

# Innovative lexicography: new products and methods in the generative AI age

PEDRO A. FUERTES-OLIVERA

International Centre for Lexicography,  
University of Valladolid (Spain)

pedroantonio.fuertes@uva.es

<https://orcid.org/0000-0003-3831-5377>

SVEN TARP

Universitat Rovira i Virgili, Tarragona  
(Spain) and Aarhus University (Denmark)

st@cc.au.dk

<https://orcid.org/0000-0003-1941-9082>

## Abstract

This paper revolves around the interaction of lexicography with generative AI. It defends that this interaction may be very positive for our independent science if we follow three basic principles of this interaction. Firstly, lexicographers must be always on a “learning mode”, i.e. ready to learn how they can extract adequate lexicographic data with generative AI chatbots, e.g. by learning to make adequate prompts with the chatbots. Secondly, lexicographers must always have the last word, they must validate the data AI chatbots generate. Finally, lexicographers must work in cooperation with AI developers, e.g. devising special types of parallel corpora that can be used for training language models that will offer short and long explanations in the form of glosses of, say, grammar rules, meaning explanations, and so on. We illustrate how this interaction currently works in two projects that offer a way forward for the future of the theory and practice of lexicography: the *Diccionario Digital del Español (DIDES)*, and a generative AI model that is being created for producing glosses that will be working in a writing assistant that aims at meeting some of the language needs of Spanish learners, e.g. explaining some of the rules that may have been violated.

## Keywords

Generative AI; databases; online dictionaries; writing assistants; digital glosses.

# Lexicografía innovadora: nuevos productos y métodos en la era de la IA generativa

## Resumen

Este trabajo analiza la interacción que puede existir entre la lexicografía y la Inteligencia Artificial (IA) generativa. Creemos que esta relación puede ser muy positiva aceptando tres principios básicos. En primer lugar, los lexicógrafos deben estar dispuestos a “someterse” a un proceso de “aprendizaje continuo” gracias al cual puedan ir “aprendiendo” a crear prompts y “adaptarlos” a las diversas situaciones de uso de la lexicografía, por ejemplo, para crear datos lexicográficos que puedan usarse para construir diccionarios, glosas para asistentes de escritura, etc. En segundo lugar, la relevancia de la tecnología nunca nos debe confundir y nunca debemos olvidar que los lexicógrafos deben tener la “última palabra”, es decir, deben validar los datos generados por los chatbots. Finalmente, los lexicógrafos deben centrarse en cooperar con programadores y expertos en IA, por ejemplo, para ir guiándolos para que desarrollen programas específicos que puedan funcionar con corpus paralelos necesarios para entrenar los modelos lingüísticos. Además, en este trabajo ilustramos este tipo de relación entre lexicografía e Inteligencia Artificial generativa con referencias a dos proyectos concretos: el *Diccionario Digital del Español (DIDES)*, y un programa de asistente de escritura para aprendices de español que ofrece glosas creadas con IA generativa que explican de forma breve y sencilla las reglas gramaticales conculcadas en un momento concreto facilitando de esta forma el proceso de aprendizaje de la lengua.

## Palabras clave

IA generativa; bases de datos; diccionarios en línea; asistentes de escritura; glosas digitales.

Recibido el 31/03/2025

Aceptado el 23/06/2025

## How to cite/Cómo citar

Fuertes-Olivera, P., Tarp, S. (2025). Innovative lexicography: new products and methods in the generative AI age. *Revista Internacional De Lenguas Extranjeras / International Journal of Foreign Languages*, (24), 71-95. <https://doi.org/10.17345/rile24.4164>

## 1. Introduction

In recent years, lexicography, like a large and growing number of disciplines, has had to come to terms with artificial intelligence in one form or another. This process has taken a quantum leap forward since the introduction of ChatGPT in November 2022, followed by a number of other generative AI-based chatbots such as Claude, DeepSeek, Gwen, Grok, Gemini, Mistral and Manus, each with its own specific characteristics but undoubtedly useful for various tasks related to lexicography. Not surprisingly, this has caused some turbulence, characterised by a mixture of wonder, curiosity, over-optimism, anxiety and fear for the future of the discipline, as discussed by Gouws & Tarp (2024). Many lexicographers have begun to experiment with the new technology, approaching it in different ways and with different visions of the future, as expressed in a growing number of academic publications, among them Alonso-Ramos (2023), Arias-Arias et al. (2024), De Schryver (2023), De Schryver & Joffe (2023), Fuertes-Olivera (2024, 2025), Huete-García & Tarp (2024), Jakubíček & Rundell (2023), Lew (2023, 2024), Li et al. (2024), McKean & Fitzgerald (2024), Ptasznik et al. (2024), Rees & Lew (2024), Rundell (2023) and Tarp & Nomdedeu-Rull (2024).

In the midst of this necessary and encouraging cauldron of new, forward-looking activity, one can also observe a certain ahistorical conservatism in which the roots and core premises of lexicography seem to have been forgotten, or at least relegated to the background. This is particularly evident in an excessive focus on dictionaries, as if these were the very core of lexicography. Almost fifty years ago, Al-Kasimi (1977:1) astutely observed that “the major motives behind the rise of lexicography differ from one culture to another” and that “each culture fosters the development of dictionaries appropriate to its characteristic demands”. Al-Kasimi’s statement can be generalised to apply not only to the emergence but also to the subsequent development of lexicography, which, from this point of view, has been driven by four main factors: 1) the information needs observed in society, 2) the technology available to meet these needs, 3) the experience and increasing skill of its practitioners, and 4) the growing superstructure of reflections and theoretical statements. These have been the core premises for the development of the discipline over the centuries.

At the same time, however, Al-Kasimi’s statement quoted above is somewhat biased in that he only mentions one of lexicography’s expressions, the dictionary, which, despite the important role it has played and continues to play in human life and development, is not the only form in which the discipline has expressed itself historically. It is now well documented that in many cultures, dictionaries evolved from a tradition of inserting glosses into manuscript works; cf. McArthur (1986), Stathi (2006), Hanks (2013) and Benati & Händl (2019). Examples of this include the famous eleventh-century Silos Glosses, found in the monastery of Santo Domingo de Silos in the Spanish province of Burgos and described in detail by Ruiz Asencio et al. (2020). In this perspective, Tarp & Gouws (2019, 2023) observe that glosses have survived the test of time and are experiencing a renaissance in the digital age, and on this basis redefine lexicography as a discipline consisting of two main branches, glossography and dictionography. All this

leads to the fundamental conclusion that the new AI technology implies not only new ways of making dictionaries, but also completely new products, such as AI-based writing assistants that contain lexicographic data in the form of digital glosses.

From what has been published so far, it can be concluded that lexicographers are approaching the generative AI technology in different ways, with some primarily asking what this technology can do for them and their dictionaries, and others focusing on the interaction between themselves and artificial intelligence to determine what is required of the human lexicographer to make generative AI perform at its best. The different approaches can be seen, for example, by comparing Ptasznik et al. (2024) with Li et al. (2024), where the latter call for “a symbiosis” of human and artificial intelligence, with the human having “the last word” based on a high level of competence and knowledge of the discipline. This production process is in some respects similar to the already existing post-editing of so-called “semi-automatic dictionaries”, which also requires a sound lexicographic craftsmanship, as discussed by Baisa et al. (2019), Blahuš et al. (2023) and other scholars who have used “corpus query tools” and other NLP techniques to compile their dictionaries. However, it also differs in that the outcome is highly dependent on the lexicographers’ ability to write good prompts, a completely new technique that can only be learned through practice and dedicated interaction with the generative AI tools, as shown with numerous examples by Tarp & Nomdedeu-Rull (2024).

Generative AI is just the latest shoot on the digital trunk that has changed lexicography forever over the last two or three decades, with huge implications for the underlying business model, the production process, the end product, and the way in which both old and new user needs are met, often customised and adapted to specific user needs. In the digital world, the lexicographers’ main focus has increasingly shifted from the dictionary to the database, which allows them to constantly modify, change, expand and update any lexicographic data at any time and without significant additional cost, since there is no longer a need to produce new editions of a given dictionary. On the contrary, the data stored in the lexicographic database can be used not only to feed a number of different dictionaries but also for other purposes and devices, e.g. linking them to elements in a text in the form of modern glosses (cf. Tarp & Gouws 2023). All this increases the possibility of finding alternative sources of funding for lexicographic projects.

The final proof of the pudding regarding the usefulness of AI and other digital technologies is that they can help to increase either the productivity or the quality of the lexicographic work, or even better, both. This will be discussed in the following sections. Section 2 looks at the requirements of modern lexicographic databases using the *Diccionario Digital del Español* (DIDES) as an example. Section 3 explains how generative AI can be used to produce lexicographic data, which are mainly but not exclusively prepared for a digital dictionary. Section 4 will then discuss how generative AI can contribute to the generation of glosses that are not related to specific words, but to different types of grammatical problems, and therefore stored in other types of lexicographic databases, and Section 5 will finally summarise the results.

## 2. New requirements to lexicographic databases

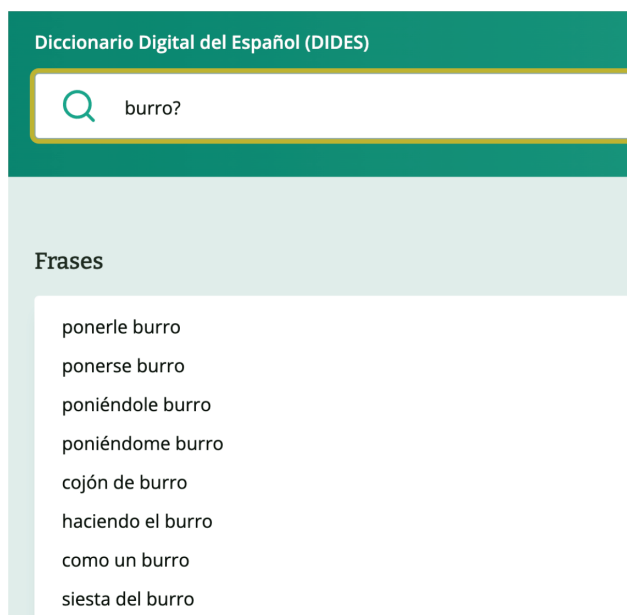
The term “lexicographic database” refers to a piece of specialised software designed to assist lexicographers in creating, editing and managing lexicographic data. This software offers a range of tools and features tailored to the specific needs of, for example, dictionary compilation, to ensure that the process is efficient, consistent, and collaborative. It might include an editor, a database, a web interface, various management tools, and a “dictionary grammar”, which according to Kilgarriff (2006: 7) is a kind of structure of the tool that lexicographers aim to produce. For instance, a lexicographic database with 20 slots, which contemplates the possibility of accessing and presenting the data stored in each slot only by itself or in combination with two or more of the rest of slots, should accommodate it to technological and economic possibilities. This accommodation is the “dictionary grammar” and it is necessary because it is not possible to propose 20! (20 factorial), because the access and presentation possibilities with 20! would be in the trillions. The dictionary grammar is then the result of deciding how to link each slot of the database, how to access these slots, and how to present the data to their potential users. If we offer users the possibility of extracting glosses, we can, for example, link the lemma and the definition, or the lemma, the definition and the synonyms, and so on. This means that when we search for a lemma, we extract only the definition (or definitions) and its synonyms.

An example of this philosophy is the *Diccionario Digital del Español (DIDES)*. Its lexicographic database is prepared for accessing and presenting several types of data. It allows three main types of access and data presentation. Firstly: searching for the lemma or part of it in the search engine, without (Boolean) operators. This retrieves the following:

1. A default screen showing the lemma, its grammar (e.g. part of speech and accompanying articles for nouns), definition preceded by geographical and diastatic tags (if any), a complete list of synonyms, a clickable note indicating that the lemma is part of several word combinations, and the clickable button “Ver mas” (See more).
2. An extended screen (the See more button) that displays sentence and phrase examples, synonyms and antonyms attached to its meaning, and links to external data such as figures, video and audio clips, and so on. It also has a clickable “Ver menos” (show less) button that allows users to return to the default screen.

Secondly, searching for the lemma or part of it with (Boolean) operators. This retrieves a list of clickable word combinations (see Figure 1). This is very relevant in this dictionary because it lemmatizes them, thus helping to eliminate or mitigate the *user burden*, that is, the challenges and difficulties that users of information systems may encounter when trying to find and understand the information they need (Fuertes-Olivera, 2025).

Figure 1. List of expressions accessed with the search string “burro?” in *DIDES* (excerpts)



Finally, internal cross-referencing, especially from clickable synonyms and word combinations, makes it easier to move around the system and explore the quantity and quality of its lexicographic data.

In summary, the above projects signal a way ahead for accommodating the theory and practice of lexicography to technological possibilities. As far as lexicographic databases are concerned, we assume that they should be ad-hoc and in-house systems that are flexible and adaptable to the characteristics of the lexicographic project. They should allow different ways of accessing and presenting the lexicographic data as well as facilitating a continuous and unstoppable process of compiling, modifying, checking and updating new data. Generative AI can be used for this, just as corpora and corpus query tools have been used in the semiautomatic projects mentioned above.

### 3. Using generative AI to produce lexicographic data

The basic method of producing lexicographic data with generative AI chatbots, such as ChatGPT 4o and Claude 3.5, is to dialogue with them by writing prompts that produce lexicographic data that are true, fair, and adequate for the aims and objectives of the lexicographic project in question. This implies a proper understanding of the process and its potential shortcomings. Fuertes-Olivera (2025) has proposed a three-stage methodology aimed at designing prompts whose responses are reliable and free of hallucinations or, if they are present, easily detectable.

Firstly, the lexicographer applies the concepts of “semantic entropy”, “lexical entropy” and “multi-agent systems” to generate prompts to be used for the same lexicographic task. Semantic entropy (Farquhar et al., 2024) assumes that the degree of

semantic inconsistency or unpredictability in the meaning of words or expressions in a given context can be mitigated (or eliminated) by repeating the same prompt(s) several times and clustering the responses according to their meanings. Lexical entropy (Karimi et al., 2024) defends that the unpredictability or variability associated with word choice in a given context can be “attacked” by varying the wording of prompts that are semantically similar. Finally, multi-agent systems (Wang et al. 2024) are based on the idea that hallucinations and similar problems can be spotted by giving the same prompts to two or more chatbots.

Secondly, the data should be validated for their veracity and accuracy. This process, typically referred to as curating the data, should not be concerned with offering “what is more characteristic of the way a given word behaves” (Rundell, 2023:7), but with what is more relevant to the project, e.g. short definitions as “glosses”, the figurative sense(s) of the lemma, and so on.

Validation is a process that requires ordering the data for inspection and declaring their validity. This process is typically initiated by elaborating a list of general prompts and arranging their responses in a document for visual inspection (cf. Appendix 1 uses a bi-text because it includes prompts and answers of two chatbots). In less than 10 minutes, we can scan the data of Appendix 1 and validate the following data, relying only on introspection, deduction and knowledge of Spanish:

- *Burro* is a noun, with several main “neutral” senses: ‘donkey’, ‘easel’, ‘trestle’, ‘pommel horse’ and ‘ironing board’ (they are neutral). The definitions, synonyms and example sentences accompanying the different senses are “useful”, i.e. they can be easily edited (if necessary).
- *Burro* is also a noun used in informal context to refer to a ‘fool or dumb person’. In this sense, it can be used as an adjective or as a noun, through a process of conversation or zero derivation (the adjective functions as a noun when it is preceded by articles).

Then, we need to do some investigation to validate data that we may not know, e.g. the geographical distribution of some senses and some expressions, including quotes, sayings and proverbs. This process is also very mechanical and does not take much time. For instance, googling [“burro” + planchar + México] yields more than 500 pages confirming that “burro” and “burro de planchar” (ironing board) are used in Central America and Mexico.

The third step occurs when we come across data that may need more or different investigation, e.g. by creating more focused prompts. An example of this can be seen in Appendix 1, which contains expressions such as “burro de carga”, “burro semental”, “burro de carpintero”, “cargar como un burro”, “al burro muerto, la cebada al rabo”, “no ver tres en un burro”, “burro grande ande o no ande”, ¡qué burrada!, and “trabajar como un burro”. With the exception of “burro semental”, we can validate the rest of the expressions as they are. At this stage we usually have two options. One of them is to create specific prompts for each of them, e.g. “Can you please describe lexicographically

“burro de carga”? Please answer me in Spanish and give me its literal and figurative meaning(s), synonyms, antonyms, part of speech, example sentences, and so on.

A related option is to explore, for instance, possible word combinations with *burro*. This can be done by creating more focused prompts (Example 1), and “validating” their responses using the methodology illustrated in Appendix 1:

- (a) We have the Spanish word [*burro*]. Can you please give us a list of multiword combinations with this Spanish word (and their English equivalents)? Please answer us in Spanish and do not forget to include lexicographic data for each multiword combination, especially meanings, synonyms, grammar, examples, variants, and so on.
- (b) Which Spanish word combinations with [*burro*] do you know? Can you please give us a list of them, as well as their English equivalents? Please answer us in Spanish and do not forget to contextualise each multiword expression lexicographically, e.g. give us meanings (literal and figurative), synonyms, antonyms, examples, grammar, variants, and so on.
- (c) What about the Spanish word [*burro*]. Can you please give us a list of Spanish expressions with this word and their English equivalents? Please answer us in Spanish and do not forget to contextualise each expression, e.g. give us meanings (literal and figurative), synonyms, antonyms, examples, grammar, variants, and so on.
- (d) We want to find Spanish expressions with [*burro*] that are formed by two or more orthographic words. Can you please give us a list of them as well as their English equivalents? Please answer us in Spanish and do not forget to contextualise each expression lexicographically, e.g. give us meanings (literal and figurative), synonyms, antonyms, examples, grammar, variants, and so on.
- (e) Many Spanish expressions with Spanish [*burro*] are formed with the preposition [*de*] in the middle, e.g. [*burro de planchar*]. Can you please give us a list of Spanish word formed in this way and their English equivalents)? Please answer us in Spanish and do not forget to contextualise each expression lexicographically, e.g. give us meanings (literal and figurative), synonyms, antonyms, examples, grammar, variants, and so on.
- (f) We have come across Spanish expressions initiated with [*burro*] and followed by one preposition [*de, para*], e.g. [*burro de carga*]. We think this pattern is common in Spanish and it may be used for creating word combinations with [*burro*]. If you agree, can you please give us a list of them, each with their English equivalent? Please answer us in Spanish and do not forget that we need them lexicographically described, e.g. with their meanings, examples, variants, synonyms, antonyms, grammar, and so on.

**Example 1.** List of specific prompts focused on word combinations with “*burro*”

In summary, this methodology has led to the creation and validation of all the senses of *burro* stored in the lexicographic database of *DIDES*, as well as 22 lemmatised expressions containing *burro* that can be accessed by entering the search string “burro?” in the search engine of *DIDES* (Figure 1). This methodology has proved to be productive and sound, both lexicographically and economically.

#### 4. Using generative AI to produce modern glosses

As mentioned in the introduction, lexicography is experiencing a certain return to its roots, as the digital environment increasingly allows for ways of satisfying user needs other than using dictionaries for this purpose, in particular through new types of glosses. In the classical tradition in cultures such as those of ancient Greece and China and the medieval Iberian Peninsula, glosses were used to explain unfamiliar or obsolete words, as was the case in manuscript copies of Homer's writings, or to help readers interpret religious and literary texts; see Yong & Peng (2008), Hanks (2013) and Nomdedeu-Rull & Tarp (2024), among others.

Modern glosses retain some of the original features, but also display new forms and content, as discussed by Tarp & Gouws (2019) and Tarp (2023). They appear above all in digital writing and reading assistants, where users can visualise them by clicking on words or expressions, in some cases already highlighted by the tool. In this sense, they not only explain the meaning of certain words, but also draw attention to grammatical problems that arise in specific contexts, offering alternative solutions and often explanations of the nature of the problem, as is the case in writing assistants like Grammarly and ProWritingAid.

These explanations are nothing more than glosses in digital clothing. Shifting or broadening the focus from words to grammar in context is clearly a challenge and a task for lexicographers if they are to continue their fine tradition of meeting user needs. It is certainly something new to learn, involving cutting-edge techniques and methodologies, but in many ways it is also a far better reflection of how language works, since lexicography (in the form of dictionaries) is essentially “an unnatural occupation” that “consists of tearing words from their mother context and setting them in rows” (Bolinger, 1985:69). Taking up the challenge can therefore give lexicography a new impetus and even allow it to become more natural.

An example of the relevance of lexicography to some new digital tools is a Spanish writing assistant currently under construction. In this project, lexicographers with experience in teaching Spanish to non-native speakers are working in close interdisciplinary collaboration with computer experts, contributing in one way or another to the accomplishment of five central tasks, some more or less similar to what they have already been doing, others completely new but nonetheless lexicographic in nature. These tasks are:

1. Compilation of a Spanish corpus to be used for training the language model that supports the writing assistant. This is a common task for lexicographers

in many dictionary projects, but in this case it is actually a new type of parallel corpus, which, unlike the usual parallel corpora, is monolingual with two identical sets of Spanish texts, where some errors have been introduced in one and corrected in the other (cf. Huete-García & Tarp, 2024).

2. Introduction of synthetic data into the model – i.e. all inflected full forms of nouns, verbs and adjectives together with their respective grammatical categories (part of speech, gender, number, person, tense and mood) – uploaded from a pre-existing lexicographic database (cf. Huete-García & Tarp, 2024). Of course, this presupposes that the database contains all these data categories, which, according to analyses by Fuertes-Olivera & Tarp (2020) and Tarp (2024), is by no means always the case.
3. Generation of validation data to check the performance of the writing assistant before it is presented to users. These data consist of two sets of sentences, one completely correct and the other containing specific grammatical errors. As far as we know, this is undoubtedly a new but relevant task for lexicographers in the digital age (cf. Tarp & Nomdedeu-Rull, 2024).
4. Writing of both short and long explanations of relevant problems, i.e. glosses, to help users understand and correct the errors they have made. This follows a long and time-honoured lexicographic tradition, but is new in the sense that the explanations do not focus on word meanings, but on types of grammatical problems, which in most cases are common to a smaller or larger group of words and word combinations (cf. Li et al., 2024).
5. Testing the writing assistant with real users to see if it works or if something needs to be modified. This is a common practice in many dictionary projects, although the testing methods have to be adapted to the specific tool (cf. Tarp, 2023).

To solve some of these tasks, the lexicographers used generative AI (until recently only ChatGPT 4o). This was the case for the parallel corpus (1), the validation data (2) and the long explanations (4), each of which – as explained by the authors quoted above – had its own particular challenges, required specially tailored prompts, produced results of varying quality, and represented a different relationship between the human lexicographer and the chatbot. In general, the researchers reported a significant increase in productivity with either the same or higher quality. However, only in one case, the generation of validation data, was it possible to make a precise comparison of productivity before and after the introduction of generative AI, showing a surprising and at the same time convincing 20-fold increase without compromising quality (cf. Tarp & Nomdedeu-Rull, 2024).

In the process of learning how to handle the generative AI chatbot, the researchers were inspired by some of its responses. When the chatbot was instructed to produce validation data, it sometimes gave an initial and spontaneous explanation of the grammatical problem in question (which it was not asked to do). The structure of some

of these explanations was then used as an inspiration to develop a model for the long explanations that was simple and didactic, making it easier to prompt the chatbot for this purpose. Figure 2 shows an explanation of grammatical disagreement between singular coordinate adjectives and a plural noun.

At the beginning of the project, the plan was to link the explanations to internal labels, or codes, which were automatically generated by the underlying language model after training on Spanish texts; see Tarp (2023). However, the number of labels soon grew to well over a million, most of which represented either random errors that would be extremely rare in real texts, or trivial errors that did not qualify for explanation. Nevertheless, there were still thousands, if not tens of thousands, of labels that would sooner or later require an explanation. It was therefore necessary to prioritise, starting with the most frequent ones, but this was much more complicated than it sounds. As documented by Li et al. (2024), there is no Spanish grammar book that approaches grammar in this way, i.e. from the perspective of the most common errors made in written texts, nor is there a Spanish learner corpus that can provide the necessary information. Moreover, although the lexicographers themselves had a long experience of teaching Spanish to non-native speakers and were able to suggest a number of common errors through introspection, the list was far from complete. They therefore decided to interact and experiment with generative AI to develop a methodology that would allow them to produce lists of relevant error types. The chatbot then generated lists of possible candidates, which the lexicographers either selected or rejected based on their knowledge and teaching experience, thus having the final say.

Appendix 2 reproduces the list of subcategories of gender agreement errors in Spanish generated using this methodology. As can be seen, the list is not a traditional linguistic one, since “the starting point is not grammatical rules as such, but only those

Figure 2. Explanation of disagreement between coordinate adjectives and a plural noun

Coordinate adjectives such as ‘español e inglés’, placed after a plural noun like ‘imperios’, must be written in the singular when one of them modifies only one of the entities to which the noun refers.

**INCORRECT EXAMPLE:**

- Los imperios españoles e ingleses fueron entre los primeros en colonizar América.

**CORRECT EXAMPLE:**

- Los imperios español e inglés fueron entre los primeros en colonizar a América.

**EXPLANATION:**

In the incorrect example, the adjectives ‘españoles’ and ‘ingleses’ are plural, meaning that there are two or more Spanish empires as well as English empires.

For the sentence to mean that there is one Spanish empire and one English empire, the adjectives must be in the singular, i.e. ‘español e inglés’.

error types that can and should be explained separately with the specific purpose of being displayed in a digital writing assistant” (Li et al., 2024:151). The three researchers also note that, although the list may not include all possible subcategories of the overall category of gender disagreement, they still consider it to be “a good starting point for the further development of the language model, until it can be tested by its target users and their errors used to fill in any gaps”.

Based on the experience of the above researchers with generative AI, Tarp (2024) developed a new way of writing short explanations that differs from the traditional generic use of expressions such as “may be”, “possible”, “it seems” and “it appears” as in the short explanation “The word X may be used incorrectly” taken from Grammarly (cf. Tarp, 2023). Instead, they designed a set of explanation templates and used AI to fill in the words in question. In this way, they managed to generate short glosses that not only describe the problem, but also explain why it is a problem, as in the following example:

The verb **cantan** is plural, but must be singular  
to agree in number with the subject **niño**.

#### **Example 2.** Short semi-automatically generated explanation

By combining this short explanation with a longer one that can be accessed by a simple click, they not only outlined a new type of grammar consisting of a large set of explanations like the one in Figure 2, but also a completely new approach to teaching and learning grammar inspired by the Hegelian concepts of “the universal”, “the particular”, and “the individual”, as explained in detail by Tarp (2025) in another article in this issue.

## **5. Conclusions and perspectives**

At the turn of the millennium, Grefenstette (1998) famously asked his colleagues whether there would be lexicographers in the year 3000. In response to Rundell’s (2012:18) optimistic prediction that lexicographers would still be around but doing something different, Li & Tarp (2024:398) added that not only would these lexicographers be using “different methods and techniques from those of today”, but the results of their work were likely “to be presented to future users in entirely new ways”. As the discussion in this paper shows, all these predictions have already come true, although the year 3000 is still a long way off.

When Grefenstette asked his famous question, digital technology was still in its infancy, while artificial intelligence was something rarely talked about outside the narrow circles where a handful of farsighted and dedicated scientists like 2024 Nobel laureate Geoffrey Hinton were confidently researching and developing this revolutionary technology. Less than three decades later, the situation has completely changed, even in a millennial discipline like lexicography.

The *DIDES* project, discussed in Sections 2 and 3, is evidence of this. It clearly shows that the lexicographer's focus is shifting from the dictionary as such to the supporting database, which now not only feeds the planned dictionary but also serves other purposes, and that even the online dictionary, with its multiple access and search options, represents a gigantic qualitative leap forward compared to its predecessors of barely two decades ago. The project also demonstrates that creative use of generative AI can add a new dimension to the compilation process, allowing faster and easier selection of less common or regional words, meanings, and other relevant lexicographic data types that are rarely included in the traditional corpora used for data selection, thus increasing not only productivity but also the quality of the final product.

Similarly, the Spanish Writing Assistant project, described in Section 4, shows how the digital technologies are giving rise to entirely new products that have paved the way for the glorious return of an old lexicographic expression like the gloss, but in a much more user-friendly, didactic, flexible and technologically advanced form.

In both projects, the new AI technology, and in particular generative AI, was more than just an object of observation and study. Much attention was also paid to the different types of relationships that emerge between the human lexicographer and the generative AI chatbot when the former decides to interact with the latter to solve different production tasks of varying nature. These relationships include:

- **“general review and control”** of the AI-generated corpora, where a few errors are not a relevant issue;
- **“meticulous revision and proofreading”** of the validation data, which must be 100% correct to serve their purpose;
- **“inspiration”** to (1) perform extended data selection, (2) identify error types that qualify for explanation, and (3) write long explanations, where the chatbot in all cases makes suggestions that the lexicographer then builds upon based on personal knowledge, skills and experience.

In these three types of relationship, the human always makes the final decision. This also implies that there is an impermeable wall between the chatbot and the users of the lexicographic product, who are never directly exposed to any AI-generated data, unless these data have already been approved, edited or processed by the lexicographer, who is the supreme filter and authority. As can be seen, the statement that the human lexicographer has the last word, is by no means an empty one.

All this allows us to say something essential about the lexicographers of the future. In order to fulfil their mission and interact with an artificial intelligence that is likely to become more and more advanced, lexicographers cannot rest on their technological laurels and allow themselves to be less qualified, knowledgeable and competent than their current colleagues. On the contrary, the demands on them are likely to be even greater than today if the users of their products are to be satisfied. They will need to be trained and prepared to recognise all cases where a chatbot is hallucinating and

providing fake data, and to have enough knowledge, experience and creativity to take over when it reaches its intrinsic technological limits.

In other words, whether or not lexicography has a bright future depends largely on the lexicographers themselves. If they do not do their job properly, AI may take over, and that may not be as fun as it sounds.

## References

- AL-KASIMI, Ali M. (1977). *Linguistic and Bilingual Dictionaries*. Leiden: EJ Brill Publishers.
- ALONSO-RAMOS, Margarita (2023). El papel de ChatGPT como lexicógrafo. In Cecilio GARRIGA-ESCRIBANO, Sandra IGLESIA-MARTÍN, José Antonio MORENO-VILLANUEVA & Antoni NOMDEDEU-RULL (eds.) (2023). *Lligams: Texts dedicats a María Bargalló Escrivà*. Tarragona: Publicacions URV, 15-27.
- ARIAS-ARIAS, Iván, DOMÍNGUEZ-VÁZQUEZ, María José & VALCÁRCEL-RIVEIRO, Carlos (2024). Der Effizienz- und Intelligenzbegriff in der Lexikographie und künstlichen Intelligenz: kann ChatGPT die lexikographische Textsorte nachbilden? *Lexikos*, 34(1), 51-76. <https://doi.org/10.5788/34-1-1879>
- BAISA, Vít, BLAHUŠ, Marek, CUKR, Michal, HERMAN, Ondřej, JAKUBÍČEK, Miloš, KOVÁŘ, Vojtěch, MEDVEĎ, Marek, MĚCHURA, Michal, RYCHLÝ, Pavel & SUCHOMEL, Vít (2019). "Automating Dictionary Production: a Tagalog-English-Korean Dictionary from Scratch". *Proceedings of the 6th Biennial Conference on Electronic Lexicography*. [https://elex.link/elex2019/wp-content/uploads/2019/09/eLex\\_2019\\_45.pdf](https://elex.link/elex2019/wp-content/uploads/2019/09/eLex_2019_45.pdf)
- BENATI, Chiara & HÄNDL Claudia (eds.) (2019). *From Glosses to Dictionaries: The Beginnings of Lexicography*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- BLAHUŠ, Marek, CUKR, Michal, HERMAN, Ondřej, JAKUBÍČEK, Miloš, KOVÁŘ, Vojtěch, KRAUS, Jan, MEDVEĎ, Marek, OHLÍDALOVÁ, Vlasta & SUCHOMEL, Vít (2023). Rapid Ukrainian-English Dictionary Creation Using Post-Edited Corpus Data. In M. MEDVEĎ, M. MĚCHURA, C. TIBERIUS, I. KOSEM, J. KALLAS, M. JAKUBÍČEK & S. KREK (eds.) (2023). *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference*. Brno: Lexical Computing, 613-637. 114.pdf (elex link)
- BOLINGER, Dwight (1985). Defining the undefinable. In R.F. ILSON (ed.) (1985). *Dictionaries, lexicography and language learning*. Oxford/New York/Toronto/Sydney/Frankfurt: Pergamon Press, 69-73.
- DE SCHRYVER, Giles-Maurice (2023). Generative AI and Lexicography: The Current State of the Art Using ChatGPT. *International Journal of Lexicography*, 36(4), 355-387. <https://doi.org/10.1093/ijl/ecad021>
- DE SCHRYVER, Giles-Maurice & JOFFE, David (2023). *The end of lexicography, welcome to the machine: On how ChatGPT can already take over all of the dictionary maker's tasks*. Paper presented at the 20th CODH Seminar, Tokyo. <http://codh.rois.ac.jp/seminar/lexicography-chatgpt-20230227>

- DIDES: Diccionario Digital del Español*: <https://diesgital.com/>
- FARQUHAR, Sebastian, KOSSEN, Jannik, KUHN, Lorenz & GAI, Yarin (2024). Detecting Hallucinations in Large Language Models Using Semantic Entropy. *Nature*, 630, 625-630. <https://doi.org/10.1038/s41586-024-07421-0>
- FUERTE-OLIVERA, Pedro A. (2024). Making Lexicography Sustainable: Using ChatGPT and Reusing Data for Lexicographic Purposes. *Lexikos*, 34(1), 123-140. <https://doi.org/10.5788/34-1-1883>
- FUERTE-OLIVERA, Pedro A. (2025). *A Guide to Practical Online Lexicography*. London: Routledge. <https://doi.org/10.4324/9781003426998>
- FUERTE-OLIVERA, Pedro A. & TARP, Sven (2020). A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica*, 36: 257-286. <https://doi.org/10.1515/lex-2020-0014>
- GRAFENSTETTE, Gregory (1998). The Future of Linguistics and Lexicographers: Will there be Lexicographers in the year 3000? In T. FONTENELLE, P. HILIGSMANN, A. MICHIELS, A. MOULIN & S. THEISSEN (eds.) (1998). *Proceedings of the Eighth EURALEX Congress*. Liège: University of Liège, 25-41.
- GOUWS, Rufus H. & TARP, Sven (2024). Despite Current Challenges: Lexicography has a bright future. *Lexicographica*, 40, 187-212. <https://doi.org/10.1515/lex-2024-0009>
- HANKS, Patrick (2013). Lexicography from Earliest Times to the Present. In K. ALLAN (ed.) (2013). *The Oxford Handbook of the History of Linguistics*. Oxford: Oxford University Press, 503-536.
- HUETE-GARCÍA, Ángel & TARP, Sven (2024). Training an AI-based Writing Assistant for Spanish Learners: the Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos*, 34, 21-40. <https://doi.org/10.5788/34-1-1862>
- JAKUBÍČEK, Miloš & RUNDELL, Michael (2023). The End of Lexicography? Can ChatGPT Outperform Current Tools for Post-editing Lexicography? In M. MEDVEĎ, M. MĚCHURA, C. TIBERIUS, I. KOSEM, J. KALLAS & M. JAKUBÍČEK (eds.) (2023). *Electronic Lexicography in the 21st Century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 Conference*. Brno: Lexical Computing, 518-533. 102. pdf (elex.link)
- KARIMI, Hossein, WEBER, Pete & ZINN, Jaden (2024). Information Entropy (not impedes) Lexical Processing during Language Comprehension". *Psychonomic Bulletin & Review*, 31, 2102-2117. <https://doi.org/10.3758/s13423-024-02463-x>
- KILGARRIFF, Allan (2006). Word from the Chair. In G.M. DE SCHRUYVER (eds.) (2006). *DWS 2006: proceedings of the Fourth International Workshop on Dictionary Writing Systems*, 7. Pretoria: SF Press.
- LEW, Robert (2023). ChatGPT as a COBUILD Lexicographer. *Humanities and Social Sciences Communications*, 10(704). <https://doi.org/10.1057/s41599-023-02119-6>
- LEW, Robert (2024). Dictionaries and Lexicography in the AI Era. *Humanities and Social Sciences Communications*, 11(426). <https://doi.org/10.1057/s41599-024-02889-7>
- LI, Qian & TARP, Sven (2024). Using Generative AI to Provide High-quality Lexicographic Assistance to Chinese Learners of English. *Lexikos*, 34, 397-418. <https://doi.org/10.5788/34-1-1944>

- LI, Qian, TARP, Sven & NOMDEDEU-RULL, Antoni (2024). The Necessary Symbiosis: How ChatGPT Co-authored a New Type of Learner's Grammar. *Círculo de lingüística aplicada a la comunicación*, 100: 145-160. <https://doi.org/10.5209/clac.97409>
- MARTHUR, Tom (1986). *Worlds of Reference. Lexicography, Learning and Language from the Clay Tablet to the Computer*. Cambridge: Cambridge University Press.
- MCKEAN, Erin & FITZGERALD, Will (2024). The ROI of AI in lexicography. *Lexicography*, 11(1), 7-27. <https://doi.org/10.1558/lexi.27569>
- NOMDEDEU-RULL, Antoni & TARP, Sven (2024). *Introducción a la Lexicografía en Español: Funciones y Aplicaciones*. London: Routledge.
- PTASZNIK, Bartosz, WOLFER, Sascha & LEW, Robert (2024). A Learners' Dictionary Versus ChatGPT in Receptive and Productive Lexical Tasks. *International Journal of Lexicography*, 37 (3), 322-336. <https://doi.org/10.1093/ijl/ecae011>
- REES, Geraint Paul & LEW, Robert (2024). The Effectiveness of OpenAI GPT-Generated Definitions Versus Definitions from an English Learners' Dictionary in a Lexically Orientated Reading Task. *International Journal of Lexicography*, 37(1), 50-74. <https://doi.org/10.1093/ijl/ecad030>
- RUIZ ASENCIO, José Manuel, RUIZ ALBI, Irene & HERRERO JIMÉNEZ, Mauricio (2020). *Las Glosas Silenses. Estudio crítico y edición facsímil. Versión castellana del Penitencial*. Burgos: Instituto Castellano y Leonés de la Lengua.
- RUNDELL, Michael (2012). The Road to Automated Lexicography: An Editor's Viewpoint. In S. Granger & M. Paquot (eds.) (2012). *Electronic Lexicography*. Oxford: Oxford University Press, 15-30.
- RUNDELL, Michael (2023). Automating the Creation of Dictionaries: Are we Nearly There? In *Asialex 2023 Proceedings. Lexicography, Artificial Intelligence, and Dictionary Users*. Seoul: Yonsei University, 9-17.
- STATHI, Ekaterini (2006). Greek Lexicography, Classical. In K. BROWN (ed.) (2006). *Encyclopedia of Language and Linguistics. Vol. 5*. Second Edition. Oxford: Elsevier, 145-146.
- TARP, Sven (2023). Eppur si muove: Lexicography Is Becoming Intelligent! *Lexikos*, 33(2), 107-131. <https://doi.org/10.5788/33-2-1841>
- TARP, Sven (2024). Danish Lexicography between Tradition and Innovation: Digital Dictionaries and Needs-adapted Glosses. *Lexicographica*, 40, 95-120. <https://doi.org/10.1515/lex-2024-0006>
- TARP, Sven (2025). Lexicografía innovadora: Generando una gramática de abajo arriba para acercarse a los aprendices de lengua. *Revista Internacional de Lenguas Extranjeras*, 23, 76-94. <https://doi.org/10.17345/rile23.4197>
- TARP, Sven & GOUWS, Rufus H. (2019). Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos*, 29, 250-268. <https://doi.org/10.5788/29-1-1520>
- TARP, Sven & GOUWS, Rufus H. (2023). A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and the Implications for the Future. *Lexikos*, 33, 425-447. <https://doi.org/10.5788/33-1-1826>

- TARP, Sven & NOMDEDEU-RULL, Antoni (2024). Who Has the Last Word? Lessons from Using ChatGPT to Develop an AI-based Spanish Writing Assistant. *Círculo de lingüística aplicada a la comunicación*, 97, 309-321. <https://doi.org/10.5209/clac.91985>
- WANG, Qineng, WANG, Zihao, SU, Ying, TONG, Hanghang & SONG, Yangqiu (2024). Rethinking the Bounds of LLM Reasoning: Are Multi-Agent Discussions the Key? *ArXiv*. <https://arxiv.org/abs/2402.18272>
- YONG, Heming & PENG, Jing (2008). *Chinese Lexicography. A History from 1046 BC to AD 1911*. Oxford: Oxford University Press.

## Appendix 1

Methodological Stages commonly used for working with generative AI for generating lexicographic data for *DIDES*

**Stage one:** Lists of initial prompts elaborated for finding out general lexicographic data for the Spanish word “burro”:

- (a) we think that Spanish [*burro*] is a noun. Can you please confirm it? Do you know if [*burro*] is also an adjective, adverb, verb, preposition, and on? Can you please answer me in Spanish?
- (b) We also think Spanish [*burro*] has several meanings. Can you please give us a list of them, each contextualized, i.e. with examples, synonyms, antonyms, related words, etc. In other words, we need lexicographic data on Spanish [*burro*]. Can you please answer us in Spanish? Please give us also English equivalents for each meaning.
- (c) We think that some of the meanings of Spanish [*burro*] you mention are only used in some Spanish-speaking countries. Can you please give us a list of them (countries) and their English equivalents? Please answer us in Spanish, and do not forget to contextualize lexicographically each meaning.

**Stage two:** Creation of a bi-text document for validating the data generated with both chatbots (for reasons of space, we will only include two prompts and their answers):

Prompt with ChatGPT 4o (20, November 2024)	Prompt with Claude 3.5 (Pro) (20 November, 2024)
we think that Spanish [ <i>burro</i> ] is a noun. Can you please confirm it? Do you know if [ <i>burro</i> ] is also an adjective, adverb, verb, preposition, and on? Can you please answer me in Spanish?	we think that Spanish [ <i>burro</i> ] is a noun. Can you please confirm it? Do you know if [ <i>burro</i> ] is also an adjective, adverb, verb, preposition, and on? Can you please answer me in Spanish?

<p><b>Answer</b></p> <p>Sí, puedo confirmar que <b>[burro]</b> en español es un sustantivo. Se utiliza para referirse a un animal, específicamente al asno. Además, <b>[burro]</b> no se utiliza como adjetivo, adverbio, verbo ni preposición en español. Sin embargo, en sentido figurado, a veces se emplea como adjetivo coloquial para describir a una persona que se percibe como poco inteligente o torpe, por ejemplo: "Eres un burro por no entender esto".</p> <p>Por lo tanto, su uso principal es como sustantivo, y ocasionalmente se usa como un adjetivo en contextos informales o figurativos. No se emplea en otras categorías gramaticales.</p> <p>¿Hay algo más en lo que pueda ayudarte? 😊</p>	<p><b>Answer</b></p> <p>Sí, analizaré las diferentes categorías gramaticales de "burro":</p> <ol style="list-style-type: none"> <li>1. Como sustantivo (uso principal):             <ul style="list-style-type: none"> <li>• Animal (asno)</li> <li>• Aparato/mueble (burro de planchar)</li> <li>• Caballete (para trabajos) Ejemplo: "El burro carga la mercancía"</li> </ul> </li> <li>2. Como adjetivo:             <ul style="list-style-type: none"> <li>• Significado: torpe, ignorante, poco inteligente</li> <li>• Uso coloquial y a menudo despectivo Ejemplo: "Es muy burro para las matemáticas"</li> </ul> </li> <li>3. Formas verbales relacionadas:             <ul style="list-style-type: none"> <li>• "Aburrar(se)": volverse torpe</li> <li>• "Emberrar(se)": obstinarse tontamente (Estos verbos son poco comunes y muy coloquiales)</li> </ul> </li> </ol> <p>No funciona como:</p> <ul style="list-style-type: none"> <li>• Adverbio</li> <li>• Preposición</li> <li>• Conjunción</li> <li>• Interjección</li> </ul> <p>Usos principales:</p> <ol style="list-style-type: none"> <li>1. Sustantivo:             <ul style="list-style-type: none"> <li>• Animal</li> <li>• Mueble/aparato</li> <li>• Herramienta</li> </ul> </li> <li>2. Adjetivo:             <ul style="list-style-type: none"> <li>• Cualidad negativa</li> <li>• Característica de torpeza</li> <li>• Falta de inteligencia</li> </ul> </li> </ol> <p>Es importante notar que el uso como adjetivo es considerado coloquial y puede ser ofensivo en algunos contextos.</p> <p>Los usos más formales y aceptados son como sustantivo.</p>
--	--

<p><b>Prompt with ChatGPT 4o (20, November 2024)</b></p> <p>We also think Spanish [<i>burro</i>] has several meanings. Can you please give us a list of them, each contextualized, i.e. with examples, synonyms, antonyms, related words, etc. In other words, we need lexicographic data on Spanish [<i>burro</i>]. Can you please answer us in Spanish? Please give us also English equivalents for each meaning.</p>	<p><b>Prompt with Claude 3.5 (Pro) (20 November, 2024)</b></p> <p>We also think Spanish [<i>burro</i>] has several meanings. Can you please give us a list of them, each contextualized, i.e. with examples, synonyms, antonyms, related words, etc. In other words, we need lexicographic data on Spanish [<i>burro</i>]. Can you please answer us in Spanish? Please give us also English equivalents for each meaning.</p>
<p>Answer</p> <p>Claro, aquí tienes un análisis lexicográfico del término español “burro” con sus diferentes significados, ejemplos, sinónimos, antónimos, palabras relacionadas y equivalentes en inglés:</p> <ol style="list-style-type: none"> <li>Significado principal: Animal (Asno) <ul style="list-style-type: none"> <li>Definición: Mamífero doméstico de la familia de los équidos, similar al caballo pero de menor tamaño, orejas largas y naturaleza resistente.</li> <li>Ejemplo: <i>El burro lleva la carga a través de las montañas.</i></li> <li>Sinónimos: Asno, jumento.</li> <li>Antónimos: Caballo (aunque no es opuesto, es una especie relacionada).</li> <li>Palabras relacionadas: Mula, équidos, ganado.</li> <li>Equivalente en inglés: <i>Donkey</i>.</li> </ul> </li> <li>Uso figurado: Persona torpe o ignorante (coloquial) <ul style="list-style-type: none"> <li>Definición: Persona considerada poco inteligente o con dificultad para aprender.</li> <li>Ejemplo: <i>Eres un burro si no estudias para el examen.</i></li> <li>Sinónimos: Tonto, necio, ignorante.</li> <li>Antónimos: Inteligente, listo, sabio.</li> <li>Palabras relacionadas: Torpeza, ignorancia, falta de habilidad.</li> <li>Equivalente en inglés: <i>Fool, dumb, ignorant person</i>.</li> </ul> </li> <li>Adjetivo coloquial: Persona poco hábil en algo <ul style="list-style-type: none"> <li>Definición: Usado como adjetivo para describir a alguien que tiene dificultades con una tarea o habilidad.</li> <li>Ejemplo: <i>Soy burro para las matemáticas.</i></li> <li>Sinónimos: Malo, torpe, incapaz.</li> <li>Antónimos: Hábil, talentoso, diestro.</li> <li>Palabras relacionadas: Incapacidad, falta de destreza.</li> <li>Equivalente en inglés: <i>Clumsy, bad at something</i>.</li> </ul> </li> </ol>	<p>Answer</p> <p>Por supuesto. Aquí tiene el análisis lexicográfico de “burro”:</p> <ol style="list-style-type: none"> <li>Animal (sustantivo): Español: Mamífero équido, asno doméstico English: Donkey, ass <ul style="list-style-type: none"> <li>Sinónimos: asno, pollino, jumento</li> <li>Antónimos: N/A</li> <li>Familia léxica: burra, burrito, borriquito</li> <li>Ejemplo: “El burro carga las mercancías por el sendero”</li> <li>Colocaciones: burro de carga, burro semental</li> <li>Contexto: Rural, agrícola</li> </ul> </li> <li>Persona torpe/ignorante (adjetivo): Español: Persona de poco entendimiento, torpe English: Stupid, dumb, dense <ul style="list-style-type: none"> <li>Sinónimos: tonto, necio, ignorante</li> <li>Antónimos: listo, inteligente, avisado</li> <li>Ejemplo: “Es muy burro para las matemáticas”</li> <li>Uso: coloquial, puede ser ofensivo</li> <li>Expresiones: “Más burro que un arado”</li> </ul> </li> <li>Mueble/Soporte (sustantivo): Español: Armazón que sirve de soporte English: Stand, support, trestle <ul style="list-style-type: none"> <li>Tipos: burro de planchar, burro de carpintero</li> <li>Sinónimos: soporte, caballete</li> <li>Ejemplo: “Pon la tabla sobre el burro para serrar”</li> <li>Contexto: Carpintería, tareas domésticas</li> </ul> </li> <li>Dispositivo gimnástico (sustantivo): Español: Aparato de gimnasia English: Gymnastics horse, pommel horse <ul style="list-style-type: none"> <li>Sinónimos: potro</li> <li>Contexto: Deportivo, gimnástico</li> <li>Ejemplo: “Los gimnastas practican en el burro”</li> </ul> </li> </ol>

4. Objeto: Caballete o soporte
- Definición: Estructura o herramienta en forma de cruz utilizada para sostener materiales, especialmente en carpintería o pintura.
  - Ejemplo: *Coloca la madera sobre el burro para cortarla con precisión.*
  - Sinónimos: Caballete, soporte.
  - Antónimos: —
  - Palabras relacionadas: Herramientas, carpintería, trabajo manual.
  - Equivalente en inglés: *Sawhorse, trestle.*
5. Objeto escolar: Mochila o bolsa grande (en desuso)
- Definición: Bolsa o mochila que se usaba para cargar libros escolares.
  - Ejemplo: *En mi época de niño llevaba los libros en un burro al colegio.*
  - Sinónimos: Mochila, cartera.
  - Antónimos: —
  - Palabras relacionadas: Escolar, equipaje.
  - Equivalente en inglés: *School bag* (aunque el término es más regional y en desuso).
6. Expresión regional: Café muy cargado
- Definición: En algunas regiones, se usa para referirse a un café fuerte o muy concentrado.
  - Ejemplo: *Por la mañana siempre tomo un burro para despertarme.*
  - Sinónimos: Café fuerte.
  - Antónimos: Café suave.
  - Palabras relacionadas: Bebidas, café.
  - Equivalente en inglés: *Strong coffee.*
7. Diminutivos y aumentativos con matices figurativos
- Burro grande (aumentativo): Persona extremadamente torpe o grande físicamente.
    - Ejemplo: *Ese burro grande rompió la silla al sentarse.*
  - Burrito (diminutivo): Puede usarse afectivamente o para referirse a un pequeño burro o al alimento (en contextos de América Latina).
    - Ejemplo: *El burrito juega en el campo.*
    - En América Latina: *Voy a comer un burrito de carne.*

¿Hay algún significado o contexto adicional que quieras explorar más a fondo? 😊

5. En juegos (sustantivo): Español: Persona que pierde en ciertos juegos English: Loser (in certain games)

- Contexto: Juegos infantiles
- Ejemplo: "Le tocó ser el burro en el juego"

Expresiones idiomáticas:

- "Cargar como un burro"
  - Significado: Trabajar excesivamente
  - English: To work like a dog
  - Ejemplo: "Lleva todo el día cargando como un burro"
- "Al burro muerto, la cebada al rabo"
  - Significado: Ayuda que llega tarde
  - English: Too little, too late
  - Uso: Refrán tradicional
- "No ver tres en un burro"
  - Significado: Ver muy mal
  - English: To be as blind as a bat
  - Contexto: Coloquial

Notas de uso:

- Regional: Algunas expresiones varían según el país
- Registro: Mayormente coloquial en usos figurados
- Cultural: Importante en el refranero español

Campos semánticos:

- Animales de carga
- Herramientas/muebles
- Cualidades intelectuales (uso peyorativo)
- Deportes/gimnasia

<p><b>Prompt with ChatGPT 4o (21 November 21, 2024)</b></p> <p>We think that some of the meanings of Spanish [<i>burro</i>] you mention are only used in some Spanish-speaking countries. Can you please give us a list of them (countries) and their English equivalents? Please answer us in Spanish, and do not forget to contextualize lexicographically each meaning.</p>	<p><b>Prompt with Claude 3.5 (21 November 21, 2024)</b></p> <p>We think that some of the meanings of Spanish [<i>burro</i>] you mention are only used in some Spanish-speaking countries. Can you please give us a list of them (countries) and their English equivalents? Please answer us in Spanish, and do not forget to contextualize lexicographically each meaning.</p>
<p><b>Answers</b></p> <p>¡Claro! A continuación, detallo los significados regionales de <b>“burro”</b> que tienen un uso específico en ciertos países hispanohablantes, junto con sus equivalentes en inglés y su contextualización lexicográfica.</p> <p><b>1. Burro como alimento (México y regiones de influencia)</b></p> <ul style="list-style-type: none"> <li>• <b>Significado:</b> En algunas zonas de México y el suroeste de Estados Unidos, <b>“burro”</b> es una variante del <b>burrito</b>, un platillo hecho de tortilla de harina rellena de carne, frijoles y otros ingredientes.</li> <li>• <b>Países:</b> México, Estados Unidos (zonas con población hispanohablante).</li> <li>• <b>Equivalente en inglés:</b> Burrito (Mexican dish).</li> <li>• <b>Ejemplo:</b> “En Sonora, el burro percherón es un plato típico muy popular.”</li> <li>• <b>Nota cultural:</b> El nombre proviene del tamaño más grande de estos burritos en comparación con los normales.</li> </ul> <p><b>2. Burro como soporte (Latinoamérica, España)</b></p> <ul style="list-style-type: none"> <li>• <b>Significado:</b> En carpintería o limpieza, un <b>“burro”</b> es un caballete o soporte para sostener objetos.</li> <li>• <b>Países:</b> España y la mayoría de los países de América Latina.</li> <li>• <b>Equivalente en inglés:</b> Sawhorse, trestle.</li> <li>• <b>Ejemplo:</b> “En Argentina, los carpinteros llaman burros a los caballetes para serrar madera.”</li> <li>• <b>Nota cultural:</b> Este uso es casi universal en los países hispanohablantes.</li> </ul>	<ul style="list-style-type: none"> <li>• Sí, tiene razón. Aquí está la distribución geográfica de los diferentes significados de “burro”:</li> </ul> <ol style="list-style-type: none"> <li>1. Animal (sustantivo):             <ul style="list-style-type: none"> <li>• Universal en todos los países hispanohablantes</li> <li>• English: Donkey, ass</li> <li>• Uso común y estándar</li> </ul> </li> <li>2. Persona torpe/ignorante:             <ul style="list-style-type: none"> <li>• España, México, Argentina, Chile, Colombia</li> <li>• English: Stupid, dumb</li> <li>• Uso coloquial</li> </ul> </li> <li>3. Burro de planchar:             <ul style="list-style-type: none"> <li>• México: muy común</li> <li>• Venezuela: uso frecuente</li> <li>• Colombia: algunas regiones</li> <li>• República Dominicana</li> <li>• English: Ironing board</li> <li>• Otros países usan “tabla de planchar”</li> </ul> </li> <li>4. Persona terca/obstinada:             <ul style="list-style-type: none"> <li>• México: uso muy común</li> <li>• Colombia: frecuente</li> <li>• Venezuela: común</li> <li>• Argentina: menos frecuente</li> <li>• English: Stubborn person</li> </ul> </li> <li>5. En juegos de cartas:             <ul style="list-style-type: none"> <li>• España: común</li> <li>• Argentina: frecuente</li> <li>• Uruguay: usual</li> <li>• English: Loser (in card games)</li> </ul> </li> <li>6. Mal estudiante:             <ul style="list-style-type: none"> <li>• México: muy común</li> <li>• España: frecuente</li> <li>• Colombia: usual</li> <li>• Chile: común</li> <li>• English: Poor student, dunce</li> </ul> </li> </ol>

**3. Burro como sinónimo de ignorante (España y América Latina)**

- **Significado:** Persona torpe o ignorante, usualmente en contextos educativos o intelectuales.
- **Países:** Uso generalizado en España y América Latina, pero con mayor frecuencia en España, México, Argentina y Colombia.
- **Equivalente en inglés:** Dummy, idiot.
- **Ejemplo:** “En España, se escucha mucho decir ‘Ese niño es un burro para las matemáticas.’”
- **Nota cultural:** Puede ser ofensivo dependiendo del tono y la intención del hablante.

**4. Burro de carga (coloquial, figurativo)**

- **Significado:** Persona que realiza tareas pesadas o numerosas sin quejarse.
- **Países:** Uso extendido en España, México, Perú y Colombia.
- **Equivalente en inglés:** Pack mule.
- **Ejemplo:** “En México, un trabajador muy dedicado a menudo es apodado ‘burro de carga.’”
- **Nota cultural:** Expresa tanto admiración como crítica según el contexto.

**5. Burro como juego infantil (América Latina)**

- **Significado:** Juego en el que un niño se agacha (como un burro) para que los demás salten sobre él, similar al “leapfrog.”
- **Países:** México, Guatemala, Perú y otros países de Centroamérica y Sudamérica.
- **Equivalente en inglés:** Leapfrog (base or player).
- **Ejemplo:** “En Perú, los niños juegan al burro castigado en el recreo.”
- **Nota cultural:** Este juego tiene variaciones regionales y nombres diferentes según el país.

**6. Burro en frases hechas (España y América Latina)**

- **Frase:** “Burro grande, ande o no ande.”
  - **Significado:** Se refiere a la preferencia por objetos grandes, aunque no sean prácticos.
  - **Países:** España, México, Argentina, Venezuela.
  - **Equivalente en inglés:** Bigger is better (literalmente: “Big donkey, whether it walks or not”).
  - **Ejemplo:** “En España, usan ‘burro grande’ para bromear sobre quienes compran coches enormes sin necesitarlos.”

7. “Caer de burro”:

- México: uso común
- Colombia: frecuente
- Venezuela: usual
- English: To finally get it

8. “No ver tres en un burro”:

- España: muy común
- México: frecuente
- Argentina: poco común
- English: To be as blind as a bat

9. Burro (aparato gimnástico):

- España: común
- Argentina: frecuente
- Chile: usual
- English: Gymnastics horse

10. Burro (soporte/caballete):

- México: muy común
- España: frecuente
- Colombia: usual
- English: Trestle, stand

11. “Trabajar como un burro”:

- Universal en países hispanohablantes
- English: To work like a dog/donkey

Notas importantes:

1. Variación regional:

- Los significados pueden variar incluso dentro de un mismo país
- Algunas expresiones son más comunes en zonas rurales
- El uso urbano/rural puede diferir

2. Registro:

- La mayoría son usos coloquiales
- Algunos son considerados vulgares en ciertos países
- El nivel de aceptación social varía

3. Frecuencia de uso:

- Varía según región y contexto social
- Algunas acepciones están cayendo en desuso
- Nuevos significados emergen en diferentes zonas

4. Contexto cultural:

- Las expresiones reflejan aspectos culturales locales
- El grado de ofensa varía según el país
- Algunos usos son más aceptados socialmente que otros

Esta distribución es aproximada y puede variar con el tiempo y según zonas específicas dentro de cada país.

<ul style="list-style-type: none"> <li>• <b>Frase:</b> “¡Qué burrada!” <ul style="list-style-type: none"> <li>◦ <b>Significado:</b> Exclamación para referirse a un error; exageración o acción torpe.</li> <li>◦ <b>Países:</b> España, México, Argentina, Chile.</li> <li>◦ <b>Equivalente en inglés:</b> What a blunder!</li> <li>◦ <b>Ejemplo:</b> “En Chile dicen ‘¡Qué burrada!’ cuandalguien comete un error ridículo.”</li> </ul> </li> </ul> <p><b>7. Burro en Argentina y Uruguay (persona terca)</b></p> <ul style="list-style-type: none"> <li>• <b>Significado:</b> Persona obstinada que no cambia de opinión.</li> <li>• <b>Países:</b> Argentina y Uruguay.</li> <li>• <b>Equivalente en inglés:</b> Stubborn person.</li> <li>• <b>Ejemplo:</b> “En Uruguay dicen ‘es más burro que una mula’ para alguien muy terco.”</li> <li>• <b>Nota cultural:</b> Esta metáfora es similar al inglés “stubborn as a mule.”</li> </ul> <p><b>8. Burro en contextos rurales (Ecuador, Perú, Bolivia)</b></p> <ul style="list-style-type: none"> <li>• <b>Significado:</b> Persona sencilla o trabajadora, con frecuencia en comunidades agrícolas.</li> <li>• <b>Países:</b> Ecuador, Perú y Bolivia.</li> <li>• <b>Equivalente en inglés:</b> Simple hard worker (with rural connotations).</li> <li>• <b>Ejemplo:</b> “En las comunidades andinas de Perú, se respeta al ‘burro’ como símbolo de esfuerzo y humildad.”</li> <li>• <b>Nota cultural:</b> El burro es un animal esencial en las labores del campo, lo que da al término un significado respetuoso.</li> </ul> <p><b>Etimología regional y variación cultural</b></p> <p>La mayoría de las extensiones metafóricas de “burro” derivan de la asociación del animal con características como la terquedad, la resistencia y la simpleza. Sin embargo, su interpretación varía según el país, el contexto social y el tono del hablante. Algunas expresiones pueden ser afectuosas, mientras que otras resultan despectivas.</p> <p>¿Te gustaría que explore algún significado o región en particular? 😊</p>	
---	--

**Stage three:** validation of the data generated with the chatbots.

## Appendix 2

List of sub-categories of gender agreement errors in Spanish

### 1. Noun and adjective

- 1.1. Disagreement between noun and postposed adjective.
- 1.2. Disagreement between singular noun and postposed non-variable adjective
- 1.3. Disagreement between singular noun and anteposed adjective
- 1.4. Disagreement between same-gender coordinated nouns and adjective
- 1.5. Disagreement between different-gender coordinated nouns and adjective
- 1.6. Disagreement between singular coordinate adjective and plural noun

### 2. Noun and determiner

- 2.1. Disagreement between article and noun
- 2.2. Disagreement between masculine singular article and feminine noun
- 2.3. Disagreement between coordinated nouns and article
- 2.4. Disagreement between noun and demonstrative determiner
- 2.5. Disagreement between noun and definite determiner
- 2.6. Disagreement between noun and indefinite determiner
- 2.7. Disagreement between noun and possessive determiner

### 3. Noun and pronoun

- 3.1. Disagreement between noun and atonic personal pronoun
- 3.2. Disagreement between antecedent and possessive pronoun
- 3.3. Disagreement between antecedent and relative pronoun

### 4. Noun and participle

- 4.1. Disagreement between noun and participle

### 5. Pronoun and adjective

- 5.1. Disagreement between personal pronoun and adjective
- 5.2. Disagreement with adjective after pronouns *alguien* and *nadie*

### 6. Adverb and adjective/adverb

- 6.1. Disagreement with adjective after quantifier adverb *medio*
- 6.2. Disagreement with adjective/adverb after indefinite adverbs *algo* and *nada*

**7. Number and noun**

7.1. Disagreement with noun after cardinal or decimal

7.2. Disagreement with cardinal ending in *un* or *una* before noun

7.3. Disagreement with partitive constructions such as '*uno de los* + noun'

**8. Adjective and subject in *que*-clause**

8.1. Disagreement in constructions with '*lo* + adjective + *que*-clause'

Quoted from Li et al. (2024:151)