey. I miss our nightly dinners, conversations and hot pot sharing nights. And to my new roommate Lisa Ji, who would bring me potato pancakes when I was writing my thesis frantically. Additionally, special thanks go to Catalina Corbo for providing me with my first home in Barcelona and her dog Yai who always likes to lie down on the sofa next to me.

Moreover, I extend my gratitude to my dearest friends in China, Qilin Zou, LingMin Liu and Keqing Li. Your unwavering support and companionship have been a guiding light during my darkest hours.

I would also like to express my sincere gratitude to Wojtek's father and Hania for making me feel at home in Norway with their meticulous care, attention and love. The blessing of the lucky clover allows me to continue being a lucky girl. More importantly, to Wojciech Jan Piela, your company and support mean a lot to me.

Last but not least, I want to express my deepest appreciation to my father Guibo Lu and my mother Chaohua Lin, my beloved younger brother Ruijun Lu, my grandparents: Weiliang Lu, Fengqun Cen, Zhen'ai Lin and Jianyin Feng, my aunts Liufen Lu and Liuyan Lu, and uncles Haiguang Fang and Guosen Lu, my two closest cousins: Zijian Fang and Junyong Lu, and my sisters-in-law Huan'en Yao and Zhuoyi Dai, my nieces Jingyuan Fang and Yueying Lu, and so on. To my entire extended family, you have taught me how to love others and shaped me into a better person.

There was a moment when I felt that ancient Chinese poetry was so graceful and romantic during this unforgettable May month and I would like to cite one part of the poem 《水调歌头》 (*Prelude to Water Melody*) until the end.

人有悲欢离合，月有阴晴圆缺。

*Men have sorrow and joy; they part or meet again;*

*The moon is bright or dim and she may wax or wane.*

此事古难全，

*There has been nothing perfect since the olden days.*

但愿人长久，千里共婵娟。

*So let us wish that man*

*Will live long as he can!*

*Though miles apart, we'll share the beauty she displays.*

(Translated by a Chinese famous scholar: Yuanchong Xu 许渊冲)

So far, to my whole family, to all my friends, to all the persons I care about, and to the hardworking one who works day by day in Norway, Vestland.

Thank you, from the bottom of my heart, truly.

Thank you, for lighting up my world.

from Yingfei Lu 陆盈斐

## Table of contents

Abstract.....	i
Acknowledgements.....	ii
1. Introduction.....	1
2. Historical Review of the Two Neologisms Projects .....	3
2.1 <i>Among the New Words - American Speech</i> .....	3
2.2 <i>El banco de neologismo - Centro Virtual Cervantes</i> .....	5
3. Methodology .....	6
3.1 Data collection: English.....	6
3.2. Data collection: Spanish .....	11
3.2 Classification of data.....	15
4 Analysis of English Neologisms in <i>Among the New Words</i> .....	17
4.1 Formal neologisms.....	17
4.1.1 Difficulties of the classification .....	20
4.1.2 Suffixation.....	25
4.1.3 Prefixation.....	27
4.1.4 Compounding.....	30
4.1.5 Blending.....	31
4.1.6 Neoclassical compounding .....	33
4.1.7 Conversion .....	34
4.1.8 Other minor word formations .....	35
4.2 Semantic neologisms .....	36
4.3 Lexicalization of phrases .....	39
4.4 Borrowing .....	39

4.5 Other characteristics: the role of analogy and word elements .....	40
5 Analysis of Spanish Neologisms in <i>Banco de Cervantes</i> .....	42
5.1 Formal neologisms.....	42
5.1.1 Suffixation.....	44
5.1.2 Prefixation.....	46
5.1.3 Compounding.....	47
5.1.4 Neoclassical compounding .....	48
5.1.5 Conversion .....	50
5.1.6 Other minor word formation.....	51
5.2 Semantic neologisms .....	53
5.3 Lexicalization of phrases .....	56
5.4 Borrowing .....	57
5.5 Other characteristics: syntactic neologisms .....	60
6. Comparative Analysis of English and Spanish Neologisms in Two Projects .....	61
6.1 Formal neologisms.....	61
6.2 Semantic neologisms .....	64
6.3 Borrowing .....	65
6.4 Lexicalization of phrases .....	67
6.5 Other characteristics.....	68
7. Conclusions.....	69
7.1 Observations about the word formation processes in these two languages projects.....	69
7.2 Reflections about the difficulties during the process of our study.....	71
References.....	75
Appendix.....	78

## 1. Introduction

The emergence of neologisms, i.e. new words or new meanings for existing words, is intricately linked to the culture of a society. Neologisms reflect how various events and situations in domains such as politics, religion, art, sports, science, and others shape language (Alves, 2007). Scholars have shown increasing interest in studying neologisms, analyzing them from different theoretical perspectives, including morphology (Akut, 2020; Hormigo, 2022) and semantics (Onyedum, 2012; Adelstein, 2019), as well as in specific domains such as the Internet (Pratama et al., 2021), medicine (Porrás Garzón, 2016; Asif et al., 2021), and entertainment (Alvarez-Bolado Sanchez et al., 2013), among others. However, there are varying opinions among scholars regarding the exact definition of neology.

English and Spanish, as two of the most widely spoken languages in the world, have been the focus of recent neologism research. However, despite the growing body of literature on this topic, comparative research between English and Spanish neologisms remains limited. Therefore, this study aims to address this research gap by conducting a comparative analysis of neologisms in these two languages. By examining both lexicological characteristics and lexicographic representation, our analysis mainly aims to shed light on the similarities and differences in the word formation processes of neologisms within these two linguistic communities.

For this study, we explored the creation of new words in English and Spanish, utilizing the regular contributions to the journal *American Speech* entitled *Among The New Words* as our source for English data, and *El banco de neologismo - Centro Virtual Cervantes*<sup>1</sup> for collecting Spanish data. The investigation covers the period from 2005 to 2015, providing a

---

<sup>1</sup> We will refer to it as "Banco de Cervantes" in the subsequent references.



preliminary overview of the types of neologisms and the word formation process in both languages within these two projects.

Through these resources, our primary objective is to analyze the morphological processes involved in creating new words in English and Spanish, as well as examine the use of borrowed words and semantic neologisms. Furthermore, we seek to uncover innovative observations regarding the characteristics of neologisms in these two languages. By focusing on the selected timeframe, we aim to investigate recent developments in neology and gain insights into the linguistic trends and preferences that shape word formation.

The research questions guiding this study are as follows:

1. What are the similarities and differences in the general word formation processes of English and Spanish neologisms?
2. What factors contribute to the observed similarities or differences in the productivity of word formation processes in these two languages' neologisms?
3. How do English and Spanish neologism projects differ in their conceptualization of neology?

In order to address these questions, we must identify the most representative word formation processes observed in English and Spanish neologisms. In addition, it will be useful to point up any additional characteristics in the data collected that are particularly noteworthy.

Our preliminary hypotheses suggest that, in general, English and Spanish new words may share similar word formation processes such as prefixation, suffixation, compound formation, neoclassical compound and so on. Conversely, differences in word formation processes between the two languages' neologisms may lie in the distinct productivity of each process. For example, due to the fact that English is a Germanic language, compounds may be more prevalent, while Spanish neologisms may feature a higher proportion of neoclassical

compounds as a result of Spanish being derived from Latin. Moreover, we anticipate that borrowing in Spanish neologisms may exhibit higher productivity compared to English.

To address the research questions in more detail, a literature review and analysis are presented, along with a historical review of the two language-specific neologism projects in Chapter 2. Chapter 3 presents the methodology adopted in this study. Chapter 4 presents the in-depth analysis of English neologisms from *Among The New Words*, while Chapter 5 focuses on the analysis of Spanish neologisms from *Banco de Cervantes*, drawing upon the applied theoretical framework discussed throughout the review. The comparative analysis of English and Spanish neologisms within these two projects is presented in Chapter 6, and Chapter 7 provides conclusions, reflections, and suggestions for future research based on the findings.

## **2. Historical Review of the Two Neologisms Projects**

### **2.1 *Among the New Words - American Speech***

*Among the New Words* is the semiregular column of *American Speech*, which holds the distinction of being the longest-running record of newly coined English words. The mission of *Among the New Words* is defined by its unique approach, which prioritizes thorough documentation of its entries. To be considered a new word by *Among the New Words*, a term or phrase must not be included in general dictionaries. However, they can be interesting forms of making, or even a witness of society transformation and so on (Algeo & Algeo, 1991).

Originated by Dwight Bolinger in 1941, the column *Among the New Words* has undergone changes in leadership throughout its history. Bolinger conceived the idea and introduced it to *American Speech* after its adaptation from another publication. Due to his primary academic focus lying elsewhere, Bolinger relinquished control of the article in 1944, having contributed nine installments encompassing over 250 lexemes (Algeo & Algeo, 1997).

However, it is highly unlikely that Bolinger could have anticipated the enduring longevity of the feature, as it continues to thrive nearly 80 years later.

*Among the New Words* has seen a succession of editors across three generations. To begin with, I. Willis Russell assumed the role in 1944 and served as the chief editor for an impressive 41 years until 1983. He continued as a supporting editor until the final issue in 1985. During his tenure, Russell produced a remarkable 84 installments, featuring nearly 2,300 lexemes. In his last years as chief editor, Mary Gray Porter provided assistance and took over as chief editor in 1984, holding the position for two years and publishing six installments containing approximately 160 lexemes (Algeo & Algeo, 1997).

In 1987, as the second generation editor, John Algeo became the chief editor, with his wife, Adele, initially assisting and later being recognized as a co-editor due to her substantial contributions. Over a decade, they produced 40 installments of *Among the New Words*, publishing one in every issue of *American Speech* during their tenure. These installments comprised approximately 3,200 lexemes. The regularity and extensive content of their contributions were made possible by their joint efforts and the invaluable support received from contributors (Algeo & Algeo, 1997).

Wayne Glowka took over in 1997, transitioning from traditional citation methods to electronic databases. The current editors, Charles Carson, Benjamin Zimmer, and Jane Solomon, have expanded the sources to include print, digital, and spoken media, such as social media, audiovisual content, and song lyrics (Zimmer, Carson & Soloman, 2016).

Algeo & Algeo (1991) indicate that *every aspect of the life of a people is reflected in the words they use to talk about themselves and the world around them. As their word changes - through invention, discovery, revolution, evolution, or personal transformation- so does the language. Like the growth rings of a tree, our vocabulary bears witness to our past.* Furthermore, throughout the changes in methodology and sources to continue to report on

English neologisms, the mission of *Among the New Words* remains constant, with a commitment to adapt to future linguistic transformations.

## **2.2 *El banco de neologismo - Centro Virtual Cervantes***

*El banco de Neologismos* developed by the *Centro Virtual Cervantes* is a collection of lexical neologisms extracted from written and spoken media in Spanish and Catalan dating back to 1988. This database is derived from the databases of neologisms of the *Observatori de Neologia* (OBNEO) project at the Institut de Lingüística Aplicada (IULA) (<https://www.upf.edu/es/web/obneo>). Unlike an evaluative resource, *Banco de Cervantes* serves as an inventory to aid researchers in diagnosing and analyzing the use and implementation of neologisms in both languages. The collaboration between Pompeu Fabra University and the Cervantes Institute has made it possible to compile a list of neologisms identified in a large media corpus during the period of 2004-2018.

Therefore, as *Banco de Cervantes* can be seen as a product of the OBNEO neologism project, we now present a historical review of the OBNEO project.

According to Cabré (2006), the OBNEO project underwent four distinct stages, each with its own focus and objective. The initial stage, starting in 1989, aimed to establish the project's methodology and organizational structure. This involved defining objectives, determining the research scope, and developing a methodological framework. Data detection and collection were key activities during this stage, which continued until around 1993 (Cabré, 2006). From the beginning, providing dictionary publishers with data on neologisms has been an important goal of the OBNEO Project, both for Spanish and for Catalan.

In the second stage, the project shifted its focus to data analysis and interpretation, particularly in the study of neologisms (Cabré, 2006). Meanwhile, a classification system was established, providing a theoretical framework for the *Banco de Cervantes* project.

The third stage involved expanding the project's research scope based on a suggestion made by Dr. Antoni M. Badia Margarit. It included lexical neologisms from written press, informal written press, and audiovisual media, enriching the database and enabling comparisons with previous findings (Cabr , 2006).

Simultaneously, the fourth stage aimed to maximize automation in project workflows. The SEXTAN system, currently being replaced by SEXTAN II, facilitated the semi-automatic extraction of formal neologisms. Additionally, the LEGIMUS system, currently undergoing replacement, supported the project's objectives (Cabr , 2006).

In summary, these stages represent the historical progression of the OBNEO project and its significant contributions to the *Banco de Cervantes*. The evolution of OBNEO, characterized by methodological refinement, expanded research horizons, and the integration of automation, greatly enhanced the efficiency and productivity of the *Banco de Cervantes*, leading to a more comprehensive neologism project for dictionary supplementation.

Additionally, based on our historical review, we predict that English data will contain a larger percentage of phrases compared to Spanish due to the differing nature of the two neologism projects. Consequently, we further speculate that these distinct language projects hold different conceptualizations of neology for their respective project purposes.

### **3. Methodology**

We collected English data from *Among the New Words* and Spanish data from *Banco de Cervantes* during the 11-year period encompassing 2005-2015.

#### **3.1 Data collection: English**

We were able to access a total of 35 articles published under *Among the New Words* in *American Speech* between 2005-2015. Firstly, we collected all the new words and expressions

from the articles in chronological order and assigned numbers 1-50+ to the items for each year.

The total number of all entries collected is 689.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Total	72	88	65	72	59	51	87	58	32	78	27 <sup>2</sup>

Table 1: Total number of words collected for each journal before filtration.

To ensure work with a manageable number of entries given the practical limitations of this study, our goal was to collect random 30 expressions for every year. To ensure the random selection of the data, we followed these steps for data filtration. In odd-numbered years (e.g., 2005, 2007, 2009, etc.), we selected the odd-numbered words up to thirty first, such as 1, 3, 5, 7, and so on. If there were not enough words to reach thirty, we continued adding words from the even numbers in order. Similarly, in even-numbered years (e.g., 2006, 2008, 2010, etc.), we selected the even-numbered words first, such as 2, 4, 6, 8, and so on. If there were not enough words, we added words from the odd numbers in order. In total, we gathered 327 new words from *Among the New Words* after the filtration process.

Year	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Total	30	30	30	30	30	30	30	30	30	30	27

Table 2: The total number of words for each journal we collected after filtration.

---

<sup>2</sup> Normally, *Among the New Words* appears in three to four issues of *American Speech* each year, but in 2005, only two were published. As a result, the total number of words collected before filtration for that year was less than 30. Therefore, we did not filter any data for that year.

As a result of this process, we generated two distinct Excel sheets for each year, namely "Year" and "Year (filtered)". Within each Excel table, we designated the following columns as table headers: *Number, Entry, Year, Vol Issue Am. Speech, Morpheme classification, Semantic neologism, Borrowing, Definitions and examples*. It is important to note that the "Year" Excel table does not include information related to morpheme classification, semantic neology, borrowing, definitions, and examples. Instead, it serves solely to present the total number of words collected for each year prior to the filtration process. The "Year (filtered)" Excel sheet, on the other hand, includes comprehensive data, providing detailed insights into the morphological classification, semantic neologism, borrowing, definitions, and examples associated with each word. The Excel sheets are available for consultation: <https://kdocs.cn/l/coIZcHtfTKBb>.

Our attention at this stage was primarily directed towards completing the required fields of *Number, Entry, Year, Vol Issue Am. Speech, Definitions and examples* for each word that had undergone the filtration process.

In order to conduct a comparative analysis between the English neologisms project and the Spanish neologisms project, it was necessary to classify the morphological process to create the new form of each new word (or expression) in more detail. This information is provided by *Banco de Cervantes*, but is not included in *Among the New Words*.

It is important to recognize the inherent challenge associated with this endeavor, given the potential divergence among authors in their categorization of word formation in English. For example, the criteria used for differentiating compounds from phrases is not clearcut (Plag 2003). Consequently, the classification of word formation processes may exhibit a subjective nature in certain instances. It is evident that each classification approach carries its own strengths and limitations. After conducting a thorough examination of various scholarly sources, we have decided to adopt Laurie Bauer's definition of word formation as outlined in

his influential work, *English Word-formation* (1983), as well as that in his chapter “*Lexical Word-formation*” (co-authored with Rodney Huddleston) to the *Cambridge Grammar of the English Language* (2002). This selection provides a relatively standardized framework for our classification. The detailed classification criteria and the comprehensive results are thoroughly discussed in Section 4.1 Formal Neologisms.

Since neologisms may be formal or semantic in nature, and *Banco de Cervantes* identifies this parameter for Spanish, it was necessary to classify the English data accordingly.

During this step, we thoroughly examined the definitions and sentence examples associated with each word intended for inclusion in the *Among the New Words* database. To determine their status as semantic neologisms, we referred to reputable dictionaries such as the *Oxford English Dictionary (OED)*, *Cambridge Online Dictionary*, and *American Heritage Online Dictionary*. Our objective was to ascertain whether these words already existed in the lexicon. If they were confirmed as pre-existing words and *Among the New Words* provided a new definition, we categorized them as having a new meaning, thereby classifying them as semantic neologisms. Furthermore, we would also check if the new meaning of the existing word was included in the dictionaries. These sources were also consulted to determine if the English expression could be considered a borrowing.

However, it is important to acknowledge that certain older words may not be listed in dictionaries for various reasons. Furthermore, *Among the New Words* presents different definitions and examples for a given word across different years. In such cases, we also considered them as having a new meaning.



All remaining words<sup>3</sup> that did not fit into the aforementioned categories were classified as new words, indicating that they were entirely novel creations.

Figure 1: screenshot of the "2005 (filtered)" Excel sheet of English data.

Figure 2: screenshot of the "Final list" Excel sheet of English data (1).

<sup>3</sup> In a subsequent phase, we observed that *Among the New Words* incorporated affixes and combining forms within their journal entries. To indicate this inclusion, we employed a "/" symbol in the semantic neologism column of our records.



## 559 arcilla

---

Contexto: [...] nosotros no estamos tan acostumbrados por el hecho de jugar más en \*arcilla\* que en indoor, comentó uno de los técnicos del equipo, Sergio Elías.

Categoría gramatical: Nombre femenino (f)

Tipo de neologismo: Formado por abreviación (FTABR)

Aspecto tipográfico: Sin marca tipográfica (smt)

Fuente: El Mercurio (Chile) (X1)

Corpus de vaciado: Corpus escrito de prensa (pr)

Figure 4: An example entry of the newly detected word *arcilla* (item 1) in the *Banco de Cervantes*' system.

We collected information on the types of neologisms from *Banco de Cervantes* and entered the data into our each "Year" Excel sheet. The spreadsheet includes details<sup>5</sup> such as morphological classification, indication of semantic neologism (marked as yes or no), indication of borrowing (marked as yes or no), the language of origin for borrowed words, and the contextual information provided for semantic neologisms.

---

<sup>5</sup> In the post-processing stage, we included a column for Syntactic neologism in the Spanish Excel sheet and marked 'yes' or 'no' accordingly.

Entry	Year	Morpheme classification	Semantic n	Borrowing	Syntactic neologism	Language origin for the borrowing words	Context for the semantic neologisms
arcilla	2005	abbreviation	no	no	no		
bolón de pelo	2005	phrase	no	no	no		
congratular	2005	compound	no	no	no		
dipstick	2005	acronym	no	no	no		
estatuyente	2005	suffixation	no	no	no		
fotoperiódico	2005	neoclassical compound	no	no	no		
gramsciano-n	2005	suffixation	no	no	no		
histórico	2005	conversion	no	no	no		
interoperativid	2005	prefixation	no	no	no		
judeidad	2005	suffixation	no	no	no		
knock down	2005	/	no	yes	no	préstamo del inglés	
lirico	2005	/	yes	no	no		Hace tiempo que las autoridades de este teatro estaban buscando un tenor con cor
minipaga	2005	neoclassical compound	no	no	no		
nivel de Sama	2005	phrase	no	no	no		
organoclorar	2005	neoclassical compound	no	no	no		
prealerta	2005	prefixation	no	no	no		
quimioterapéu	2005	neoclassical compound	no	no	no		
resocializar	2005	prefixation	no	no	no		
spaghetti-west	2005	/	no	yes	no	préstamo del inglés	
toy	2005	/	no	yes	no	préstamo del inglés	
unilateralismo	2005	suffixation	no	no	no		
vintage	2005	/	no	yes	no	préstamo del inglés	
winner	2005	/	no	yes	no	préstamo del inglés	
xilocollage	2005	neoclassical compound	no	no	no		
yihadista	2005	suffixation	no	no	no		
zona cero	2005	phrase	no	no	no		

Figure 5: screenshot of the "2005" Excel sheet of Spanish data.

Notably, *Banco de Cervantes* employs a systematic classification approach for neologisms, which primarily encompasses five distinct types: formal neologisms, syntactic neologisms, semantic neologisms, and a residual category with cases that present difficulties in classification. It is important to highlight that certain categories, such as syntactic neologisms, borrowed words, semantic neologisms, and cases posing challenges in classification are not further tagged under their respective morpheme classifications.<sup>6</sup> Consequently, a literal interpretation of this classification framework would entail the exclusivity of the established types, thus precluding a neologism classified as a formal neologism from also being classified as a semantic neologism (Cabré, 2006). Therefore, during our data collection and analysis, we undertook the task of identifying and labeling all neologisms lacking a specific morpheme classification, as well as those categorized as *others* in *Banco de Cervantes* utilizing the "/" symbol to differentiate them from the systematically classified counterparts.

<sup>6</sup> It is important to note that during our English data processing, we classified the new words based on their morphological word formation process before labeling the semantic neologisms and borrowing columns. As a result, there may be instances in English where the types of neologisms can overlap. We used the symbol "/" to mark those English new words that could not be further classified in terms of their morpheme classification.

Furthermore, we conducted a thorough review of all the morphological process information provided by *Banco de Cervantes*, utilizing the theoretical framework established by Cabré (2006) for the *Observatori de Neologia*. While we generally accepted all the word formation process labels assigned by *Banco de Cervantes*, we identified a few instances that we deemed potentially incorrect.

The absence of word definitions in *Banco de Cervantes'* data posed challenges for us in comprehending the true semantic shift of the new words classified as semantic neologisms. As a result, we relied on the provided context to gain insight into the meaning of these semantic neologisms. Additionally, we consulted authoritative dictionaries such as the *Diccionario de la lengua española* to validate whether the interpreted meanings of the words were registered in the dictionary.

While the majority of borrowing words were accompanied by information about their language origins, there were instances where the origin was labeled as "otros préstamos" (other loanwords) or "Otros préstamos adaptados" (other adapted loanwords). In order to conduct a comprehensive analysis of borrowing words in Spanish, we consulted additional academic resources such as *Merriam-Webster Online Dictionary* to determine the accurate origins of these uncertain origin types of borrowing words in *Banco de Cervantes'* data.

All the existing data from 2005 to 2015 were transferred to the Spanish Excel sheet, according to chronological order. Each item of the final list was assigned a number from 1 to 284. The Excel sheets are available for consultation: <https://kdocs.cn/l/cj5yvfqjFpj8>.

Number	Entry	Morpheme classifi	Semantic need	Borrow	Syntactic neologism	Language origin for the borrowing words	Context for the semantic neologisms
1	arcilla	2005 abbreviation	no	no	no		
2	bolón de pobreza	2005 phrase	no	no	no		
3	comprañolos	2005 compound	no	no	no		
4	digipack	2005 acronym	no	no	no		
5	estatuyente	2005 suffixation	no	no	no		
6	fotoperiódico -ca	2005 neoclassical compound	no	no	no		
7	gramsciano -na	2005 suffixation	no	no	no		
8	histórico	2005 conversion	no	no	no		
9	interoperatividad	2005 prefixation	no	no	no		
10	judeidad	2005 suffixation	no	no	no		
11	knock down	2005 /	no	yes	no	préstamo del inglés	
12	lirico	2005 /	yes	no	no		Hace tiempo que las autoridades de este teatro estaban buscando un te
13	minipaga	2005 neoclassical compound	no	no	no		
14	nível de Samadhi	2005 phrase	no	no	no		
15	organoclorar	2005 neoclassical compound	no	no	no		
16	presletra	2005 prefixation	no	no	no		
17	quimioterápico	2005 neoclassical compound	no	no	no		
18	resociatizar	2005 /	no	no	no		
19	spaghetti-western	2005 /	no	yes	no	préstamo del inglés	
20	toy	2005 /	no	yes	no	préstamo del inglés	
21	unilateralismo	2005 suffixation	no	no	no		
22	viataje	2005 /	no	yes	no	préstamo del inglés	
23	wimer	2005 /	no	yes	no	préstamo del inglés	
24	silocollage	2005 neoclassical compound	no	no	no		
25	yhadista	2005 suffixation	no	no	no		
26	zona cero	2005 phrase	no	no	no		
27	arquitectura blanca	2006 phrase	no	no	no		
28	boxeos nois	2006 /	no	yes	no	préstamo del catalán	
29	concertista	2006 conversion	no	no	no		
30	dictocrata	2006 neoclassical compound	no	no	no		
31	estado fallido	2006 phrase	no	no	no		
32	fotofinish	2006 /	no	yes	no	préstamo del inglés	
33	goya	2006 /	yes	no	no		El festival rendirá homenaje al guionista Rafael Azcona, que tras obten

Figure 6: screenshot of the "Final list" Excel sheet of Spanish data.

### 3.2 Classification of data

For the analysis of English neologisms, we relied on the framework proposed by Laurie Bauer (1983) in his influential work *English Word-formation*, as well as on Bauer and Huddleston (2002). These works provided us with a solid foundation and framework to examine word formation processes in English.

For the analysis of Spanish neologisms, we drew upon the *Nueva gramática de la lengua española* published by *Real Academia Española* (2009), which served as a comprehensive reference for the Spanish language. Additionally, we incorporated Teresa Cabré's (2006) theory on *La clasificación de neologismos: una tarea compleja* as a guiding framework for the *Observatori de Neologia* (OBNEO) project.

By integrating these established frameworks and scholarly works, we ensured a rigorous and informed approach to analyzing neologisms in both English and Spanish languages.

We utilized the filtering function in the Excel sheet to facilitate the calculation of percentages related to different word formation processes and types of neologisms. This approach allowed us to generate visually informative charts that accurately represent the

distribution and composition of these linguistic phenomena. We then compare the English data with the Spanish data in terms of five key aspects: formal neologisms, semantic neologisms, borrowing, lexicalization of phrases, and other linguistic characteristics.

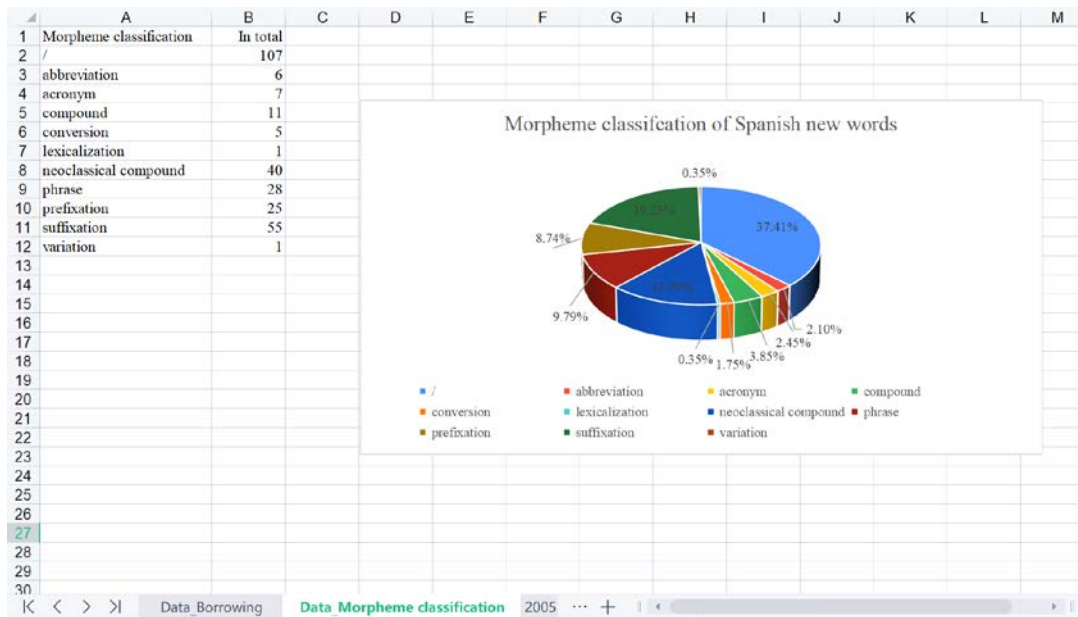


Figure 7: example of the screenshot about the data processing for Spanish data (morpheme classification).

Number	suffix	Entry	Year	Morphem	Semantic	Borrowin	syntactic	neologism
5	ente	estatuyente	2005	suffixati	no	no	no	no
7	ano, ana	gramsciano -na	2005	suffixati	no	no	no	no
10	dad	judeidad	2005	suffixati	no	no	no	no
21	ismo	unilateralismo	2005	suffixati	no	no	no	no
25	ista	yihadista	2005	suffixati	no	no	no	no
37	ista	kirchenista	2006	suffixati	no	no	no	no
50	ico, ica	xenofóbico -ca	2006	suffixati	no	no	no	no
55	al	cnkursal	2007	suffixati	no	no	no	no
59	ero	grafitero	2007	suffixati	no	no	no	no
61	ero	interrailero	2007	suffixati	no	no	no	no
62	ón, ona	jugón -ona	2007	suffixati	no	no	no	no
63	ismo	kirchnerismo	2007	suffixati	no	no	no	no
64	ado	lixiviado	2007	suffixati	no	no	no	no
69	ato	quinquenato	2007	suffixati	no	no	no	no
72	able	toreable	2007	suffixati	no	no	no	no
73	ista	uniformista	2007	suffixati	no	no	no	no
77	ista	yihadista	2007	suffixati	no	no	no	no
81	miento	completamiento	2008	suffixati	no	no	no	no
82	ción	dictaminación	2008	suffixati	no	no	no	no
85	esco	gorileSCO -ca	2008	suffixati	no	no	no	no
88	orio, ria	jubilatorio -ria	2008	suffixati	no	no	no	no
89	ismo	kirchnerismo	2008	suffixati	no	no	no	no
93	dad	ordinalidad	2008	suffixati	no	no	no	no
96	ante	resaltante	2008	suffixati	no	no	no	no
98	able	transable	2008	suffixati	no	no	no	no
99	ista	unilateralista	2008	suffixati	no	no	no	no
100	ar	villanizar	2008	suffixati	no	no	no	no
104	ista	zenista	2008	suffixati	no	no	no	no
110	ón	frailón	2009	suffixati	no	no	no	no
111	ano, ana	gramsciano -na	2009	suffixati	no	no	no	no
125	ción	uniformización	2009	suffixati	no	no	no	no

Figure 8: example of the screenshot about the data processing for Spanish data (suffixation analysis).

## 4 Analysis of English Neologisms in *Among the New Words*

### 4.1 Formal neologisms

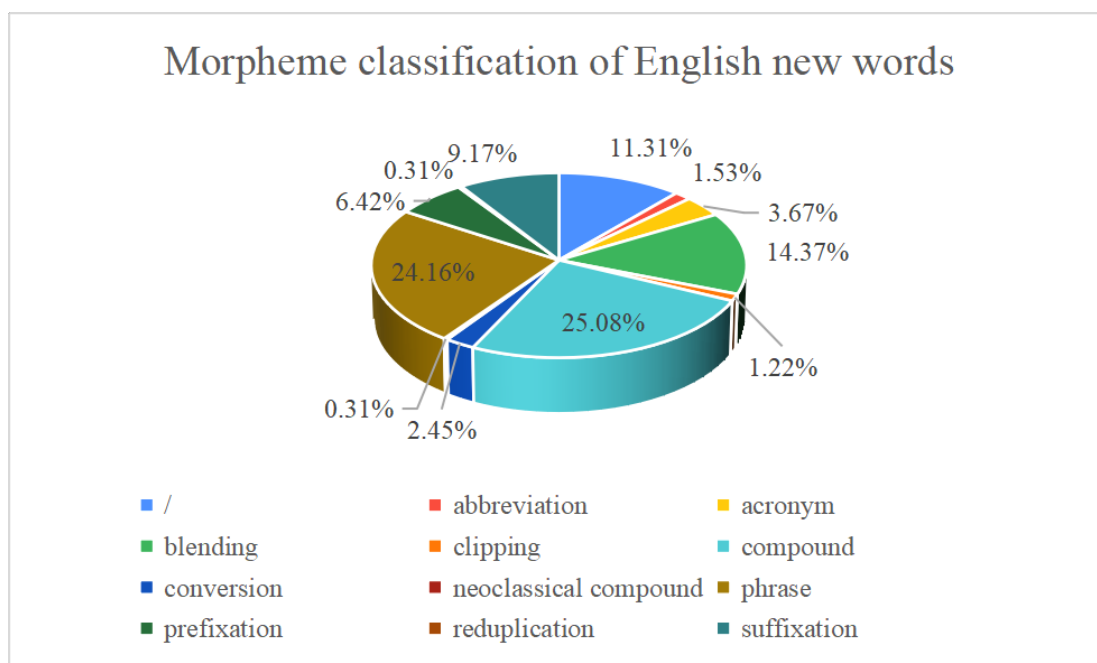


Figure 9: Distribution of different word formation processes' productivity for English data.

In this section, we present the distribution of word formation processes in newly coined English words in our dataset.

In general terms, the predominant word formation processes observed are compound formation and phrases, accounting for 25.08% (82 items) and 24.16% (79 items) respectively. Notably, blends constitute 14.37% (47 items) of the total (327 items) and represent a relatively significant proportion of the newly created English words. This finding is intriguing, as blends have received relatively limited attention in linguistic circles. However, our data demonstrate a remarkable diversity and complexity within this category, which is discussed in greater detail in Section 4.1.5.



Additionally, our investigation indicates that 9.17% (30 items) of the new words undergo the process of suffixation, while 6.42% (21 items) exhibit prefixation. Within these two categories, certain affixes emerge as more productive within the dataset, such as the suffix *-er* and the prefix *un-*. These specific affixes are further analyzed in Section 4.1.2 and 4.1.3 respectively.

Approximately 11.31% (37 items) of the data is marked with the symbol "/", indicating that these words do not conform to standard vocabulary formation methods and cannot be classified accordingly. This category encompasses various combining forms such as *-coin* and *-sauce*, as well as the productive utilization of proper nouns in new words, for instance, *Mel Gibson*<sup>7</sup> (item 69) and *Mellencamp* (item 216). These terms defy classification into standard word formation types. Moreover, the dataset comprises borrowed words from other languages, such as *luanqibaozhao* (item 21) borrowed from Chinese, meaning *complicated mess* and *sudoku* (item 40) borrowed from Japanese, denoting *a puzzle involving filling slots or boxes*. Due to the inherent disparities between foreign languages and English, determining the precise form of these borrowed words remains challenging, even when adapted to the English language.

Furthermore, certain new words exhibit multiple forms due to their lack of standardization. For instance, *whale tail* (item 53) and *whale's tail* (item 53) represent different variations of the same item. While *whale tail* can be categorized as a compound, *whale's tail* falls under the phrase category due to the presence of genitive 's. Since it is not feasible to classify multiple forms simultaneously, such words are also placed in the "/" category.

---

<sup>7</sup> Since *Among the New Words* is not a systematic dictionary supplement oriented neologism project, and multiple forms of a single neologism may occur, we cite only the first form for enumeration among those that occur in multiple forms.

It is worth noting that we even encountered new words that defy further classification as prefixation, suffixation, or compounding, such as the word *quadraboobage* (item 103). This term refers to bulging flesh resembling four breasts, resulting from wearing an ill-fitting brassiere. There are multiple possible analyses of its internal structure: [[quadra + boob] age], [quadra [boob + age]], or [quadra] + [boob] + [age]. Although the third combination is highly unlikely, and we can reasonably dismiss it, distinguishing between the first and second possibilities remains inconclusive. Thus, in such cases, these words are also included in the "/" category. Besides, this category also encompasses certain old words with new meanings that defy further classification based on morphemes, such as *angel* (item 13) and *even* (item 309), among others.

Conversion is also considered a lexical word-formation processes according to Bauer and Huddleston (2002), as it involves the creation of new words. In our English dataset, we identified a conversion rate of 2.45% (8 items) for newly formed words. Additionally, we observed instances of minor word-formation processes, such as initialisms. Acronyms and abbreviations are two prominent types of this word-formation process, both of which are present in our dataset, accounting for 3.67% (12 items) and 1.53% (5 items) respectively. Another noteworthy word formation process observed in our English data corpus is clipping. Despite the historical significance of a base being formed through the process of clipping not being grammatically relevant in present-day language (Bauer and Huddleston, 2002), it remains a productive process within the minor categories of word formation. In our dataset, clipping accounts for 1.22% (4 items) of the overall data.

Neoclassical compounds constitute a relatively small portion, accounting for only 0.31% (1 item) of the total. The limited prevalence of neoclassical compounds in our data can be attributed to the challenge of discerning between regular compounds and neoclassical compounds. However, to ensure consistency in classification and facilitate a meaningful

comparison with the Spanish dataset, we established a criterion that necessitates both forms of a combination to exhibit neoclassical characteristics in order to be categorized as a neoclassical compound.

We also found an occurrence of the reduplication process, a category that is not explicitly addressed in Bauer and Huddleston's classification (2002). This category accounts for a mere 0.31% (1 item) of the overall total.

Overall, our English neologism data demonstrate the richness of word formation processes, the most representative feature of which is the demonstration of the strong vitality of compound word formation. Lexicalized phrases are also salient in the data. As we mentioned earlier, *Among the New Words* is not a dictionary supplement-oriented neologism project, we could find many words that are on the informal side and feature a predominance of phrases. In fact, lexicographers do not generally include phrases in dictionaries, so the relatively large number of phrases found in a collection of new words that is not dictionary supplement-oriented is consistent with our expectations. In addition, we have many interesting findings in our data, but since the main nature of our study is to compare it with the Spanish neologism project, we only discuss in detail some of the points we consider most worthy of discussion in this section, while the rest is summarized and analyzed in a more concise and general way.

#### **4.1.1 Difficulties of the classification**

In this section, we would like to elaborate on the difficulties we encounter when classifying word formation processes. As we mentioned earlier, different authors use different groupings for classification, and we have chosen what we consider to be the relatively most authoritative criteria as our framework. However, we still encountered significant difficulties in carrying out the classification process.

Phonological characteristics play a crucial role in distinguishing compounds from phrases. For example, in English, compound nouns typically exhibit primary stress on the first

element, while adjective + noun combinations typically feature primary stress on the second element. The issue of stress placement in compound words and phrases has been explored by researchers such as Quirk et al. (1985).

However, it is worth noting that there is no consensus regarding the consistent treatment of stress patterns in compounds and phrases in English. Yurtbaşı (2017) highlights the lack of a standardized approach to the stressing of compounds and phrases, both in terms of writing and pronunciation. Bauer and Huddleston (2002) also discuss the distinction between morphological compounds and syntactic constructions, particularly when the final component is a noun rather than an adjective or verb. They provide examples to illustrate this contrast.

1.a. *greenhouse sweetheart cotton-plant newspaper* [morphological compound]

b. *green house sweet taste cotton shirt quality paper* [syntactic construction]

(Bauer and Huddleston, 2002: 1644)

In example [1.a], the criteria align: *greenhouse* and *sweetheart* are written as single words and have primary stress on the first component, among other characteristics. In contrast, examples in [ii], like *green house* and *sweet taste*, are written as word sequences and have primary stress on the second component. They also allow a broader range of modification, similar to phrases headed by adjectives or nouns. However, it is worth noting that the boundary between morphological compounds and syntactic constructions can be uncertain in certain cases, indicating the complexities involved in classification.

The reliance on spelling alone for distinguishing compounds from phrases can be problematic due to the variable ways in which compounds may be written. They can appear as one word (e.g., item 20: *Jesusland*), two separate words (e.g., item 58: *holiday tree*), or hyphenated words (e.g., item 26: *bikini-ready*). In contrast, phrases are typically written as separate words (e.g., item 36: *pig and the python*). However, there are instances where spelling alone is insufficient to differentiate between compounds and phrases.

Moreover, since the neologism project *Among the New Words* does not provide phonological information for each new word, we are unable to apply phonological features to differentiate compounds from phrases. Consequently, our classification efforts rely primarily on spelling, which presents significant challenges. For example, in our dataset, item 119 includes the forms *big foot* and *bigfoot*. As previously mentioned, such variations in form can complicate the classification process. While *bigfoot* can be confidently classified as a compound, it is unclear whether *big foot* places primary stress on the first element *big* or the second element *foot*. Therefore, without phonological information, our ability to discern between compounds and phrases solely based on spelling is limited. Nonetheless, we have made diligent efforts to classify these instances to the best of our abilities.

Another challenge we have encountered is the distinction between blends and compounds containing combining forms, as the boundaries between these two types of word formation processes are not always clear. A blend refers to a word formed by combining parts of two or more words to create a new word with a new meaning. However, the difficulty arises when trying to determine whether a word is a blend or a compound containing a combining form solely based on its form.

For instance, consider the example *Obamakinz* (item 128). It is uncertain whether *-kinz* is a potential new combining form combined with the proper noun *Obama*, or if it is a blending of *Obama* with another word containing the *-kinz* element. Similarly, for the word *pairage* (item 319), which is defined as [*pair + marriage*] denoting a legally recognized same-sex union equivalent to marriage in *Among the New Words*, there is ambiguity in its classification.

According to Lehrer (2003: 371), a full word followed by a splinter (word part), e.g. *oldraulic* < *oil + hydraulic*, is one type of blending. Applying this definition, *pairage* appears to be a blend. Bauer and Huddleston (2002: 1700), however, describe *-age* as a suffix that occurs in various nouns, although it is no longer productive. It is predominantly attached to

nouns and verbs, with occasional use in adjectives. Their examples demonstrate different meanings associated with *-age* usage, such as collectivity, state or condition, result, place, amount or rate, and charge.

2. a. *baggage, coinage, fruitage, leafage, wordage* [collectivity]
- b. *bondage, parentage, peerage, pupilage, shortage* [state, condition, rank]
- c. *breakage, marriage, stoppage, wastage, wreckage* [result]
- d. *anchorage, hermitage, orphanage, parsonage, vicarage* [place]
- e. *acreage, dosage, mileage, tonnage, voltage* [amount or rate]
- f. *anchorage, cartage, corkage, haulage, postage* [charge]

(Bauer and Huddleston, 2002: 1700)

We note that *-age* functions as a suffix in the word *marriage*. This raises the question of whether the element *-age* in *pairage* can be considered a suffix as well. Complicating matters further, both *age* and *pair* are existing words in the dictionary, which might suggest that *pairage* could also be classified as a compound word. We suggest that *pairage* is similar in structure to those words in [2b], which implies that *-age*<sub>[affix]</sub> still has limited productivity in English.

Another example illustrating the challenges in distinguishing between affixes and combining forms is the word *podcast* (item 48), which combines *iPod* and *broadcast*. Coined in 2004, *podcast* describes a digital media format allowing users to download audio files for portable devices like iPods. The term merges *pod* from *iPod* with *cast* from *broadcast*, representing the distribution of audio or video content to a dispersed audience.

Lehrer (2003: 371) proposes a type of blend combining *two splinters* (e.g., *Spanglish* < *Spanish* + *English*); *sitcom* < *situation* + *comedy*). This implies that *podcast* can be classified as a blend. However, the challenge lies in the interpretation of *-cast*. It can be viewed as a combining form, as seen in *broadcast*, but it can also function as a suffix when added to words

to create new ones with related meanings, such as *forecast*. Consequently, another issue arises in distinguishing between affixes and combining forms.

The diachronic development of word formation is a complex process that continues to be investigated by linguists. As language evolves, new words emerge, while old words may decline in usage or undergo shifts in meaning. This dynamic nature contributes to the creation of new affixes and combining forms, as language users manipulate existing words to express novel concepts. However, the underlying mechanisms driving these changes remain incompletely understood and remain the focus of ongoing research.

Additionally, there is an ongoing debate concerning the relationship between derivation and compounding, for example, when the suffix *-er* is attached to a compound. This situation prompts the question of whether the resulting word should be categorized as a compound or as a product of suffixation. This issue has been analyzed in a variety of ways in the literature (e.g., Anderson, 1992; Lieher, 1992; Booij, 2010) and discussion of the differing standpoints goes beyond the scope of this paper.

In general, we believe that the difficulties we have encountered are related to the definition of compound, whether it is the delineation between compound and phrase, or between compounding and blending, or between compound and derivation, or even the distinction between affix and combining form, which is to some extent unclear due to the role of diachronic development of word formation. Although the definition of compound words has been a topic of debate among scholars, with various definitions proposed and revised over time in response to new challenges (e.g., Bauer, 1983; Plag, 2003), it may need to be further refined in the future since the data exhibit challenges in terms of classification.

### 4.1.2 Suffixation

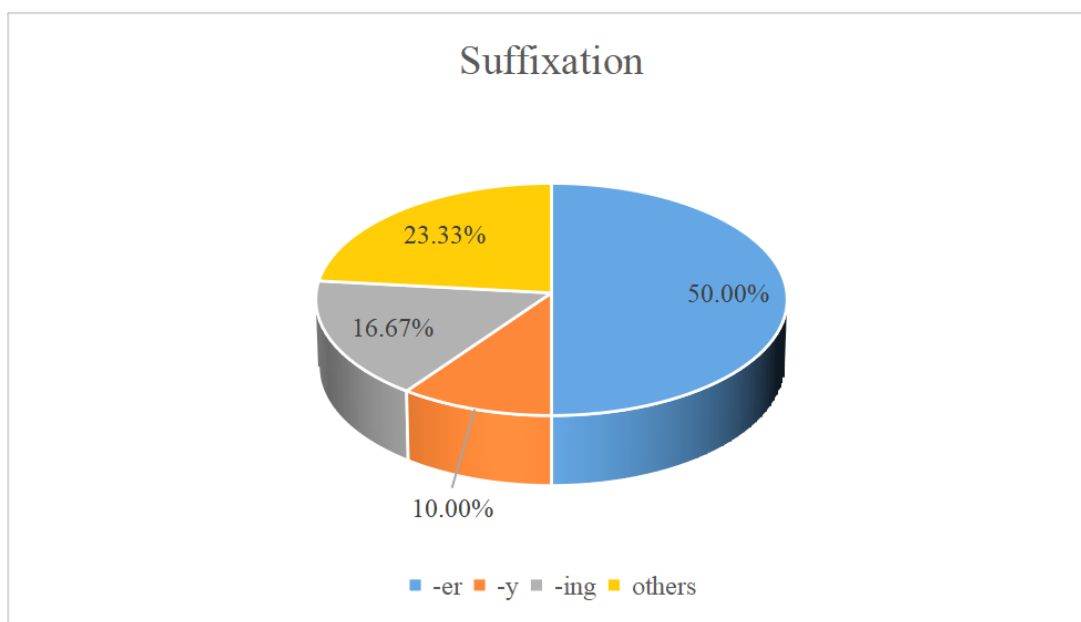


Figure 10: Distribution of different suffixes' productivity in the English suffixation group.

Within the domain of suffixation, which constitutes 9.17% (30 items) of the dataset, a salient pattern emerges, particularly concerning words adorned with the suffix *-er*. This specific suffix accounts for 50% (15 items) of all words formed through suffixation, while an additional 16.67% (5 items) of words are formed by attaching the suffix *-ing* and 10% (3 items) are attributed to words featuring the *-y* suffix. The remaining 23.33% (7 items) encompasses relatively less common suffixes, collectively classified under the category of *others*. Consequently, the subsequent section aims to provide a deeper analysis of the *-er* suffix.

Bauer and Huddleston (2002) highlight the extensive productivity of the *-er* suffix. They state (2002: 1698) that nouns with the *-er* suffix encompass a wide spectrum of meanings, with the prototypical instance being a deverbal noun designating a person occupying the agent role in relation to the corresponding verb. In addition, *-er* attaches to many bases that are not verbs. Notable examples include nouns or proper names such as *executioner*, *golfer*, *freighter*, *petitioner*, *Londoner*, and *New Yorker* (2002: 1698). Furthermore, instances exist where the *-er* suffix attaches to numerals, adverbs, adjectives, and dephrasal compound bases, as



exemplified by words like *fiver*, *oncer*, *southerner*, *dogooder*, *fast-tracker*, and *nine-to-fiver* (2002: 1698). Notably, our analysis of the database indicates that words with the *-er* suffix predominantly attach to nouns, such as *mooner* (item 168) and *podcaster* (item 49), as well as numerals, such as *thirteenth* (item 174) and *twelfth* (item 175), or compound words, such as *game-changer* (item 133), *Big banger* (item 165), and *grassy-knoller* (item 167).

The agentive function is often exhibited by the *-er* suffix, as evidenced by words such as *podcaster* (item 49), *game-changer* (item 133), *Big banger* (item 165), *grassy-knoller* (item 167), *mooner* (item 168), *prepper* (item 170), *shout-downer* (item 171), *swift boater* (item 172), *tea partiers* (item 173). These examples might strongly suggest a high likelihood of the *-er* suffix being attached to a compound when it assumes an agentive role within a word. According to Bauer and Huddleston (2002: 1698), “deverbal *-er* nouns figure very productively in compounds such as *hairdresser*, *stage-manager*, etc.,” which suggests that when the suffix *-er* is attached to a compound, the resulting word form may still be considered a compound word rather than a word formed through suffixation from their view.

The analysis of these intriguing data warrants further investigation into the characteristics of derivation and compounding within the grammatical framework. As noted previously, there is a persistent debate in the literature on morphology regarding the relationship between derivation and compounding, as extensively discussed by notable scholars such as Anderson (1992), Lieher (1992), and Booij (2010). Although the present paper does not delve into this specific topic, we propose that the examination of neologisms, akin to those scrutinized in this study, holds significant value in informing and enriching this scholarly debate.

### 4.1.3 Prefixation

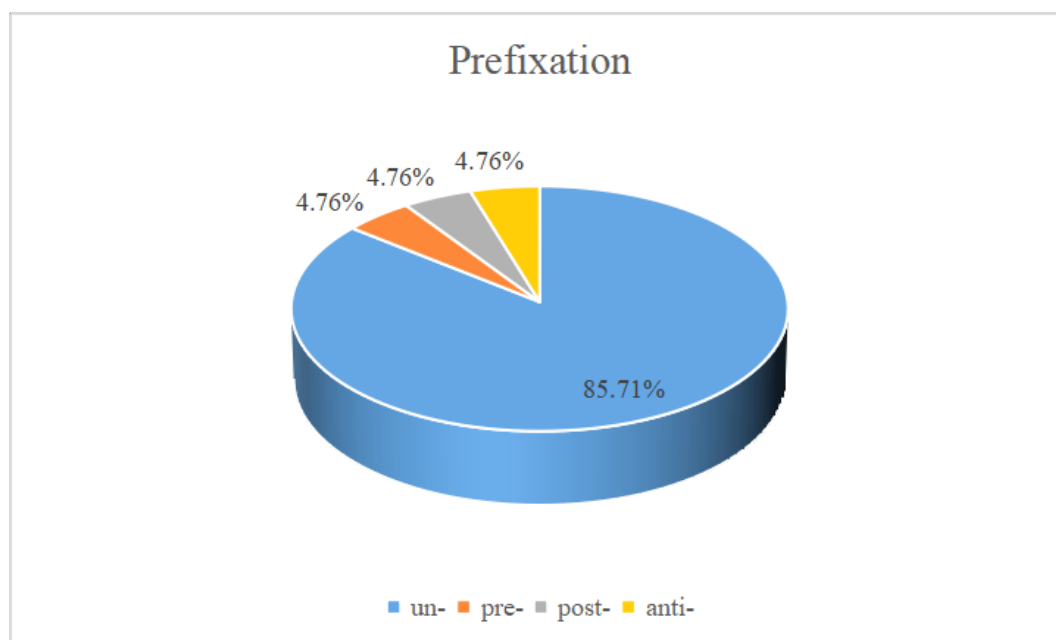


Figure 11: Distribution of different prefixes' productivity in the English prefixation group.

The data from prefixation reveal a significant characteristic, constituting 6.42% (21 items) of the whole dataset, with a specific focus on the prefix *un-*. Out of the 21 newly formed words with a prefix, a remarkable majority of 18 words (85.71% of prefixed words) exhibit the *un-* prefix. Another 4.76% (1 item for each prefix) of the words contain the prefixes *post-*, *pre-*, and *anti-* respectively. These findings emphasize the dominance of the *un-* prefix in the dataset analyzed.

The *American Heritage Online Dictionary* identifies two prefixes spelled *un-* in the English language. Primarily affixed to adjectives and participles used as adjectives, one of the *un-* prefixes conveys the meaning of *not* (e.g., *unable*, *unclean*, *unequal*, *unripe*, *unsafe*). While less commonly attached to nouns (e.g., *unbelief*, *unconcern*, *unrest*), a considerable proportion of the newly coined words contain *un-* attached to nouns; examples include *unagency* (item 193), *uncampaign* (item 195), *unconference* (item 196), *uncurriculum* (item 197) and *unschool*

(item 200). Meanwhile, the other prefix *un-* attaches to verbs and means reversal, as seen in *unclick* (item 204), *unpause* (item 208), *untag* (item 210).

Bauer and Huddleston (2002: 1687-1688) discuss negation in word formation and identify five prefixes that convey a negative meaning: *a-*, *dis-*, *in-*, *non-*, and *un-*. Their analysis reveals that the prefix *in-* is no longer productive in creating new words, while *a-* and *dis-* are rarely used for this purpose. However, the prefix *non-* exhibits a relatively high degree of productivity, whereas *un-* demonstrates the highest level of productivity among these prefixes. The data group analyzed in our investigation provides substantial evidence for the considerable productivity of the *un-* prefix.

Zimmer, Carson and Horn (2011) extensively examine the prefix *un-*. They explain that this prefix combines two etymologically distinct origins, functioning both as a negative prefix derived from Old English (*un-*, cognate with German *un-*, Latin *in-*, and Greek *a [n]-*) and as a reversative prefix derived from Old English (*on[d]-*, cognate with German *ent-* and Greek *anti-*). The *un-* prefix, encompassing both negative and reversative elements, has been creatively employed and highly productive since early Modern English. In the Oxford English Dictionary (2nd ed.), out of the 314 words credited to Shakespeare with the prefix *un-*, the majority consists of adjectives and adverbs conveying a negative meaning, while a few instances include negativized nouns and reversative verbs. However, the inclusion of negative nouns and reversative verbs in contemporary word formation continues to enrich the linguistic landscape. This productivity is also evident in our dataset, with newly introduced *un-* words predominantly classified into these two types:

3.a. Negative nouns: *unagency*, *unbanking*, *uncampaign*, *unconference*, *uncurriculum*, *unmarketing*, *unpolitician*, *unschool*, *unschooling*.

b. Reversative verbs: *unban*, *unbrick*, *unclick*, *unerase*, *unfragment*, *unfuck*, *unpause*, *unselect*, *untag*.

Zimmer, Carson and Horn (2011) also highlight the influence of 7UP's *Uncola* advertising slogan on the emergence of new *un*-nouns in American English. Originating in 1967, this campaign positioned 7UP as an alternative to major competitors like Coke and Pepsi. This influence extends to other domains, including politics, where terms like *uncandidate* and *unpolitician* have emerged to refer to alternative candidates who employ *uncampaign* strategies. Moreover, the concept of *unagency* (item 193) involves *unmarketing* (item 198), while *unschool* (item 200) offers an alternative educational approach through an *uncurriculum* (item 197). These linguistic innovations demonstrate the enduring impact of the *Uncola* advertising campaign in various domains.

A notable subset of the newly introduced *un*-nouns in our study follows a deverbal pattern, adopting the structure *un-X-ing*. Examples such as *unbanking* (item 194) and *unmarketing* (item 198) exemplify this form, which is understood as [*un-[X-ing]*] rather than [*un-X*]-*ing*. In this structure, the *un-* prefix negates the verbal noun, rather than creating a reversative *un*-verb that is subsequently nominalized (Zimmer, Carson & Horn, 2011). Although considered "rare" in contemporary English, this type of word formation has shown recent productivity, as evidenced by our data. Additionally, intransitive *un*-verbs can be derived from *un-X-ing* verbal nouns through back-formation (Zimmer, Carson & Horn, 2011). For example, an individual knowledgeable in *unschooling* (item 201) possesses the ability to *unschool* (item 200).

Furthermore, it is worth noting that while back-formed *un*-verbs are relatively uncommon, the reversative category has experienced significant growth in recent years, particularly within the realm of technology (Zimmer, Carson & Horn, 2011). Examples include *unban* (item 202), meaning the removal of restrictions preventing individuals, such as interactive video game players or online forum participants, from engaging due to their misconduct; *unbrick* (item 203), referring to the repair of electronic devices that have become

unresponsive; and, *unpause* (item 208), involving the resumption of an interrupted state, whether it be an electronic device or software. These instances exemplify the productivity of the reversative category with the subject domain of technology.

Overall, the analysis of prefixation reveals the prominent role of the *un-* prefix in word formation. The *un-* prefix combines negative and reversative meanings, and both of them are productive. The impact of advertising campaigns, such as 7UP's *Uncola*, has contributed to the emergence of new *un-*nouns in American English. Furthermore, the structure of *un-X-ing* verbal nouns and the formation of intransitive *un-verbs* through back-formation add to the diversity of word formation patterns observed in contemporary English.

#### 4.1.4 Compounding

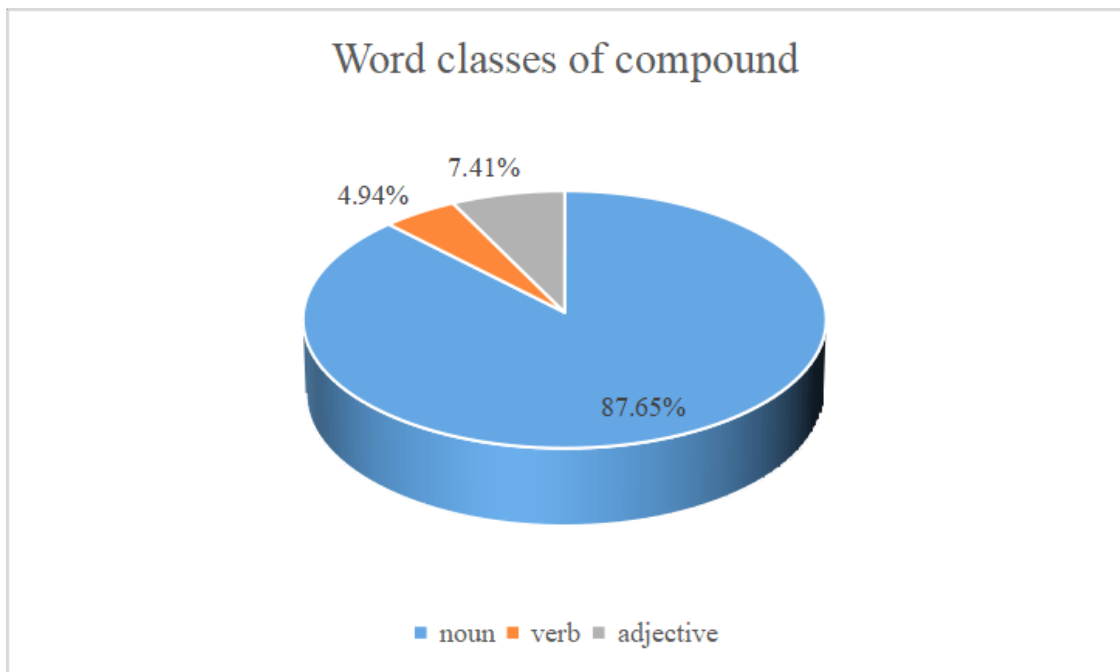


Figure 12: Distribution of word classes of English compound words.

Compounding emerges as the predominant characteristic within our English database, representing a significant proportion of 25.08% of the entire dataset. As indicated by Bauer and Huddleston (2002: 1646), nouns constitute the most abundant and diverse group of compounds. This observation is further supported by the data in our database, in which nouns constitute

87.65% of the total. Moreover, from a morphological standpoint, a salient characteristic of these compound nouns in our database is the consistent use of the combining form *-core*.

The combining form *-core* has been previously documented by the *Oxford English Dictionary* (OED3), derived from the term *hardcore*. According to the OED3, *-core* serves as the second component in various compounds that designate subgenres of popular music, particularly those characterized by heightened intensity or extremity, such as punk, grunge, techno, and heavy metal. Historical evidence traces back to the emergence of *speedcore* in 1985 (Zimmer and Carson, 2012). Within our English data corpus, notable instances include *crunkcore* (item 222), *electronicore* (item 223), *nerdcore* (item 227), and *Raggacore* (item 228).

Another noteworthy word-forming element in compound nouns that has captured our attention is the lexical component *-gate*, originating from the term *Watergate*. Our analysis reveals the presence of newly coined words like *Boobgate* (item 1) and *Katrinagate* (item 34), where *-gate* functions as a productive word element. It is important to highlight that the *noun+noun* construction constitutes the largest subgroup of compounds (Bauer, 1983, p.202; Bauer and Huddleston, 2002, p.1647). Furthermore, we have also observed instances where the component *gate* itself acts as a noun within the *noun+noun* compound structure rather than a combining form that refers to a political scandal. Examples include *gate rape* (item 184), denoting the intrusive screening of passengers by airport security, and *gate louse* (item 262), a plural form referring to individuals who impede the boarding gate while awaiting entry onto an airplane.

#### **4.1.5 Blending**

Blends, comprising 14.37% of the total dataset, demonstrate a significant presence and rank as the third most prominent category in terms of overall frequency.

One notable example of the typical blending form in our dataset is *Brangelina* (item 54), which directly combines the names of Hollywood actors Brad Pitt and Angelina Jolie to create a couple's name. Our data also includes other neologisms generated from the blending of proper personal noun names, such as *TomKitten* (item 86), a nickname for Suri Cruise, the child of Hollywood actors Tom Cruise and Katie Holmes, and *Billary* (item 108), a term that jocularly represents William Jefferson "Bill" Clinton and Hillary Rodham Clinton as members of a joint presidency.

Furthermore, it is important to acknowledge that both the presidential election cycle and the subsequent presidential administration period in the U.S. may give rise to a fresh wave of innovative blends within the political realm. There are several Obama-blends in the data; for instance, *Obamacan* (item 126) [Obama + Republican] refers to a Republican who supported Democrat Barack Obama for the US presidency in 2008. Similarly, *Obamanomics* (item 136) [Obama + economics] signifies the economic policies implemented during the Obama administration. Since 2008, the incorporation of Barack Obama's name (Obama) as a combining form has proven highly fruitful (Zimmer and Carson, 2012), resulting in a diverse array of Obama-inspired terms, including *Obamaloney* (item 234) [Obama + baloney]. It is noteworthy that this trend extended beyond Obama alone, as Mitt Romney, the Republican Party candidate for President of the United States in the 2012 election, also contributed to the lexicon with terms like *Romnesia* (item 237) [Romney + amnesia] and *Romneyshambles* (item 238) [Romney + omnishambles]. However, the extensive utilization of such name-based wordplay may have fostered a perception that the major-party candidates were interchangeable, as exemplified by blended terms like *Robama* (item 236) [Romney + Obama].

Additionally, our analysis has yielded intriguing findings that indicate the emergence of brand names through the process of blending, such as *Groupon* (item 269) [group+coupon] and *Instagram* (item 270) [instant + telegram].

The data exhibits a diverse array of blends. Noteworthy examples include *Pajamahadeen* (item 10), which combines *pajama* and *(muja)hidden*, *envirologistics* (item 42), merging *environmental* and *logistics*, and *earmarxist* (item 110), blending *earmark* and *Marxism*.

In conclusion, although blends have received some attention from scholars such as Pound (1914), Lehrer (2003), and Čolić (2015), they have not been extensively explored within the linguistic community compared to other processes of word formation. The notable prevalence of blends in our database, coupled with the diversity of blending constructions, underscores their significance in productive word-formation in English.

#### **4.1.6 Neoclassical compounding**

Our analysis reveals that neoclassical compounds constitute a relatively small proportion of the data, representing only 0.31% (1 item) of the total. The limited prevalence of neoclassical compounds in our dataset can be attributed to the challenge of distinguishing them from regular compounds.

Based on this criterion, our analysis has identified only one English neologism: *cisphobia* (item 297). This term meets the specified classification criteria as a neoclassical compound, combining the Latin prefix *cis-* meaning *on the side* with the suffix *-phobia* indicating *fear of* in Greek.



#### 4.1.7 Conversion

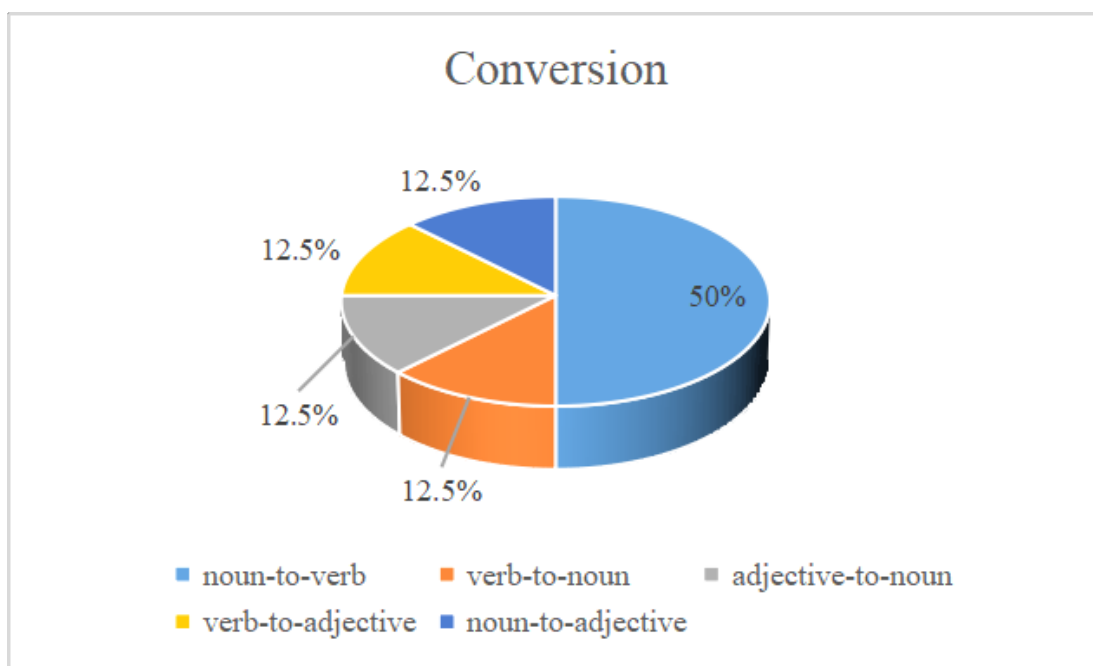


Figure 13: Distribution of the conversion patterns in English data.

Bauer and Huddleston (2002: 1640) highlight that conversion entails the alteration of a word's syntactic category without accompanying changes in form, exemplified by the transformation of the adjective *humble* into the verb *humble* and the verb *attempt* into the noun *attempt*.

According to Bauer (1983: 229), conversion manifests primarily through four major types: *noun-to-verb*, *verb-to-noun*, *adjective-to-noun*, and *adjective-to-verb*. In our English database, we have identified 2.45% (8 items) of the total as undergoing conversion. While this percentage may appear relatively small in relation to the overall dataset, our analysis has yielded interesting findings. Notably, a significant portion, comprising 50% of the identified cases, involves conversion from *noun-to-verb* without any accompanying changes in form. Illustrative examples include *lance* (item 89), *jack* (item 138), *demise* (item 274) and *drone* (item 275).

There are both examples of *verb-to-noun* and *adjective-to-noun* conversion. It is worth noting that while we could not find specific data on *adjective-to-verb* conversions, our findings did indicate that *verb-to-adjective* conversion exists and is exemplified by the case of the word *fetch* (item 18). The *noun-to-adjective* category similarly is also found and exemplified by instances such as *amoeba* (item 296).

We have observed a notable instance that highlights a shift in the word class from verb to noun. This change is exemplified by the term *occupy* (item 217), which is originally defined as a verb in the sense of *to gain access to and remain in (a building, etc.) or on (a piece of land), without authority, as a form of protest (OED3 1920–)*. However, within the context of our analysis, *occupy* has undergone a transformation and now functions as a noun. Specifically, it is used as an umbrella term referring to demonstrations and protests centered around economic inequalities. This shift in word class from verb to noun also is indicative of the evolving nature of language and the potential for lexical adaptation and semantic expansion.

We propose that the observed instances of conversion in our English corpus can be associated with semantic neologisms. Further exploration of this phenomenon is presented in Section 4.2.

#### **4.1.8 Other minor word formations**

In this section, we briefly mention other minor word formations that we have found. They are examples of acronyms (3.67%, 12 items), abbreviations (1.53%, 5 items), clippings (1.22%, 4 items) and reduplication (0.31%, 1 item).

While abbreviations are pronounced as sequences of letters and are typically written with full stops after each letter or spelled out as ordinary words, such as *GWOT* (item 43: *Global War On Terrorism*), and *MOOC* (item 266: *Massive Online Open Course*), acronyms are pronounced as ordinary words, with the letters having their characteristic phonological

value (Bauer and Huddleston, 2002, p.1632-1633), for instance, *DOMA* (item 2: *Defense of Marriage Act*) and *bogo* (item 31: *buy one get one*).

Clipping, as a linguistic phenomenon, involves the truncation of a word or phrase by removing a portion of it, resulting in a shorter sequence that retains its phonological integrity (Bauer and Huddleston, 2002, p.1634), examples include *Insta* (item 254: *clipping of INSTAGRAM*), *feels* (item 260: *clipped form of feelings*), among others.

There is a single instance of reduplication: the word *cray-cray* (item 243), where the initial syllable of *crazy* is reduplicated. This phenomenon lies beyond the parameters of Bauer and Huddleston's (2002) classification and was not explicitly addressed in their categorization.

#### 4.2 Semantic neologisms

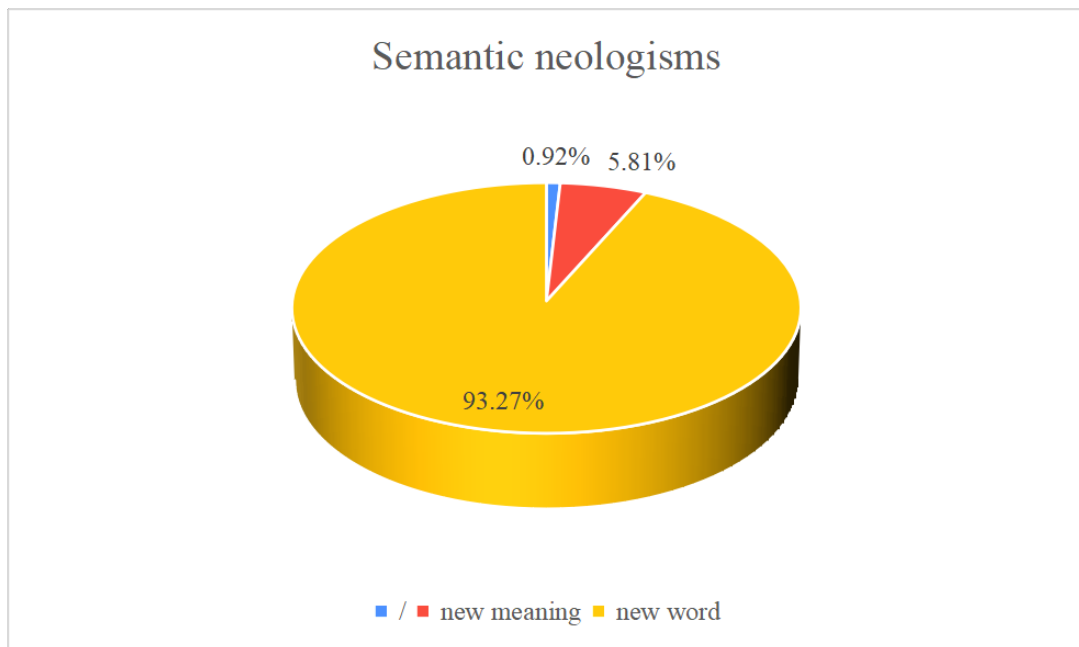


Figure 14: Distribution of the semantic neologisms' productivity in English data.

---

<sup>8</sup> In the diagram, the term *new meaning* refers to the concept of semantic neologisms, which involves assigning new meanings to existing words. On the other hand, *new word* primarily refers to formal neologisms while "/" includes the word elements such as affixes and combining forms.

Semantic neologisms play a role in revitalizing and expanding the expressive and communicative capabilities of languages (Creus and Julià-Muné, 2014). This type of neologism involves adding a new meaning to an existing, old word. In our English data corpus, we have detected semantic neologisms amounting to 5.81% of the total number of neologisms (19 items).

The identification of semantic neologisms presents a challenge, as there is often ambiguity in determining whether a new word represents a regular neologism with a novel meaning or an additional meaning attached to an existing word. While *Among the New Words* occasionally offers valuable lexicographic information from the OED regarding both the original and observed meanings of a new word, aiding in the identification of semantic neologisms with added meanings, this occurrence might be relatively infrequent. For example, in the case of words such as *angel* (item 13), *surge* (item 85), and *basic* (item 302), the original meanings are not provided in *Among the New Words*.

**surge** *n* Substantial, rapid increase in troop strength **1990** Dec 10 David A Fulghum *Aviation Week & Space Technology* 81 (LexisNexis) MAC officials justified keeping the units on active duty because the Air Force had to consider potential demands such as the unexpected outbreak of hostilities or another surge in the Middle East buildup. **2006** Nov 18 Tim Harper *Toronto Star* A11 (LexisNexis) It may seem counterintuitive, but there are calls—including one from the man who could be the next U.S. president—to flood Iraq with thousands more American troops, a so-called surge, in one last bid to win a war that looks more and more unwinnable. **2007** Jan 11 Casey Ross *Boston Herald* 4 (LexisNexis) President Bush and Sen. Edward M. Kennedy could be on a constitutional collision course over the Iraq war, as the Republican commander in chief pushes a troop surge while the Bay State Democratic powerhouse musters all his political strength to scuttle the plan in Congress.

Figure 15: One example of semantic neologisms: *surge* (item 85) that *Among the New Words* omits relevant lexicographic information (OED) in the journal.

Nonetheless, these could still be recognized as semantic neologisms by our detection since they are common words in English. However, beyond such relatively straightforward cases, our study may overlook some potential semantic neologisms. This remains a limitation

in our approach, as we did not employ computational detection tools or any other automatic procedures for the classification.

In addition, it is noteworthy that a considerable portion of these semantic neologisms exhibit meanings that have not been widely integrated into standard dictionaries. This phenomenon could be attributed to the limited prevalence of these new connotations. For instance, *Among the New Words* includes the word *angel* (item 13) with its newly defined meaning as *soldier killed in action*, which may still remain unfamiliar to a significant portion of the American population.

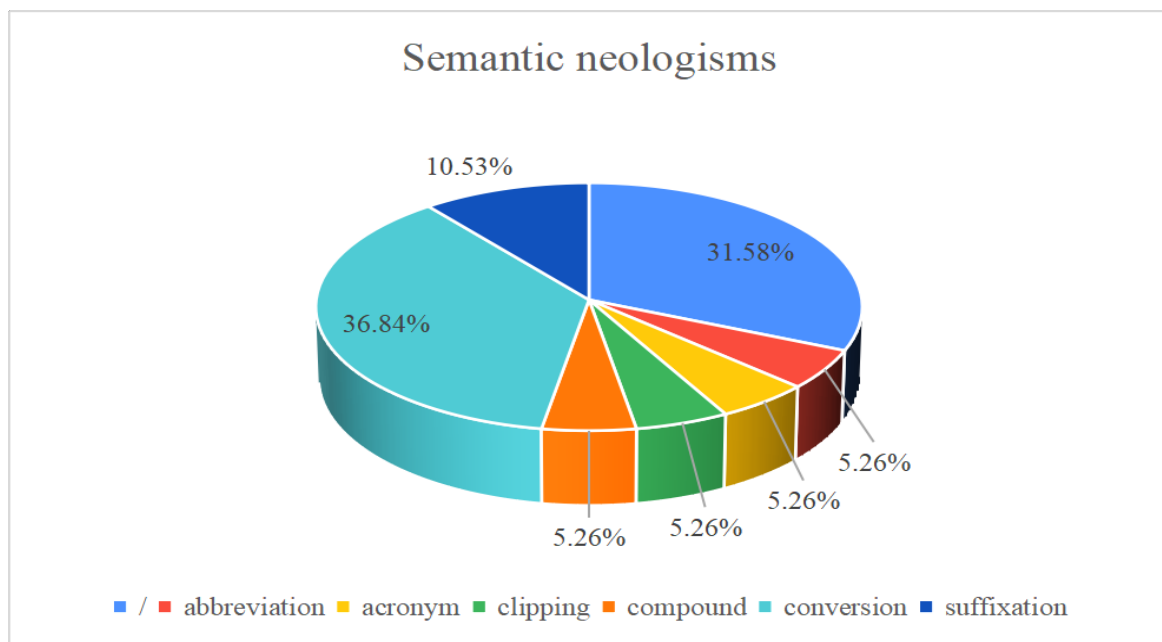


Figure 16: Distribution of the semantic neologisms' word formation processes in English data.

Moreover, upon examining the morpheme classification of these semantic neologisms, we see that the predominant lexical formation process employed is conversion, encompassing approximately 36.84% of the entire category of semantic neologisms. To illustrate this phenomenon, for example, consider the word *fetch* (item 18), which undergoes conversion to acquire a novel adjective meaning, denoting something cool, awesome, or good. This new

meaning emerges from its original definition as a transitive verb, specifically referring to the act of going in search of and conveying or bringing back.

Given that the lexical formation process of conversion facilitates the modification of word classes while retaining the same spelling form, it significantly contributes to the expansion of semantic meanings within a language. Considering this characteristic, we propose that the structural mechanism of conversion may serve as a bridge connecting formal neologisms to semantic neologisms.

### **4.3 Lexicalization of phrases**

The examination of our English database reveals a noteworthy presence of lexicalized phrases, occupying a substantial position as the second most prevalent category with a share of 24.16% in the overall dataset, closely trailing compounds in number. This observation substantiates our initial hypothesis. Considering the nature of *Among the New Words*, which encompasses a broader range of informal language rather than focusing on dictionary supplementation, the prevalence of phrases is to be expected.

Within our analysis of the phrase group, a distinctive characteristic emerges in the form of rhyming phrases comprising two nouns (*N and N*) or two verbs (*V and V*). This particular pattern represents a salient feature within the dataset, as both the *N and N* and *V and V* patterns exhibit notable phonological and morphological similarities. Examples of such rhyming phrases include *hats and bats* (item 146), *mag and bag* (item 147), *chew and screw* (item 142), *chalk and talk* (item 151), among others.

### **4.4 Borrowing**

Our analysis indicates that around 2.75% (9 items) of the total dataset consists of borrowed words. It is worth noting that a considerable number of these borrowed words originate from Asian languages, including examples such as *luanqibaozhao* (item 21) from Chinese, *freeter* and *sudoku* from Japanese, and *Gangnam Style* (item 246) from Korean.

Therefore, we suggest that it may be more or less a reflection of the increasing influence of Asian culture on American English.

#### 4.5 Other characteristics: the role of analogy and word elements

Although it is beyond the scope of this paper to delve into an exhaustive analysis of each item in the database, we feel compelled to dedicate a brief section to the presentation of select findings that are still worth additional attention.

Our primary focus in this discussion centers on the role of analogy in word formation. Analogical formations involve the creation of new words based on existing lexemes, without leading to a series of productive formations (Bauer, 1983). The term *pope-squatter* (item 50) is an example of a compound word formed by combining *pope* and *squatter*. It refers to individuals who register domain names with the intention of profiting from them, specifically targeting names associated with potential future popes. This neologism can be seen as an analogy to *cybersquatter*, as both terms describe individuals seeking financial gain through domain name registrations. While *cybersquatters* target domain names similar to existing trademarks or brand names, *pope-squatters* focus on names potentially corresponding to anticipated future popes.

Another example of analogical formation is *Mortaritaville* (item 91), a nickname for Balad US Air Base, resembling the well-known song *Margaritaville* by Jimmy Buffet. This term combines *mortar* and *Margaritaville*, and the new word *Mortaritaville* has the same syllable structure as the word it is based on.

In both cases, these analogical terms demonstrate that the analogy affects both form and meaning, which also contributes meaningful additions to the lexicon of language users.

Another significant finding that deserves to be noted is that some word elements seem to participate in more than one kind of morphological formation, for example the combining form *cis-* that appears in *ciscentric* (item 297), *cisness* (item 298), *cisphobia* (item 299),

*cissexism* (item 300), respectively. As previously mentioned, the lexical item *cisphobia* can be classified as a neoclassical compound due to its formation pattern. However, *cisness* demonstrates the process of suffixation, while *ciscentric* and *cissexism* belong to the category of compound vocabulary because neither *centric* nor *sexism* can be considered neoclassical formants.

The prefix *cis-* is a rare antonym of *trans-* according to Bauer and Huddleston (2002: 1687), and both prefixes share a close linguistic and conceptual relationship, especially within the domain of gender and sexual identity in contemporary English. The expanding meaning of *trans-* to include a broader range of gender identities, as observed by Zimmer, Solomon, and Carson (2014), has brought attention to its counterpart, *cis-*, which now serves as a combining form for describing cisgender identity and has gained prominence gradually in society.

Furthermore, we have observed the presence of combining forms and affixes, including notable instances such as *-sauce* (item 189) as a combining form and *-cation* (item 131) as a suffix. These findings underscore the significant role played by word elements in the process of generating new vocabulary.

In conclusion, our analysis highlights the remarkable diversity of word formation observed in the English database, suggesting numerous avenues for future exploration. While it is beyond the scope of this study to comprehensively examine every intricate detail, the significance of these findings emphasizes the need for thorough investigation in future research efforts.



## 5 Analysis of Spanish Neologisms in *Banco de Cervantes*

### 5.1 Formal neologisms

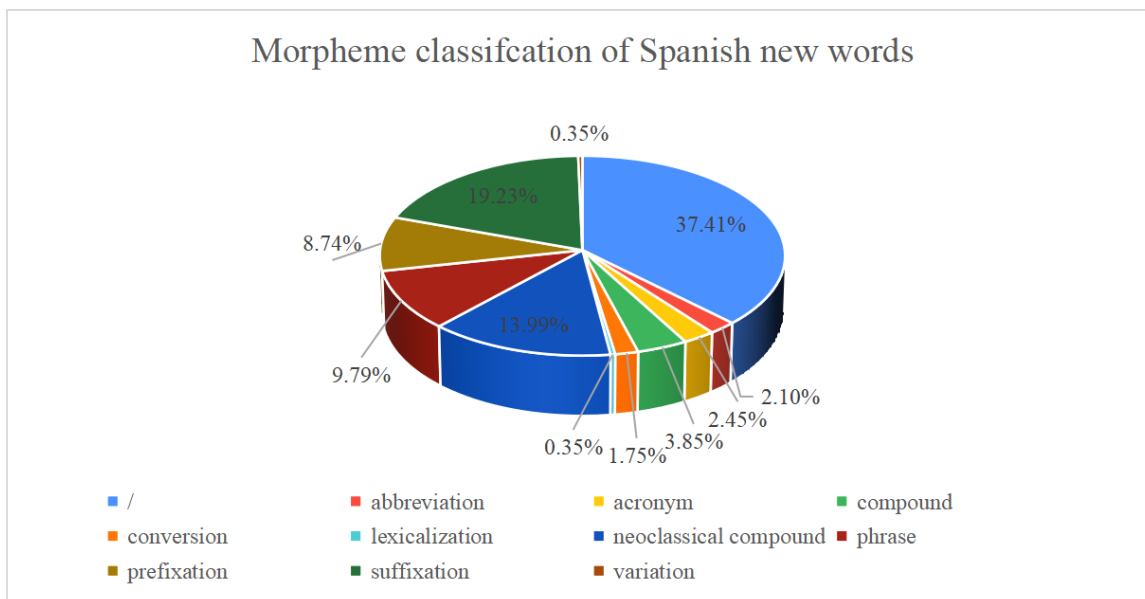


Figure 17: Distribution of different word formation processes' productivity for Spanish data.

In this section, we discuss the distribution patterns of word formation processes observed in Spanish neologisms, utilizing data extracted from the *Banco de Cervantes* database.

Our analysis reveals that a significant proportion, precisely 37.41% (107 items), of the neologisms extracted from the *Banco de Cervantes* database are designated with the symbol "/" indicating the neologisms without a specific morpheme classification, as well as those classified as *others* in *Banco de Cervantes*. In addition, we see that suffixation is the most common word formation process for neologisms. This particular category constitutes a substantial portion, accounting for 19.23% (55 items) of the entire dataset, and is nearly twice as common as prefixation, which represents 8.74% (25 items) of the dataset.

Neoclassical compounding emerges as the second most frequent word formation mechanism, accounting for 13.99% (40 items) of the dataset, while the regular compounds only amount to 3.85% (11 items) of the total.

The category of phrase<sup>9</sup>'s group exhibits a moderate representation within the overall dataset, accounting for 9.79% of the data.

Other, minor types of word formation such as acronyms and abbreviations amount to 2.45% (7 items) and 2.10% (6 items) of the total, surpassing examples of conversion (1.75%, 5 items).

In *Banco de Cervantes*, the category of *variation* is defined as an orthographic formal variant encompassing new words that do not undergo changes in morphology or syntax (Cabré, 2006). For instance, the word *infraestructura* can be considered an orthographic formal variant of *infraestructura* (Cabré, 2006). Additionally, the concept of *lexicalization* in neologisms involves the formation of new words through the lexicalization of an inflected form, typically derived from a specific verb paradigm (Cabré, 2006). However, both categories account for only 1 item of the total respectively, indicating their relatively minor significance within the entire database.

Overall, suffixation stands out as the most frequent word formation strategy among Spanish neologisms. It demonstrates a higher degree of productivity in comparison with prefixation. The presence of neoclassical compounds is also noteworthy. This distinctive pattern can be attributed to the inherent linguistic characteristics of the Spanish language, which has its linguistic roots in Latin. The propensity of Spanish to draw upon Latin as a primary

---

<sup>9</sup> We further classify the category labeled by *Banco de Cervantes* as *formado por sintagmación* as the lexicalization of phrases.

source for the derivation of new words is not only unsurprising but also an enduring linguistic tradition that has persisted throughout the course of its rich history.

Additionally, the moderate presence of phrase neologisms can be ascribed to the nature of the Spanish neologism project itself. Specifically, since the project is focused on dictionary supplementation, the inclusion of phrases becomes less probable. As mentioned earlier, when compared to neologisms undergoing other word formation processes, phrase neologisms possess a relatively lower likelihood of being incorporated into the dictionary because dictionaries are designed to define individual words, as opposed to words combined into phrases (although most dictionaries contain some definitions of fixed phrases). Hence, it is reasonable to note that the number of neological phrases reported on in *Banco de Cervantes* is limited.

### 5.1.1 Suffixation

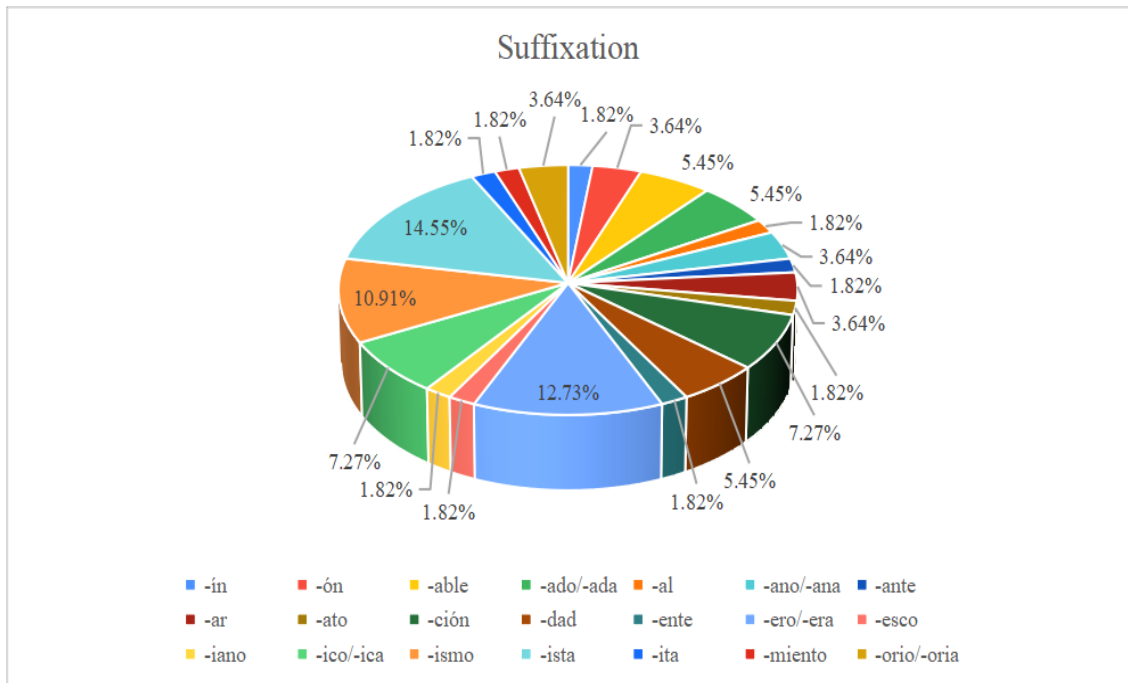


Figure 18: Distribution of the wide variety of suffixes' productivity in the Spanish suffixation group.

Several productive suffixes are present in the data. The *-ista* suffix demonstrates high productivity, similar to the *-ismo* suffix, and serves to designate individuals in Spanish nouns (Real Academia Española, 2009: 473). Among the various suffixes observed in our dataset, the *-ista* suffix emerges as the most productive, accounting for 14.55% of the entire category (8 items). Notable examples include *uniformista* (item 73), *yihadista* (item 77), *unilateralista* (item 99), *zenista* (item 104), among others. Additionally, the productivity of the *-ismo* suffix closely follows, with a rate of 10.91% (6 items). Illustrative examples include *unilateralismo* (item 21) *kirchnerismo* (item 63), *internismo* (item 191).

Another highly productive suffix observed in our database is *-ero/-era*, accounting for 12.73% of the data (7 items), ranking just behind the suffix *-ista* in prominence. The *Real Academia Española* (2009: 466) notes that the suffix *-ero/-era*, originating from the Latin *-arius*, *-a*, *-um*, is widely recognized as one of the most productive suffixes in the Spanish language for constructing terms associated with professions or occupations (e.g., *barbero*, *cocinera*, *jardinero*). An example of its usage is the term *graffitero* (item 59), which refers to individuals involved in the creation of graffiti art or graffiti artists.

A total of 21 different suffixes are present in the data, which showcases the remarkable diversity in the Spanish derivational system. Notably, the suffix *-ción* stands out for its high productivity in the derivational paradigm of deverbal action nouns (Real Academia Española, 2009: 346), exemplified by words like *dictaminación* (item 82), which is derived from the verb *dictaminar*. Additionally, the suffix *-dad* allows the formation of nouns from adjectives, and it is considered one of the most productive suffixes in the paradigm of nouns referring to qualities (Real Academia Española, 2009: 413). For instance, we find examples like *ordinalidad* (item 93) derived through this suffix. Furthermore, our analysis has also revealed the presence of other notable suffixes, including *-ico/-ica*, *-ado/-ada*, *-ano/-ana*, *-orio/-oria*, *-able*, *-miento*, *-ón*, among others.

### 5.1.2 Prefixation

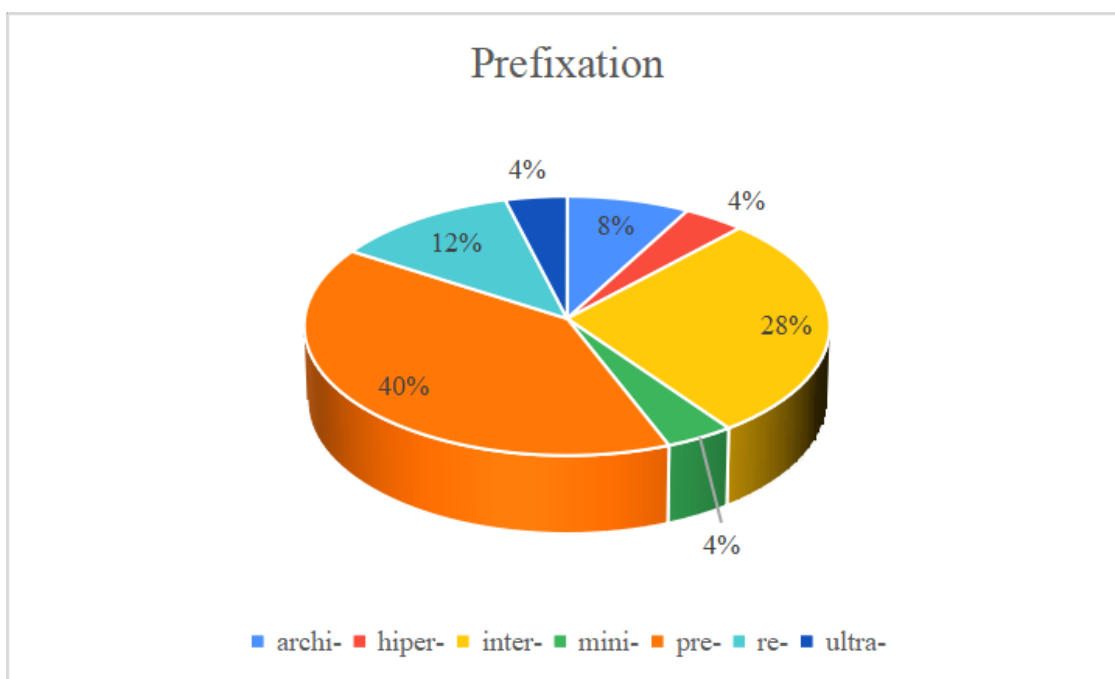


Figure 19: Distribution of the prefixes' productivity in the Spanish prefixation group.

Prefixation accounts for approximately half the number of neologisms as suffixation (8.74% of the total, 25 items). A few prefixes are very productive.

Based on our findings, which align with the observations of the *Real Academia Española*, it is evident that prefixes such as *pre-* and *inter-* are very productive in modern Spanish (Real Academia Española, 2009: 667). Specifically, the prefix *pre-* stands out as the most frequent, accounting for a substantial 40% of the entire group of prefixed forms (e.g., item 19: *prealerta*; item 42: *precandidata*; item 68: *preadolescencia*; item 94: *pre-consciente*),. The second most productive prefix is *inter-*, with 28% of all prefixed forms (e.g., item 9: *interoperatividad*; item 35: *intermodal*; item 139: *intersector*). These empirical findings reinforce the notion that these prefixes continue to play a significant, active role in the linguistic consciousness of Spanish speakers (Real Academia Española, 2009: 667).

Additionally, our analysis has revealed the presence of the prefix *re-* indicating 'reiteration' and the prefix *archi-* denoting 'intensity,' as documented by the Real Academia

Española (2009: 670). We also see relatively less productive prefixes in our dataset, including *mini-*, *hiper-*, and *-ultra*.

### 5.1.3 Compounding

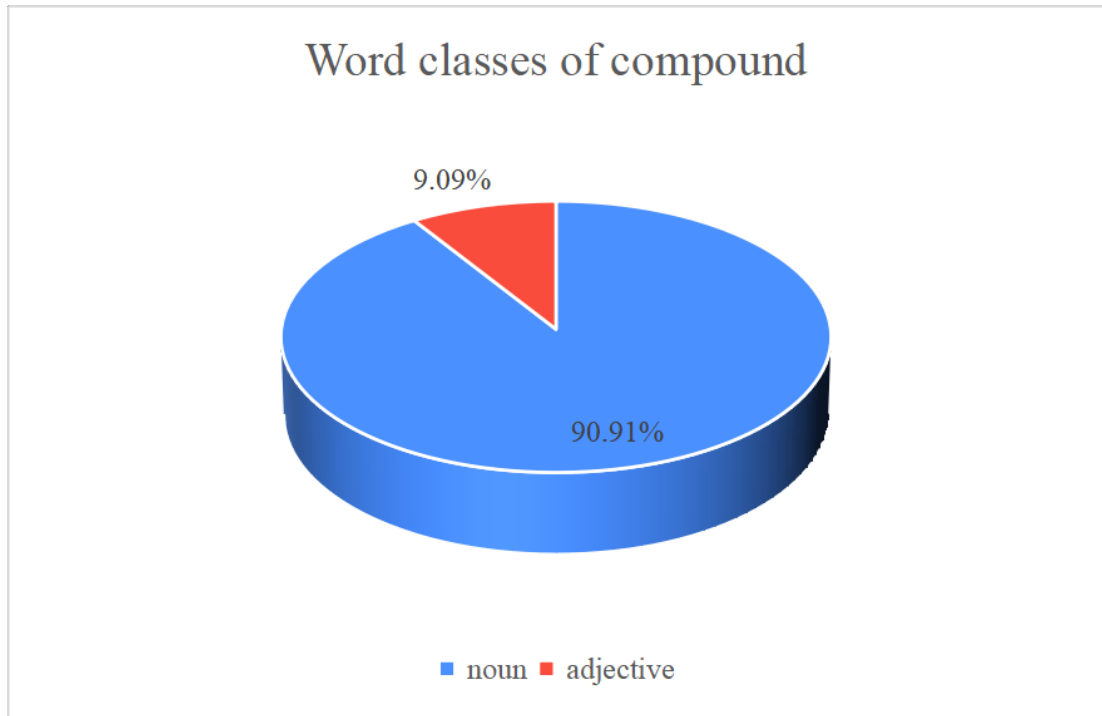


Figure 20: Distribution of word classes of Spanish compound words.

The overall amount of compounds is relatively less representative, with only 3.85% of the total (11 items). Nevertheless, our analysis reveals that among the limited number of compound words, a significant majority, amounting to 90.91% (10 item), are nouns, while the remaining 9.09% (1 item) are classified as adjectives.

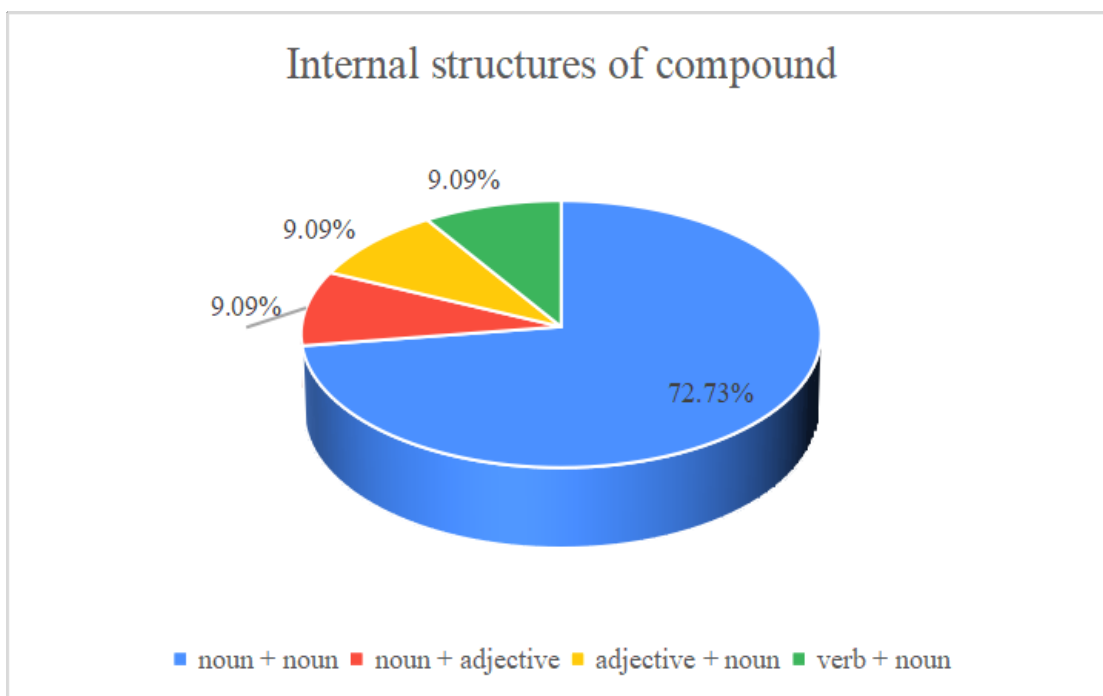


Figure 21: Distribution of the internal structures of Spanish compound words.

The data contain compounds with several distinct internal structures. The predominant structure observed is *noun+noun*, which accounts for 72.73% of the total. Examples of this structure include *traje-pantalón* (item 124), *zona euro* (item 130), and *fotoperiodismo*<sup>10</sup> (item 266). Additionally, other patterns such as *noun+adjective* (e.g., item 266: *vinotinto*), *adjective+noun* (e.g., item 150: *treceañero*), and *verb+noun* (e.g., item 154: *tragatapas*) are also found but are much less prevalent (only 1 item for each type).

#### 5.1.4 Neoclassical compounding

The presence of neoclassical compounds emerges as a notable characteristic within the database, occupying a prominent second position and encompassing 13.99% of the dataset.

---

<sup>10</sup> According to Cabré (2009: 232), *fotoperiodismo* has been classified as belonging to the category of *composición culta*. However, in *Banco Cervantes*, this word is categorized under the regular compound group.

According to Cabré (2006), neoclassical compounding can be classified into three types: (1) neoclassical prefix + neoclassical suffix; (2) neoclassical prefix + root; (3) root + neoclassical suffix:

*Por composición culta (FCULT): neologismos formado a partir de uno de los procesos siguientes:*

- *una forma prefijada culta y una forma sufijada culta. Ej.: biogenia, aerófago, megápolis;*
- *una forma prefijada culta y un radical. Ej.: autoexigencia, fotoperiodismo, microvestido, xenotrasplante;*
- *un radical (propio de la lengua o bien prestado de otra lengua) y una forma sufijada culta. Ej.: clasicómano, normógrafo, simpaticoide.*

(Cabré, 2006: 232)

Our analysis reveals that neoclassical compound neologisms primarily fall into the second category, characterized by the combination of a neoclassical prefix with a root. Notably, these neoclassical compounds feature the prefix *foto-*, derived from the Greek word *phōs* and *phōtos*, which signifies *light* (Real Academia Española, 2009: 783). Examples of these neologisms include *fotoperiódico/-ca* (item 6), *fotorama* (item 84), *fotoleyenda* (item 162), *fotolavado* (item 214), and *fotolaminado* (item 246). Additionally, the element *mini-* also demonstrates robust productivity within our dataset, as evidenced by examples such as *minipaga* (item 13), *minicasa* (item 39), *minivehículo* (item 65), *minimarket* (item 117), *minisandwich* (item 143), and *miniobra* (item 221).

In addition, *Real Academia Española* (2009: 782) further emphasizes the utilization of lexical roots from scientific and technical vocabulary in neoclassical compounds. This assertion is supported by our dataset, exemplified by roots like *organo-* in *organoclorar* (item 15), which is a term used in biology and chemistry. Similarly, the element *video-* in *videovigilancia* (item 48/item 204) is predominantly employed in fields associated with visual media, electronics, and communication technology, among others.



### 5.1.5 Conversion

Conversion represents a mere 1.75% (5 items) of the entire dataset, featuring various conversion patterns. For instance, we observe instances of *adjective-to-noun* conversion, exemplified by words like *histórico* (item 8). Conversely, we encounter *noun-to-adjective* conversion, as seen in the case of *concertista* (item 29). The identification of these conversion patterns was facilitated by the contextual information provided by *Banco de Cervantes*, enabling us to ascertain the word class. Additionally, their explicit grammatical type markings and cross-referencing with the *Diccionario de la lengua española* helped to validate our identifications.

However, it is worth noting that for certain words, despite the system directly indicating their parts of speech as new words, the absence of original definitions and word classes from *Banco de Cervantes* hampers precise determination of the specific conversion process. For instance, the word *compostar* (item 133) can be inferred to be a verb based on a root plus the suffix *-ar*. Although *Banco de Cervantes* designates it as a transitive verb undergoing conversion, this aligns with the word class provided by the *Diccionario de la lengua española*. Whereas in English conversion involves a change in lexical category with no change in form, the change from noun stem to verb stem has traditionally been categorized as conversion in Spanish, even though there is a change in form (all verbs must belong to a conjugation and have a form that is part of the verb paradigm, hence the thematic vowel /a/ and the infinitive marker /r/) (Cabré, 2006).

## 1 compostar

---

Contexto: El documento inicial del Plan de Residuos (2002) preveía \*compostar\* en 2016 un total de 21.780 toneladas de residuos orgánicos.

Categoría gramatical: Verbo transitivo (v tr)

Tipo de neologismo: Formado por conversión (FCONV)

Aspectos tipográficos: Sin marca tipográfica (smt)

Fuente: Diario Vasco (PVN3)

Corpus de vaciado: Corpus escrito de prensa (pr)

Figure 22: The entry of *compostar* (item 133) registered in the *Banco de Cervantes*.

**compostar** Conjugar

### 1. tr. Transformar residuos orgánicos en compost.

Figure 23: The entry of *composter* (item 133) in the *Diccionario de la lengua española*.

#### 5.1.6 Other minor word formation

In terms of minor word formation processes, acronyms and abbreviations constitute 2.44% and 2.09% of the dataset (7 items and 6 items), respectively. For instance, the term *wifi* (item 75) is an acronym derived from the initial letters of *Wireless Fidelity*. *Variation* and *lexicalization*, on the other hand, represent the least prominent word formation processes, accounting for a mere 0.35% of the total (1 item).

We have identified only one instance of the *variation* process, namely the word *zionismo* (item 182), which serves as a variant spelling of the more commonly used term *sionismo*. This variation deviates from the standard or prevailing spelling.

Furthermore, *lexicalization* refers to the setting of an existing lexical meaning. As explained by Cabré (2006), this process often involves the lexicalization of a particular form

within an inflected, typically derived from a specific form within the verbal paradigm. Within our dataset, we have identified a solitary case of lexicalization: *lixiviado* (item 168). This term has been lexicalized to denote the liquid resulting from the percolation of a fluid through a solid, as inferred from the contextual information provided by *Banco de Cervantes*.<sup>11</sup>

## 1 lixiviado

---

**Contexto:** El alcalde Édgar Cesario Navarro Sánchez aseguró que lo que buscan es evitar que los asistentes puedan arrojar basura a los ductos que se colocaron para la extracción de biogás y \*lixiviados\* que generaron los desechos acumulados durante 25 años en ese vertedero a cielo abierto

**Categoría gramatical:** Nombre masculino (m)

**Tipo de neologismo:** Formado por lexicalización (FLEX)

**Aspectos tipográficos:** Sin marca tipográfica (smt)

**Fuente:** El Universal (M3)

**Corpus de vaciado:** Corpus escrito de prensa (pr)

Figure 24: The entry of *lixiviado* registered in *Banco de Cervantes*.

---

<sup>11</sup> We note that the definition of lexicalization used by *Banco de Cervantes* is much narrower than that used by Bauer (1983) or especially that used by Bosque (1982), which is usually the standard for analyses of Spanish.

## 5.2 Semantic neologisms

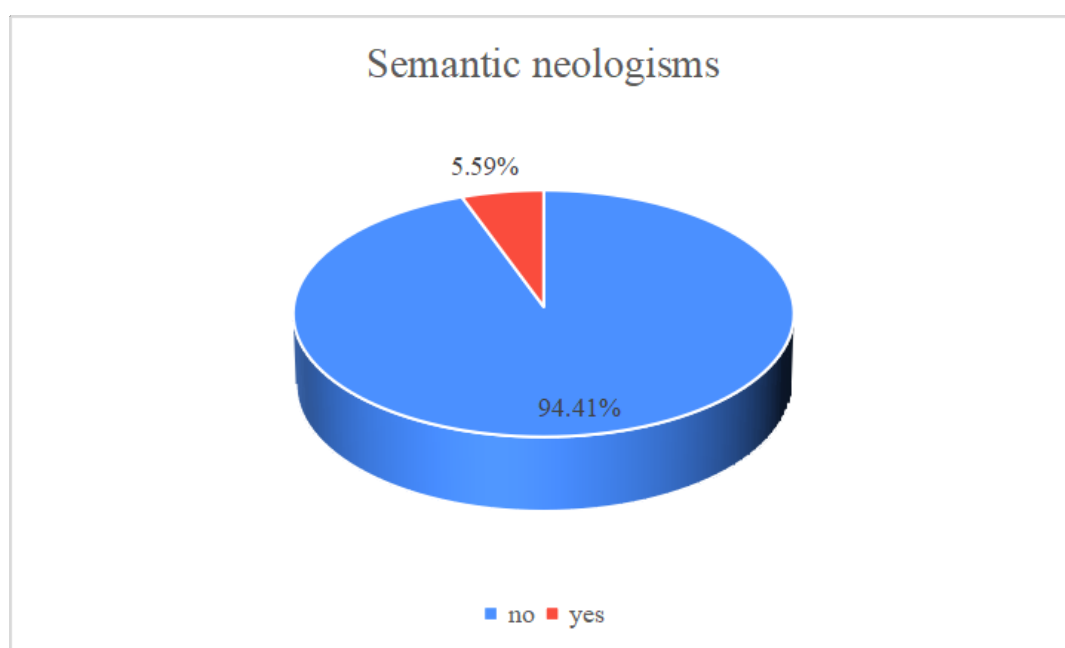


Figure 25: Distribution of the semantic neologisms' productivity in Spanish data.

Cabré (2006: 233) indicates that semantic neologisms are formed by modifying the meaning of a lexical base. Examples include *buscador* (informatics), *parquet* (basketball court), and *llegador* (football player).

In general, a high frequency of use is widely recognized as a fundamental criterion for the inclusion of a neologism in language dictionaries (Barnhart, 1985; Ishikawa, 2006; Cook, 2010; O'Donovan & O'Neill, 2008). When discussing semantic neologisms, an important aspect to consider is whether the newly assigned meaning to an existing word has been integrated into the dictionary. Within our dataset, we have identified a noteworthy case of word duplication: *goya* (item 33/item 215). This finding suggests that *goya* may exhibit a higher frequency of usage compared to other words in our dataset, making it an ideal choice for our subsequent example.

Number	Entry	Ye	Morpheme classific	Semantic neologism	Borrow	Syntactic neologism	Context
12	lirico	2005	/	yes	no	no	Hace tiempo que las autoridades de este teatro este
33	goya	2006	/	yes	no	no	El festival rendirá homenaje al guionista Rafael Az
95	quinceañera	2008	/	yes	no	no	Por segundo año consecutivo, el próximo sábado 2
105	aquejarre	2009	/	yes	no	no	El *aquejarre* de adictos a la novela negra celebra
106	bochorno	2009	/	yes	no	no	El conocimiento sobre el modo en que se efectuar
108	dilapidar	2009	/	yes	no	no	Juan Muñoz es uno de los referentes básicos en la
113	interno -na	2009	/	yes	no	no	La convocatoria de alumno *interno* se creó para
121	quintil	2009	/	yes	no	no	La asociación también detectó fichas de protecció
134	diésel	2010	/	yes	no	no	Antes de empezar el acto, le alemán
144	nicovita	2010	/	yes	no	no	Era el año 1979, almorzábamos en casa con mamá
157	armada	2011	/	yes	no	no	Nueve podios cosechó la *armada* de Concepción
171	ordeñador	2011	/	yes	no	no	Información de la Policía Federal Preventiva prop
188	fondero	2012	/	yes	no	no	Completa el grupo Spicer Boy, *fondero* que puec
215	goya	2013	/	yes	no	no	Y el Premio de interpretación en Cannes y dos *gc
248	nicho	2014	/	yes	no	no	Esos son los cinco grandes *nichos* que comprat
263	comunero	2015	/	yes	no	no	37 días llevan en huelga de hambre cuatro *comun

Figure 26: A case of word duplication: *goya* (item 33/item 215) in Spanish semantic neologisms data.

## 1 goya

**Contexto:** El festival rendirá homenaje al guionista Rafael Azcona, que tras obtener seis \*goyas\* ha accedido a subir a un escenario por primera vez a recoger un premio.

**Categoría gramatical:** Nombre masculino (m)

**Tipo de neologismo:** Neologismo semántico (S)

**Aspectos tipográficos:** Sin marca tipográfica (smt)

**Fuente:** El País (EP)

**Corpus de vaciado:** Corpus escrito de prensa (pr)

Figure 27: The entry of *goya* (item 33) registered in *Banco de Cervantes* in 2006.

## 1 goya

---

**Contexto:** Y el Premio de interpretación en Cannes y dos \*goyas\* y otro de Honor... Todo eso se apagó ayer en Madrid a los 80 años.

**Categoría gramatical:** Nombre masculino (m)

**Tipo de neologismo:** Neologismo semántico (S)

**Aspectos tipográficos:** Cursiva (cva)

**Fuente:** El País (EP)

**Corpus de vaciado:** Corpus escrito de prensa (pr)

Figure 28: The entry of *goya* (item 215) registered in *Banco de Cervantes* in 2013.

In both contexts, the word *goya* refers to the Goya Awards, which are Spain's main annual national film awards. To be precise, the Academy of Cinematographic Arts and Sciences of Spain celebrates the excellence of Spanish cinema each year through the presentation of the Goya Awards to the outstanding professionals in each of the technical and creative fields. We have consistently observed the emergence of the same novel meaning in both the 2006 and 2013 datasets, suggesting a potential sustained adoption of this new word meaning. However, our search in the *Diccionario de la lengua española* did not yield the identified semantic extension of *goya*.

Bernal, Freixa and Torner (2020) suggest that evaluating the frequency usage of neologisms is a complex criterion that further requires considering its temporal stability and distribution across different types of texts, registers, and geographical varieties. Therefore, based solely on the frequency of two instances within our database, we are unable to ascertain the extent of widespread usage of the term *goya*. However, it is plausible to infer that the newly assigned meaning of this term has the potential to gain broader recognition in the future, given the widespread popularity of the Goya Awards in Spain and beyond. Consequently, it could be considered a prospective candidate for inclusion in the dictionary in the future.

### 5.3 Lexicalization of phrases

According to Cabré (2006: 233), the process of syntagmation refers to the formation of neologisms through lexicalized syntactic structures, which is in line with the characteristics of lexicalization of phrases and is therefore further classified in depth as a phrase by our effort.

The category of phrase expressions demonstrates a moderate presence within the entire dataset, comprising 9.79% (28 items) of the total data. One of the notable structural patterns observed within this group of phrases is the combination of *noun + de + noun*, where the preposition 'de' serves to indicate modification and establishes a link between two nouns. This pattern is exemplified by phrases such as *bolsón de pobreza* (item 2), *nivel de Samadhi* (item 14), *quijada de burro* (item 43), *queso de cabeza* (item 277), and *tráfico de especies* (item 280). Another observation is the pattern of *adjective + y + adjective*, in which these two adjectives share a similar meaning and are combined through the conjunction 'y', such as the example *común y corriente* (item 237). In this case, both *común* and *corriente* can mean 'common' or 'ordinary', and when used together in this phrase, they reinforce the idea of something being unremarkable or not standing out.

In addition, we found relatively high frequencies of phrases related to the word *zona*, such as *zona cero* (item 26, item 78), *zona núcleo* (item 52), *zona no euro* (item 234), and *zona roja* (item 260).

## 5.4 Borrowing

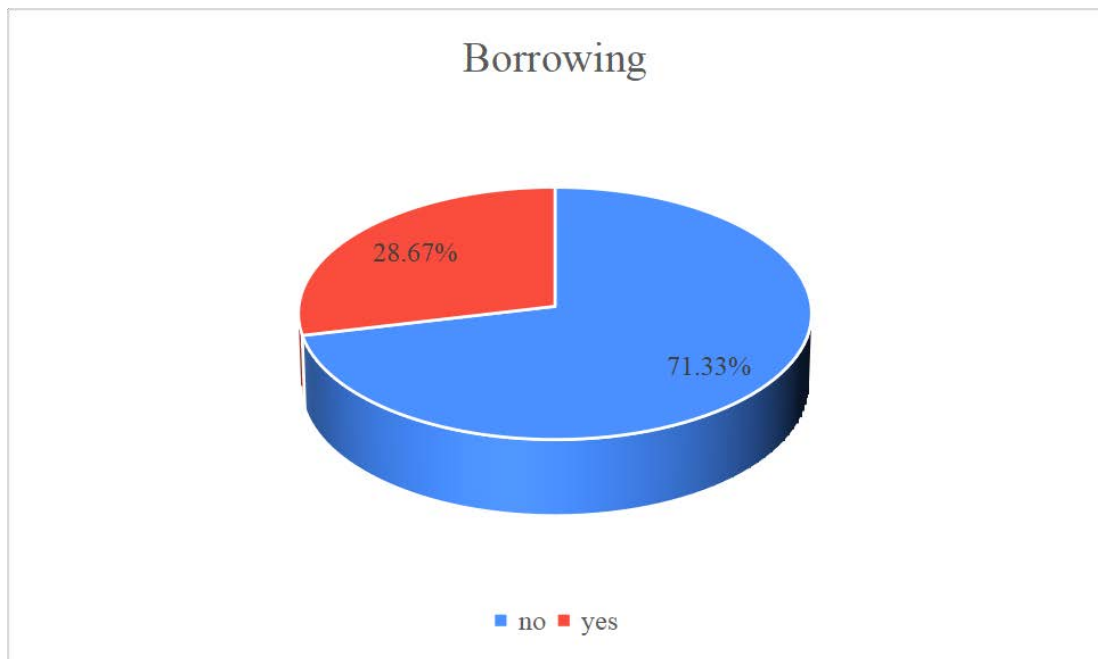


Figure 29: Distribution of the borrowing words' productivity in Spanish data.

Borrowed words represented a significant portion, accounting for 28.67% (82 items) of the entire database. This percentage surpassed that of any other category of formal neologisms, underscoring the crucial role of borrowing in fostering dynamic neologic innovation within Spanish.



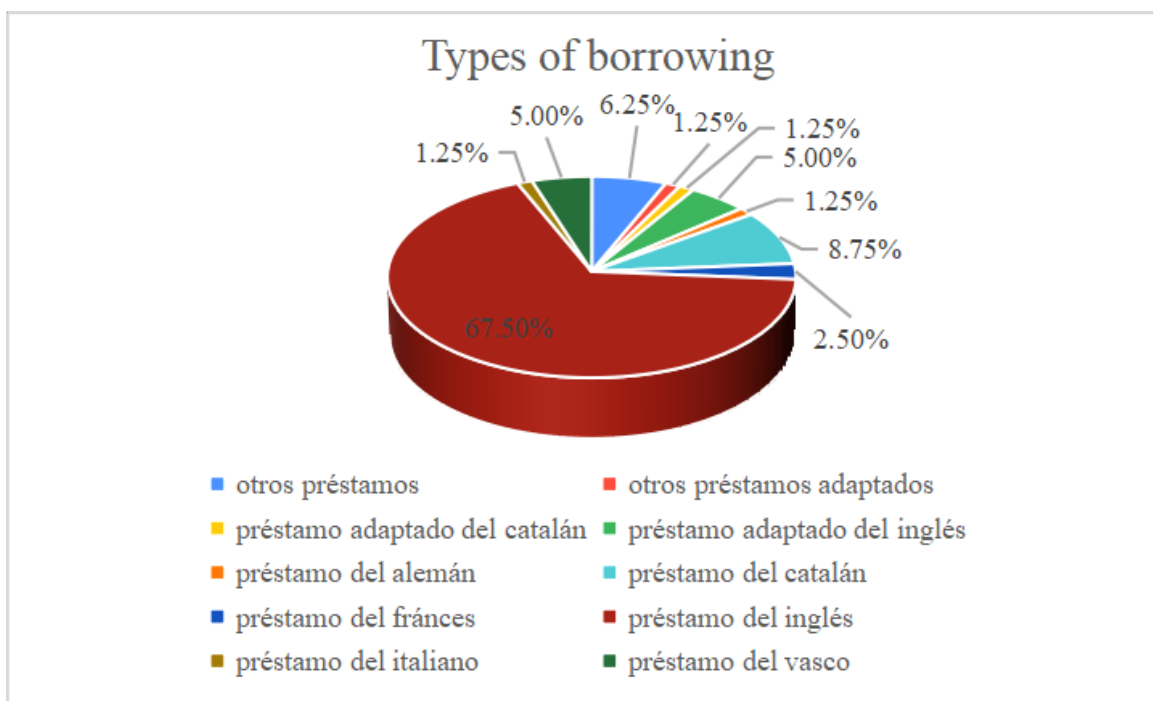


Figure 30: Distribution of different types of borrowing words in Spanish data.

When it comes to the specific types of borrowing, we discovered that 67.50% of the vocabulary was borrowed from English, with an additional 5.00% being loanwords adapted from English. In other words, a significant 72.5% of the borrowed words in the dataset can be attributed to direct or indirect borrowing from English, highlighting the considerable influence of English on the introduction of new terms in Spanish. To provide some examples, we found direct borrowings such as *toy* (item 20), *winner* (item 23), *hit* (item 34), *spa* (item 45), *wiki* (item 101), *wifi* (item 205), *diet* (item 238), and *underground* (item 281). Additionally, there are indirect borrowings with orthographical adjustments, such as *esténcil* (item 135) and *power trío* (item 224). These examples further illustrate the diverse ways in which English borrowings have made their way into the Spanish lexicon.

Furthermore, a notable finding is the prevalence of regular compounds in these English loanwords, which exemplifies their productivity. For instance, we observed compounds like *spaghetti-western* (item 19), *wild-card* (item 49), *stand-up* (item 97), *kiss-kiss* (item 141), *night show* (item 40), *new wave* (item 92), *body copy* (item 158), *living space* (item 220), *widespread*

(item 127), and *longboard* (item 194). These types of English words not only contribute to the relatively limited number of compounds in Spanish formal neologisms, but also enrich the language with their lexical diversity and complexity. Additionally, we also encountered instances of phrase expressions, such as *knock down* (item 11) and *digital rights management* (item 56).

There are also a few borrowings from German (1.25% of the loanwords in our dataset, 1 item,). An illustrative example is the term *kneipen* (item 245). Furthermore, we also encountered borrowings from other Romance languages, namely French and Italian, accounting for 2.50% (2 items) and 1.25% (1 item) respectively. These languages, derived from Latin, have contributed to the Spanish vocabulary with words such as *villeggiatura* (item 74) from Italian and *boeuf bourguignon* (item 80) from French.

Catalan and Basque are two additional languages spoken in Spain, and our analysis revealed the presence of neologisms from these languages within our dataset. Notably, Catalan demonstrates a relatively strong influence in the Spanish language, with 8.75% of the new words directly borrowed from Catalan, exemplified by terms like *xota* (item 180) and *estelada* (item 139), as well as 1.25% of indirectly borrowed words through formal adjustments, such as *xorisar* (item 128). Additionally, we identified that 5.00% of the loanword group comprises new words borrowed from the Basque language, including *ararteko* (item 79) and *zizurtarra* (item 156).

We also identified 6.25% (5 items) of the words classified as direct borrowings from other languages and 1.25% (1 item) as other indirect borrowings. While the *Banco de Cervantes* database does not provide specific language sources for these words, we looked them up in various dictionaries to ascertain their origins. For instance, the term *ninja* (item 170) is listed

in the Merriam-Webster online dictionary<sup>12</sup>, which clearly attributes its origin to the Japanese language. Similarly, the word *oranje* is identified as Dutch according to the Wiktionary<sup>13</sup>. Another example is *Yunsa*, which, according to the Spanish version of Wikipedia<sup>14</sup>, is derived from Quechua.

In general, these multiple instances of borrowing from multiple linguistic sources highlight the cross-pollination and linguistic exchange between languages, further enriching the vocabulary and cultural diversity within the Spanish language.

### 5.5 Other characteristics: syntactic neologisms

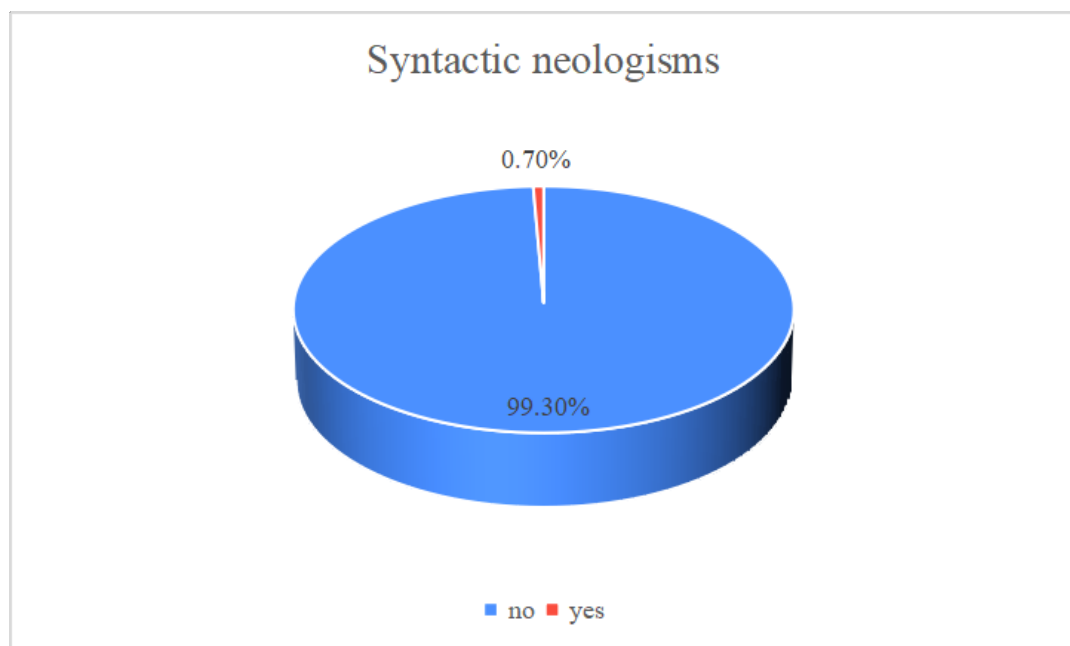


Figure 31: Distribution of syntactic neologisms' productivity in Spanish data.

According to Cabré (2006: 233), syntactic neologisms involving a change in grammatical subcategory (e.g., gender, number, change in verbal regime, etc.) within a lexical base, such as *descalificarse*, which undergoes a *verb-to-pronoun* transformation from the transitive verb *descalificar*.

---

<sup>12</sup> The URL for checking is: <https://www.merriam-webster.com/dictionary/ninja>.

<sup>13</sup> The URL for checking is <https://en.wiktionary.org/wiki/oranje>.

<sup>14</sup> The URL for checking is <https://es.wikipedia.org/wiki/Yunza>.

This particular category of neologisms is relatively infrequent within our comprehensive database, comprising only 0.70% (2 items) of the total word count. However, the instances we have identified both exhibit a distinctive transformational pattern of *verb-to-pronoun* conversion. Notable examples include *universalizarse* (item 177), derived from *universalisar*, and *internacionalizarse* (item 269), derived from *internacionalizar*. Additionally, it is worth noting that the presence of the reflexive pronoun "se" in these instances indicates the reflexive nature of the verb and implies the action being performed on oneself.

## **6. Comparative Analysis of English and Spanish Neologisms in Two Projects**

### **6.1 Formal neologisms**

Differences exist in word formation processes between English and Spanish. Our analysis revealed that compounds are the most prevalent word formation process in English, accounting for 25.08% of the entire English dataset. Conversely, in Spanish, the most representative process, excluding the "/" type that defies classification, is suffixation, which constitutes 19.23% of our complete Spanish data.

The prevalence of compounds in English neologisms potentially validates the dynamic potential of combining roots and combining forms. However, it also highlights the aforementioned challenge of distinguishing between compounds and phrases, as well as between compounds and neoclassical compounds or blends, or even between compounds and derivations. This difficulty may contribute to the productivity of compound words observed in our English data, and we still suggest the need for a revised definition of compounds in the future.

However, in contrast to English, compounds exhibit a relatively limited presence in Spanish, accounting for only 3.85% of the entire Spanish language dataset. This disparity highlights a difference in the compound's productivity of word formation between English and

Spanish. Despite this divergence, both languages share a common characteristic in terms of their intrinsic structure, namely the prominent use of nouns as the primary word class within their respective compound vocabularies. Notably, a significant and noteworthy intrinsic structure observed in compounds in both languages is the combination of two nouns.

That suffixation emerges as the most productive word formation process in Spanish is not entirely surprising given the role of derivation in the language, yet we might have expected prefixation (especially with forms like *super-* and *hiper-*) to be just as productive. Our database exhibits a rich and diverse array of suffixes employed in Spanish neologisms. Additionally, it is noteworthy that both English and Spanish demonstrate a higher productivity of suffixation compared to prefixation. This discrepancy is particularly pronounced in Spanish, with the percentage of suffixation surpassing that of prefixation by more than two-fold.

Furthermore, our data indicate a significant prevalence of neoclassical compounds in Spanish data, which account for 13.99% of the total. This discrepancy can be attributed to the inherent nature of Spanish as a Romance language, which draws from Latin and Greek origins for word formation. Cabré (2006) proposed a classification of neoclassical compounds in Spanish, encompassing three types: (1) neoclassical prefix + neoclassical suffix, (2) neoclassical prefix + root, and (3) root + neoclassical suffix. Interestingly, Cabré's (2006) classification appears to be more expansive and inclusive than our standard for English, allowing for greater flexibility in considering a new word as a neoclassical compound. This suggests that distinguishing between compounds and neoclassical compounds in Spanish may be comparatively more challenging.

In English, we employ a condition that both combining forms must originate from Latin or Greek in order to classify a word as a neoclassical compound. Although this condition has resulted in a scarcity of neoclassical compound data in our overall analysis, we adhere to it due to the relative imperfection of the definition of compounds in English. Moreover, English

exhibits the word formation process of blending, further complicating the differentiation between these three types. To be precise, blending constitutes 14.37% of our comprehensive English data and showcases the diverse combinations that can arise through this process, which is not observed in Spanish.

In our analysis of both English and Spanish, it becomes evident that acronyms and abbreviations play roles as word formation processes. However, we must acknowledge that despite sharing the same terminology, these processes exhibit distinct connotations and characteristics in each language. On the one hand, Cabré (2006: 233) defines abbreviations in Spanish as neologisms formed through the abbreviation of a lexical base, exemplified by terms like *protá* (*proagonista*) and *neocon* (*neoconservador*). On the other hand, acronyms in Spanish are created by combining segments of words to form a syntactic structure, as seen in examples provided by Cabré (2006: 233) like *cubanglish* (*cubano y English*) and *turismática* (*informática aplicada al turismo*).

In contrast, Bauer and Huddleston (2002: 1632-1633) differentiate between abbreviations and acronyms in English based on pronunciation and spelling conventions, as discussed in section 4.1.8. Therefore, specifically, in Spanish, the concept of abbreviation aligns more closely with the word formation process of clipping observed in English. Conversely, the Spanish examples of acronyms provided by Cabré (2006) demonstrate potential similarities to the process of blending in English. Consequently, these disparities in classification and criteria for neologisms contribute to variations in the final outcomes of our analysis.

Last but not least, the process of conversion can be observed in both English and Spanish, revealing certain common transformation patterns such as *adjective-to-noun* and *noun-to-adjective* conversions. However, according to our data analysis, English may exhibit a wider range of potential patterns, including *noun-to-verb* and *verb-to-noun* conversions,

among others. Furthermore, we propose conducting further investigation to explore the potential role of the structural mechanism of conversion as it may serve as a bridge connecting formal neologisms to semantic neologisms. Meanwhile, it is worth noting that the challenge of identifying additional patterns in Spanish is closely linked to the nature of the Spanish neologism project itself. While the *Banco de Cervantes* provides a systematic collection of neologisms, it does not explicitly indicate the original word class of each neologism. Moreover, the rarity of these neologisms and the limited access to them in dictionaries further complicate the distinction of their specific conversion patterns.

## **6.2 Semantic neologisms**

The proportion of semantic neologisms in relation to the total data is relatively modest in both English and Spanish, accounting for 5.81% and 5.59%, respectively, with a marginal difference between the two languages. Nonetheless, a shared challenge across both languages lies in the identification of semantic neologisms.

When dealing with English neologisms in our analysis, there is often a certain degree of ambiguity in determining whether a new word represents a regular neologism with a novel meaning or an additional meaning attached to an existing word. It would be advantageous if the editors of *Among the New Words* consistently included indications of the previous meanings of existing words in their journal entries. This practice would help us reduce the possibility of inadvertently overlooking certain semantic neologisms. However, it is crucial to acknowledge the advantages of the English neologism project, as they already provide meanings for each new word and offer example sentences. Unfortunately, this comprehensive information is lacking in the Spanish neologism project. Within the *Banco de Cervantes* system, semantic neologisms are accompanied by contextual information only, without specific definitions. Consequently, further analysis is required to ascertain the precise neologism associated with

the old word based on the given context in our study. These aspects present additional challenges to our ongoing analysis.

Lombard, Huyghe and Gygax (2021) indicate that neological intuition (NI) refers to the metalinguistic capacity to assess the novelty of words. It pertains to how individuals comprehend and process unfamiliar words encountered in daily experiences, even in the absence of extensive linguistic knowledge. While NI relies on overall linguistic proficiency, they have revealed that it varies across individuals due to a combination of linguistic and non-linguistic factors. Moreover, they put forth the notion that response times<sup>15</sup> provide evidence indicating that the identification of semantic neologisms poses greater difficulty compared to morphological ones. This argument can be more or less confirmed once again in our study, thereby further substantiating its validity.

### 6.3 Borrowing

The presence of loanwords in the Spanish neology data differs greatly from that in the English data. We have seen (see section 4.4) that loanwords have a relatively limited presence in English. In contrast, Spanish exhibits a significantly higher proportion of loanwords at 28.67%, surpassing even the formal neologism categories. This substantial disparity in loanword productivity between the two languages suggests that English tends to disseminate its linguistic features and expressions globally. In contrast, Spanish assumes a more receptive and absorptive role, actively incorporating influences from diverse countries and language families.

Based on our analysis of the Spanish database, the predominant proportion of loanword types comprises newly introduced words originating directly from English or indirectly through

---

<sup>15</sup> In their study (Lombard, Huyghe, & Gygax, 2021), response time refers to the time it takes for a person to identify a neologism.



orthographic modifications. This phenomenon serves as a microcosm, highlighting the expression of English cultures and the pervasive influence of English across various domains in other language cultures, particularly within the context of globalization. The differential absorption of loanwords between English and Spanish can be attributed to inherent structural and lexical disparities.

English exhibits a more flexible and adaptable structure, facilitating the relatively seamless incorporation of words from other languages. In contrast, Spanish, with its phonetic and grammatical rules, inflection, and greater reliance on derivational processes, encounters inherent challenges when exporting words into other languages. Nonetheless, Spanish demonstrates a more inclusive approach to assimilating loanwords by directly introducing new words or adapting foreign loanwords to align with its own cultural and linguistic features.

Furthermore, our database highlights the diversity of foreign cultural influences present in Spanish. This diversity extends beyond English to encompass loanwords from German and Dutch, as well as loanwords from French and Italian, representative of the Romance language family. Additionally, loanwords from other languages spoken within Spain, such as Catalan and Basque, contribute to the rich linguistic landscape. Moreover, our findings reveal the representation of relatively minor languages worldwide, such as Quechua, through the addition of new words in Spanish.

In contrast, loanwords in English have a relatively limited presence. However, a notable characteristic in our database is the relatively significant influence of Asian cultures on English, evident in the incorporation of loanwords from Chinese, Japanese, and even Korean. It is worth mentioning that we also discovered Japanese loanwords in Spanish, suggesting a potential gradual increase in the influence of Asian culture on both English and Spanish in the future.

In conclusion, the contrasting patterns of loanword utilization in English and Spanish highlight the distinct roles these languages play within the global linguistic landscape. English

operates as a vehicle for spreading its linguistic features and contributing to the linguistic diversity of other cultures, while Spanish embraces external influences to enrich its lexicon and cultural fabric.

#### **6.4 Lexicalization of phrases**

There are notable differences in the lexicalization of phrases between English and Spanish. In the English dataset, phrases exhibit considerable productivity, ranking second after compound words and accounting for 24.16% of the total data. Conversely, the proportion of phrases in Spanish is relatively moderate, amounting to only 9.79%. This contrast can be attributed to the distinct nature and objectives of the two neologism projects.

It is crucial to emphasize that the *Banco de Cervantes* project primarily concentrates on single-word entries rather than phrases. This deliberate focus aligns with the project's objective of expanding the dictionary. Therefore, the primary criterion for inclusion in the *Banco de Cervantes* revolves around determining whether a word merits a place in the dictionary. In contrast, *Among the New Words* operates differently, functioning as a collaborative effort driven by individuals with a passion for identifying noteworthy linguistic phenomena. While the *Banco de Cervantes* project provides a more comprehensive view to some extent and relies on a systematic classification to categorize different types of new words, *Among the New Words* does not adhere to a specific criterion or prioritize dictionary inclusion. Contributors to *Among the New Words* are motivated by curiosity, finding certain words or expressions interesting, regardless of their potential inclusion in a dictionary.

Essentially, the contrasting data highlights that each project offers a partial perspective on neology in the respective languages. However, their viewpoints diverge significantly. From the standpoint of *Among the New Words*, a new word is deemed interesting and can serve as a microcosm of a reflexive social phenomenon or a combination of new word forms, without necessarily needing to be included in a dictionary. Conversely, the *Banco de Cervantes* holds

the belief that the words included in its system have the potential to be future candidates for dictionary inclusion.

Furthermore, it is important to acknowledge that despite the divergences discussed earlier, there are noteworthy similarities in phrase structures between English and Spanish. As expounded upon in Section 4.3 and Section 5.3, English demonstrates a distinct pattern characterized by "noun and noun" and "verb and verb" constructions. In these constructions, the two constituent words exhibit similar phonological and morphological characteristics. Similarly, Spanish exhibits a pattern of "adjective y adjective", where the conjunction "y" corresponds to the English equivalent of "and". Furthermore, it is possible for the two adjectives to share similar meanings. These observations suggest that both languages share a common structure for creating new phrases, wherein two words of the same word class are combined, potentially showcasing similarities in phonological, morphological, and semantic features. Moreover, Spanish also features the intriguing "noun *de* noun" construction, which translates to "noun of noun" in English.

Considering these observations, it becomes evident that despite the previously discussed variations in productivity, English and Spanish do exhibit certain commonalities in their phrase structures. These shared patterns imply that certain aspects of language organization transcend individual languages and can be identified across diverse linguistic systems.

### **6.5 Other characteristics**

In both English and Spanish, there are intriguing and distinct characteristics observable in the realm of new words. Notably, in English, we have identified the presence of analogical processes that influence both the form and meaning of original words, resulting in the creation of a new lexicon. However, such analogical phenomena have not been observed in our Spanish

dataset. Conversely, *Banco de Cervantes* has classified a specific category of neologisms called syntactic neologisms, which remains absent from our English database.

However, despite these differences, a noteworthy similarity between the two languages emerges concerning the significance of word elements in the formation of neologisms. In English, our data reveals that certain affixes participate in various word formation processes, and some items can be categorized explicitly as affixes or combining forms. Conversely, while our Spanish data lacks independent word elements, our analysis highlights a wide range of affixes employed in either suffixation or prefixation processes. This underscores the crucial role played by such word elements in the study and analysis of neologisms.

## **7. Conclusions**

### **7.1 Observations about the word formation processes in these two languages projects**

In general, we observed differences in the productivity of word formation processes between English and Spanish. English neologisms are primarily found in compounds and phrases, whereas Spanish exhibits significant productivity in suffixation and neoclassical compounds. Additionally, other types of neologisms, such as semantic neologisms and borrowings, cannot be further classified based on the word formation process according to the qualification criteria in *Banco de Cervantes*. However, borrowed words in Spanish demonstrate a prominent presence compared to other categories of formal neologisms.

We suggest that these differences can be explained by three factors. Firstly, the nature and purpose of the two neologism projects differ, leading to distinct detection of word formation processes in the two languages. The *Banco de Cervantes*, operated by the *Observatori de Neologia* project (OBNEO), primarily aims to provide new words for dictionaries, focusing on word selection rather than phrases. Conversely, the approach in

*Among the New Words* reflects a more exploratory mindset, prioritizing interesting linguistic observations over dictionary inclusion.

Secondly, the inherent nature of the languages themselves plays a crucial role. English exhibits a remarkable level of flexibility and adaptability in its structure, allowing for seamless exportation of new words to other languages, particularly during the process of globalization and due to its worldwide influence. In contrast, Spanish faces inherent challenges in exporting words due to its phonetic and grammatical rules and stronger reliance on derivational processes. However, Spanish adopts a more inclusive approach by assimilating loanwords, either through direct introduction or adaptation to align with its cultural and linguistic features. Consequently, we propose that English functions primarily as an exporting language, disseminating its influence through borrowed words, while Spanish tends to operate as an importing language, assimilating loanwords to enrich its vocabulary. This inherent difference in language nature significantly contributes to the contrasting borrowing tendencies observed between English and Spanish.

Thirdly, the historical origins of the two languages can also be a factor. English belongs to the Germanic language family, whereas Spanish traces its roots back to Latin. This disparity in origin results in Spanish exhibiting a greater presence of newly affixed words in our data compared to English. Therefore, the disparities of the different word formation process productivity, not only highlight the distinctive nature of each language but also demonstrate how linguistic traits and historical backgrounds contribute to the evolution and enrichment of new words in different linguistic systems, for example, blends actually are characteristics of English but not Spanish.

Furthermore, it becomes evident that while there is a superficial similarity in the word formation processes observed in the neologism data of English and Spanish, this similarity may not hold true in essence. It is crucial to acknowledge that the apparent resemblance in word

formation processes between the two languages may arise from divergent criteria for classification and differing interpretations of terminology.

On the one hand, in our analysis of English neologisms, we employed Bauer and Huddleston's (2002) classification of English word formation as one of our analytical frameworks. Conversely, for the analysis of Spanish data, we adopted the theoretical framework proposed by Cabré (2006) within the *Observatori de Neologia* (OBNEO). Consequently, certain terminologies used in both classifications may bear similar names, but their interpretations and underlying principles may differ significantly. For example, Cabré's (2006) concept of abbreviation aligns more closely with Bauer and Huddleston's (2002) definition of clipping.

The definition of neoclassical compounds provided by Cabré (2006) is notably broader, encompassing any word element with neoclassical features. In contrast, our analysis of English neoclassical compounds adheres to a stricter criterion, requiring both combining forms to originate from Greek or Latin sources. This additional criterion introduces a more restricted perspective, resulting in a limited amount of data found for English neoclassical compounds. Moreover, we would like to highlight the fact that neoclassical elements in English, as in Spanish, combine with a wide range of morphological types and not only with other neoclassical formants. This fact points up the need for increased study of neoclassical formants in current English word-formation.

## **7.2 Reflections about the difficulties during the process of our study**

Throughout the process of our study, we have encountered significant challenges in classifying the word formation processes within the English data obtained from *Among the New Words*. Unlike *Banco de Cervantes*, which employs a systematic classification approach, *Among the New Words* lacks such a framework. Therefore, in order to compare these two projects, we have undertaken the task of classification ourselves. As previously mentioned,

different authors utilize varying groupings for classification, and we have selected what we deem to be the most authoritative criteria as the foundation of our framework. However, we have still faced notable difficulties during the classification process.

In general, we believe that the encountered challenges revolve around the definition of compounds. This encompasses issues such as distinguishing compounds from phrases, differentiating compounds from blends, delineating compounds from derivations, and even clarifying the distinction between affixes and combining forms, which is not very clear. The ambiguity arises from the influence of diachronic development in word formation. Although the definition of compound words has been a subject of debate among scholars, with various definitions proposed and revised over time in response to emerging complexities (e.g., Bauer, 1983; Plag, 2003), we suggest that further refinement of the compound definition may be necessary in the future, given the persisting challenges in word formation classification.

Despite our adherence to *Banco de Cervantes'* classification of word formation processes in their system, we have encountered instances where the actual application of these processes does not fully align with the theoretical framework. When we scrutinize the data and review the definitions and examples provided by Cabré (2006), discrepancies become apparent. For instance, according to Cabré's (2006) framework, the term *fotoperiodismo* clearly falls into the category of neoclassical compound. However, *Banco de Cervantes* classifies it as a compound.

Therefore, we suggest that it may be necessary to conduct a comprehensive re-evaluation and revision of the word formation process information for each newly registered word in *Banco de Cervantes* to ensure its factual accuracy. By reassessing the classification of neologisms and comparing them with the established theoretical framework proposed by Cabré (2006), we can ensure consistency and accuracy in the classification of neologisms. This will

contribute to the overall reliability and integrity of the word formation process data in *Banco de Cervantes*.

Our analysis has revealed additional challenges when examining semantic neologisms in both English and Spanish. Identifying semantic neologisms proves to be more difficult compared to morphological neologisms, primarily due to variations in neological intuition (Lombard, Huyghe, & Gygax, 2021). Specifically, in the case of English semantic neologisms, distinguishing between regular neologisms with new meanings and additional meanings attached to existing words presents a significant challenge. To address this ambiguity, it would be beneficial for readers to further understand better about the semantic shift of the new words if the editors of *Among the New Words* consistently indicated the previous meanings of existing words in their journal entries. This practice would also greatly assist in reducing the likelihood of overlooking certain semantic neologisms during related research like ours. Nonetheless, it is worth noting that the English neologism project already offers meanings and example sentences for each new word, providing valuable resources for analysis.

In contrast, the Spanish neologism project within *Banco de Cervantes* presents a distinct set of challenges. Semantic neologisms in Spanish lack specific definitions and are accompanied solely by contextual information. Consequently, our study requires further analysis to accurately determine the precise neologism associated with an old word, relying on the contextual clues provided. This limitation significantly complicates our understanding and investigation of Spanish neologisms.

To summarize, our research underscores the importance of consistent indications of previous meanings in English semantic neologism entries and highlights the advantages of the English neologism project in providing comprehensive information. Conversely, the Spanish neologism project poses challenges due to the absence of specific definitions, requiring additional analysis to unravel the intended neologism. These observations emphasize the need



for refining classification approaches and delving further into the complexities of word formation in both languages.

## References

- Adelstein, A. (2019). Hacia una definición de neologicidad en semántica. In *Palabras como puentes: estudios lexicológicos, lexicográficos y terminológicos desde el Cono Sur* (pp. 209-230). Buena Vista Editores.
- Akut, K. B. (2020). Morphological Analysis of the Neologisms during the COVID-19 Pandemic. *International Journal of English Language Studies*, 2(3), 01-07.
- Algeo, J., & Algeo, A. S. (1991). *Fifty years among the new words: A dictionary of neologisms 1941-1991*. Cambridge University Press.
- Algeo, J., & Algeo, A. (1997). Among the New Words. *American Speech*, 72(2), 183-197. Duke University Press. <https://www.jstor.org/stable/455789>
- Alvarez-Bolado Sanchez, C. M., & Alvarez de Mon Rego, I. (2013). Semantic Neology in the Domain of Videogames in Spanish. *Ibérica*, 25(25), 61-82.
- Alves, I. M. (2007). The Social Aspect of Brazilian Portuguese Neologisms. *Dictionaries: Journal of the Dictionary Society of North America*, 28(1), 149-152.
- Anderson, S. R. (1992). *A-morphous morphology* (No. 62). Cambridge university press.
- Asif, M., Zhiyong, D., Iram, A., & Nisar, M. (2021). Linguistic analysis of neologism related to coronavirus (COVID-19). *Social Sciences & Humanities Open*, 4(1), 100201.
- Barnhart, D. K. (1985). Prizes and pitfalls of computerized searching for new words for dictionaries. *Dictionaries*, 7, 253-260.
- Bauer, L. (1983). *English word-formation*. Cambridge university press.
- Bauer, L., & Huddleston, R. D. (2002). Lexical word-formation. In R. Huddleston (eds.), *The Cambridge Grammar of the English Language* (pp. 1623-1719). Cambridge University Press.
- Bernal, E., Freixa, J., & Torner, S. (2020). Criterios para la diccionarización de neologismos: de la teoría a la práctica. *Revista signos*, 53(104), 592-618.
- Booij, G. (2010). Construction morphology. *Language and linguistics compass*, 4(7), 543-555.
- Bosque, Ignacio. (1982) Más allá de la lexicalización. *Boletín de la Real Academia Española* 62, 103-158.
- Cabré, M. T. (2006). La clasificación de neologismos: una tarea compleja. *Alfa*, 50(2), 229-250.

- Čolić, A. M. (2015). Word formation of blends. *Mostariensia-časopis za društvene i humanističke znanosti*, 19(2), 21-36.
- Cook, C. P. (2010). *Exploiting linguistic knowledge to infer properties of neologisms* [Doctoral dissertation, University of Toronto]. Toronto.
- Creus, I., & Julià-Muné, J. (2014). Neologismes semantics / Semantic change. *Italian Journal of Linguistics*, 15, 179-205. <https://doi.org/10.1075/ivitra.7.10cre>
- Hormigo, M. T. D. (2022). Neología de préstamo: la adaptación morfológica de género de los neologismos léxicos procedentes del inglés. *Revista de Lingüística y Lenguas Aplicadas*, 17, 49-58.
- Ishikawa, S. (2006). When a word enters the dictionary: A data-based analysis of neologism. In JACET Society of English Lexicography (Ed.), *English Lexicography in Japan* (pp. 39-52). Bunkyo-ku: Taishukan.
- Lehrer, A. (2003). Understanding trendy neologisms. *Italian Journal of Linguistics*, 15, 369-382.
- Lieber, R. (1992). *Deconstructing morphology: Word formation in syntactic theory*. University of Chicago Press.
- Lombard, A., Huyghe, R., & Gygax, P. (2021). Neological intuition in French: A study of formal novelty and lexical regularity as predictors. *Lingua*, 254, 103055. <https://doi.org/10.1016/j.lingua.2021.103055>
- O'Donovan, R. & O'Neill, M. (2008). A systematic approach to the selection of neologisms for inclusion in a large monolingual dictionary. In E. Bernal & J. DeCesaris (Eds.), *Proceedings of the XIII EURALEX International Congress (Barcelona, 15-19 July 2008)* (pp. 571-579). Barcelona: Institut de Lingüística Aplicada – Universitat Pompeu Fabra.
- Onyedum, A. (2012). Social media Neologisms: A morpho-semantic analysis. *Lagos: University of Lagos*.
- Plag, I. (2003). *Word-formation in English*. Cambridge University Press.
- Porras Garzón, J. M. (2016). Una aproximación al uso de la neología especializada en el ámbito de las enfermedades raras. [Master's thesis, Pompeu Fabra University]. UPF Digital Repository. [https://repositori.upf.edu/bitstream/handle/10230/27684/Porras\\_2016.pdf?sequence=1](https://repositori.upf.edu/bitstream/handle/10230/27684/Porras_2016.pdf?sequence=1)
- Pound, L. (1914). *Blends, their relation to English word formation* (Vol. 42). C. Winter.
- Pratama, M. E. S., Idris, Z., Simamora, T. H., & Tampubolon, J. S. (2021). Morphological Process Of Neologism Found In Social Networking Jargon.

*Vernacular: Linguistics, Literature, Communication and Culture Journal*, 1(I), 36-41.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. Longman.

Real Academia Española. (2009). *Nueva gramática de la lengua española. Morfología. Sintaxis I*. Espasa.

Yurtbaşı, M. (2017). Stress placement on phrases and compounds in English. *International Online Journal of Primary Education*, 6(2), 39-46.

Zimmer, B., Carson, C. E., & Horn, L. R. (2011). *American Speech*, 86(3), 355-376. Duke University Press. <https://doi.org/10.1215/00031283-1503937>

Zimmer, B., & Carson, C. E. (2012). *American Speech*, 87(2), 190-207. Duke University Press. <https://doi.org/10.1215/00031283-1668208>

Zimmer, B., & Carson, C. E. (2012). *American Speech*, 87(4), 491-510. Duke University Press. <https://doi.org/10.1215/00031283-2077633>

Zimmer, B., & Solomon, J. (2014). *American Speech*, 89(4), 470-496. Duke University Press. <https://doi.org/10.1215/00031283-2908233>

Zimmer, B., Carson, C. E., & Solomon, J. (2016). *American Speech*, 91(4), 472-512. Duke University Press. <https://doi.org/10.1215/00031283-3870163>

## Appendix

Kindly note that due to the substantial magnitude of the database, it is not viable to incorporate the complete printed Excel sheet within this thesis. However, we offer online access to a document where you can access and review our database.

English neologisms database: Final Data EN\_Yingfei Lu\_Janet DeCesaris

<https://kdocs.cn/l/coIZcHtfTKBb>

Spanish neologisms database: Final Data ES\_Yingfei Lu\_Janet DeCesaris

<https://kdocs.cn/l/cj5yvfqjFpj8>