

Eppur si muove: Lexicography is Becoming Intelligent!

Sven Tarp, *Department of Afrikaans and Dutch, Stellenbosch University, South Africa; Centre for Lexicographical Studies, Guangdong University of Foreign Studies, China; International Centre for Lexicography, Valladolid University, Spain; Centre of Excellence in Language Technology, Ordbogen A/S, Denmark; and Centre for Lexicography, Aarhus University, Denmark (st@cc.au.dk)*

Abstract: The paper focuses on an ongoing R&D project which the author is conducting together with Spanish lexicographers as well as computer scientists from a high-tech company specialized in online dictionaries and language services. The objective is to develop an AI-powered Spanish writing assistant for both native and non-native writers and learners. After briefly discussing the current experiences with digital writing assistants, the paper will detail the concrete project, where the lexicographers' task is, on the one hand, to contribute to the training of the underlying language model, and on the other hand, to outline a model for good communication between the tool and its users. Based on a study of practice in existing writing assistants, the paper will then formulate a set of principles for user communication that will be implemented in the writing assistant. The article gives examples of how this implementation takes place, what new challenges it poses, and how the writing assistant will eventually work. Finally, it discusses the lexicographers' new tasks and outlines some perspectives for future work.

Keywords: WRITING ASSISTANTS, USER COMMUNICATION, LEXICOGRAPHICAL CONTEXTUALIZATION, INTEGRATED DICTIONARIES, GLOSSES, LEXICOGRAPHICAL DATA OF A NEW TYPE, INCIDENTAL LEARNING, INTENTIONAL LEARNING, LEMMA-CENTRED LEXICOGRAPHICAL DATABASES, PROBLEM-CENTRED LEXICOGRAPHICAL DATABASES

Opsomming: Eppur si muove: Die leksikografie word intelligent! In hierdie artikel word gefokus op 'n deurlopende N&O-projek wat die outeur saam met Spaanse leksikograwe, asook rekenaarwetenskaplikes van 'n hoëtegnologiemaatskappy, wat spesialiseer in aanlyn-woordeboeke en taaldienste, onderneem. Dit het die ontwikkeling van 'n KI-gedrewe Spaanse skryfhulpmiddel vir beide moeder- en niemoedertaalskrywers en -leerders ten doel. Ná 'n kort bespreking van die huidige ervaring met digitale skryfhulpmiddels word die konkrete projek, waarvan die leksikograwe se taak, aan die een kant, is om 'n bydrae te lewer tot die afrigting van die onderliggende taalmodel, en aan die ander kant, om 'n model vir goeie kommunikasie tussen die werktuig en die gebruikers daarvan te beskryf, uitvoerig uiteengesit. Gebaseer op 'n studie van die ervaring in bestaande skryfhulpmiddels, sal daarna 'n stel beginsels vir gebruikerskommunikasie geformuleer word wat in die skryfhulpmiddel geïmplementeer sal word. In die artikel word voorbeelde

gegee van hoe die implementering plaasvind, watter nuwe uitdagings teengekom word, en hoe die skryfhulpmiddel uiteindelik sal funksioneer. Laastens word die leksikograaf se nuwe take bespreek en 'n paar vooruitsigte vir toekomstige werk word uiteengesit.

Sleutelwoorde: SKRYFHULPMIDDELS, GEBRUIKERSKOMMUNIKASIE, LEKSIKOGRAFIESE KONTEKSTUALISERING, GEÏNTEGREERDE WOORDEBOEKE, GLOSSE, 'N NUWE TIPE LEKSIKOGRAFIESE DATA, TOEVALLIGE LEER, DOELBEWUSTE LEER, LEMMAGESENTREERDE LEKSIKOGRAFIESE DATABASISSE, PROBLEEMGESENTREERDE LEKSIKOGRAFIESE DATABASISSE

1. Introduction

This article is dedicated to my friend and colleague Rufus H. Gouws. His 43 years of highly productive academic work may seem a lot to many of us (and indeed it is), but it could also be argued that these 43 years represent only a short span of time from the perspective of more than 4,300 years of lexicographical history. Be that as it may, the past four decades have seen more dramatic and rapid changes than any previous period. And the transformation process shows no signs of slowing down, hence the famous quotation from Galileo included in the title.

Today, lexicography is becoming intelligent. Artificial Intelligence (AI) is penetrating the millennial discipline from various angles: data selection, preparation, storage, presentation and integration into different tools; cf. Tarp (2019). The disruption is almost total. AI-based tools are taking over more and more creative tasks, sending almost fully processed data to lexicographers, whose job is increasingly reduced to control, revision, and final editing. There are even examples of AI-generated data, which can rightly be considered lexicographical, now being presented directly to the target audience without a single lexicographer having been over it. Thus, the relationship between the human and the artificial lexicographer is constantly changing, often strained by mutual mistrust or ignorance of each other's capabilities. Time calls for both constructive thinking and *disruptive interdisciplinary collaboration* between lexicographers, on the one hand, and computer scientists who are the embodiment of the new technologies, on the other hand.

A quarter of a century ago, on the eve of the third millennium, Grefenstette (1998) asked whether there would be lexicographers in the year 3000. In the second decade of the new millennium, Rundell (2012) replied that "there will still be lexicographers, but they will no longer do the same job". A decade later, amid an unprecedented technological revolution, it can be added beyond doubt that not only their job but also the outcome of that job will be different. Until now, the vast majority of lexicographers have regarded their field to be that of dictionaries and similar reference works. In the current disruptive period, however, entirely new types of lexicographical products are emerging. As Tarp and

Gouws (2023) have suggested based on an analysis of the historical schism between *glossography* and *dictionography*, this requires a redefinition of the very subject matter of the discipline, which should not only include dictionaries, but also modern glosses and other types of lexicographical data integrated into writing assistants, reading assistants, machine translators, and other tools. It is therefore absolutely essential that today's lexicographers engage in the design of these digital tools. In this sense, their job should not only be to integrate existing lexicographical data types into them (linear innovation), but even more to develop entirely new lexicographical data types in close collaboration with computer scientists and programmers. This is what is meant by *disruptive interdisciplinary collaboration*.

From the perspective of accessibility, Nomdedeu-Rull and Tarp (2024) propose a classification of digital dictionaries into three main types: stand-alone, embedded, and integrated. The *stand-alone dictionary* is the digital equivalent of the traditional printed dictionary that can be consulted independently of the context in which the user's information needs occur. Today, this type of digital dictionary is mostly published on Internet portals or in the form of applications that can be downloaded to laptops, tablets, and mobile phones. The *embedded dictionary* is a variant of the previous one with almost the same features. The only difference is that its search engine is located in another tool, typically a learning application, and only users who have access to it can consult the dictionary. Finally, the *integrated dictionary* is characterised by the fact that the user can consult it by clicking directly on a word or phrase that appears in a digital text. This type of digital dictionary, which is much easier to consult than the other two, is becoming increasingly common. However, although it represents an important innovation, it is also a clear expression of the persistence of *linear thinking* when it comes to putting cutting-edge technology at the service of lexicography. Nomdedeu-Rull and Tarp (2024) therefore add a fourth solution, the modern *gloss*, that is, a single lexicographical data stripped of all the traditional characteristics of a dictionary: book format, lemma, article, macrostructure, and microstructure. When designed to be visualized on demand in digital texts, the short glosses make allowance for a more fluent reading and writing process, which also creates the conditions for *incidental learning* as opposed to *intentional learning*; cf. Tarp (2022a). The latter requires immersion and contemplation, which can only be achieved by providing access to additional lexicographical data, including traditional dictionary articles. This kind of integrated combination of old, new and as yet undeveloped lexicographical data types — for example, in writing or reading assistants — may prove to be highly relevant for language learning in the digital age.

With this perspective, the paper will focus on an ongoing R&D project that the author is conducting together with computer scientists at Ordbogen A/S, a Danish high-tech company specialising in online dictionaries and language services. The objective is to create a completely new digital *Spanish writing assistant* that has many similarities to *Grammarly*, but also differs in some important

aspects, such as being both *monolingual* and *bilingual*, which requires additional functionalities and features. The writing assistant under construction is primarily driven by two interconnected, AI-powered language models called GECToR and BERT, each with its own specific function. These language models are also the main ones used to support *Grammarly*; cf. Omelianchuk et al. (2020).

Section 2 provides a brief overview of existing writing assistants and their functionalities. These observations lead directly to Section 3, where the research project and its aims are presented. Section 4 then discusses the results of a critical analysis of various *Grammarly*-like writing assistants and how they communicate with their users and explain the suggestions for improving language and style. Based on the analysis, a set of principles for good user communication in writing assistants will be formulated. The following section gives examples of how these principles are implemented and what new — and unexpected — challenges this poses. Section 6 summarises the tasks, some of them entirely new, that lexicographers have in developing user-centred writing assistants, and then defines their new role. Finally, the paper offers some perspectives and challenges for contemporary lexicographers.

2. Writing assistants

Today, the writing of first- and second-language texts is increasingly done on laptops, tablets and mobile phones. At the same time, a deterioration in written language can be observed almost everywhere, especially among the younger generations. This undoubtedly calls for the development of new didactic methods that can motivate students in a new way, and it therefore seems logical to start precisely where people write, that is, on the devices mentioned. Instead of being passive writing tools, these devices — by incorporating writing assistants — can be turned into active tools that interact with users and their written language. From this perspective, it is not surprising that the last decade has seen the appearance of a number of digital writing assistants, like the ones presented by Verlinde and Peeters (2012), Granger and Paquot (2015), Tarp et al. (2017), Alonso-Ramos and García-Salido (2019), and Frankenberg-García et al. (2019), among others.

These writing assistants incorporate different technologies, with some of them, like *Grammarly* and *DeepL Write*, relying on AI-powered language models. Depending on their specific philosophy, they have different purposes, features, and functionalities. Some of them, like the *Interactive Language Toolbox*, have a pronounced didactic purpose. A few, such as *ColloCaid* and *HARenEs*, focus on a specific language problem, namely collocations. The probably best-known — such as *Grammarly*, *ProWritingAid*, and *Write Assistant* — provide more general language support for L1 and L2 writing. Most of these general writing assistants exist in only one language (usually English), whereas only a few, like *LanguageTool* and *DeepL Write*, have versions in several languages. Surprisingly,

almost all of them are monolingual, even though they claim to assist both native and non-native writers. As far as the different functionalities of the existing writing assistants are concerned, the following main types can be observed:

- (1) **Detection** of possible problems in the written text
- (2) **Correction** with suggestion of alternative solutions
- (3) **Prediction** of word terminations and next words
- (4) **Transformation** of syntax, style, sentences, etc.
- (5) **Translation** with provision of L2 equivalents
- (6) **Consultation** with look-ups in lexicographical databases

The author of this article is not aware of any writing assistant that has all the above functionalities. Most have three or four. *Write Assistant*, for instance, offers prediction, translation and consultation of lexicographical databases. *Grammarly*, *ProWritingAid*, *LanguageTool*, *DeepL Write*, and *Ginger* offer detection, correction and transformation (rephrasing). In the long run, the ideal writing assistant will probably be the one that combines the functionalities of *Write Assistant* and one of the other four, for instance, *Grammarly* (see Figure 1). Especially the possibility of consulting the meaning of suggested words and using one's own mother tongue seems to be highly relevant when writing in a second language.

WRITE ASSISTANT	GRAMMARLY
	1. DETECTION (possible problems)
	2. CORRECTION (alternative suggestions)
3. PREDICTION (word completions, etc.)	
	4. TRANSFORMATION (syntax, style etc.)
5. TRANSLATION (L2 equivalents)	
6. CONSULTATION (look-ups in databases)	6. CONSULTATION (look-ups in databases)

Figure 1: Complementary functionalities in two writing assistant concepts

As we will see in the next section, lexicographical data are in one way or another involved in all six functionalities, but it is obviously most pronounced when it comes to consultation. In this respect, there are different levels of consultation in existing writing assistants. After a text has been typed or pasted into the tool, the problems detected are usually underlined in different colours depending on their severity and the actions users are expected to perform (see Figure 2).

The traditionel dictionary is the result of a long historical development. It started, both in China and Europe, in ancient times when the old scribes copied manuscript works from earlier periods and inserted glosses to explain obsolete and difficult words. These glosses were later compiled into glossaries which, over the centuries, developed into the modern dictionary. The process had both advantages and disadvantages and saw at least three important inovations.

Figure 2: Possible problems detected by *ProWritingAid*

Users then have to click on the underlined words to find out what the problem might be and what to do about it. The window that pops up will usually offer a brief explanation of the problem as well as an alternative solution, such as a suggestion to change the spelling, inflection, wording, syntax or style. In Figure 3, *LanguageTool* suggests a correction of a spelling mistake appearing in the same text as in Figure 2. The suggested alternative words are undoubtedly *lexicographical data of a new type*, automatically generated by the underlying language model without any direct human intervention. And the same applies to the explanatory notes which are, of course, written by humans, but automatically visualized and inserted into specific contexts without their intervention and knowledge.

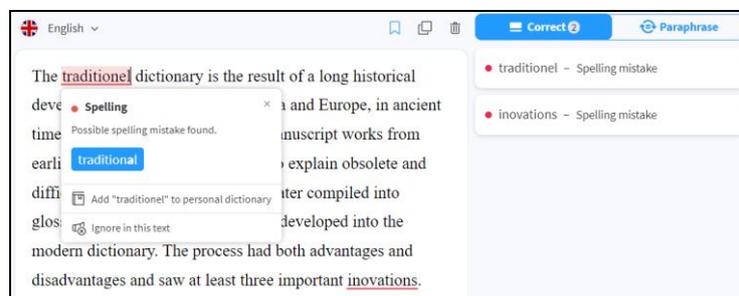


Figure 3: Pop-up window with corrections in *LanguageTool*

In Figure 3, the detected problem is briefly explained with the note *Possible spelling mistake found*. When necessary, some writing assistants also allow their users to access additional explanations. *Grammarly* and *ProWritingAid*, for example, frequently offer this option in the form of a button that users can click on to get more information. This kind of user communication, and the challenges it poses, will be discussed in detail in Section 4.

If the *Paraphrase* button in Figure 3 is activated instead of the *Correct* button, and a sentence is simultaneously marked in the inserted text, *LanguageTool* immediately lists a number of alternative AI-generated sentences with a different wording that the writer can use to vary his or her language and express a

particular idea even more clearly (see Figure 4).

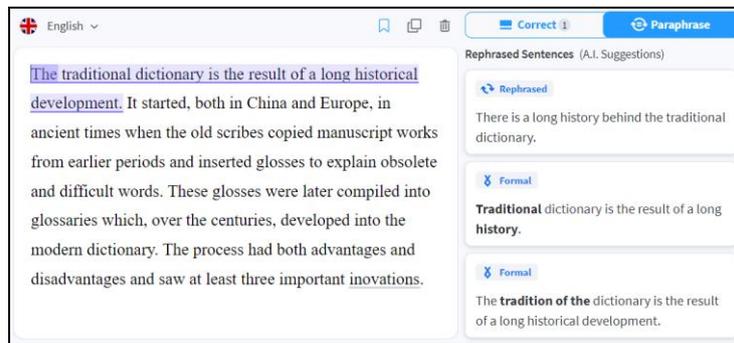


Figure 4: AI-generated rephrasing of sentences in *LanguageTool*

The beta version of *DeepL Write* also offers alternatives to single words and rewording of whole sentences, often listing more than 20 alternative suggestions (see Figure 5). The quality of these suggestions seems quite high, but the design itself, with their presentation in scrolling windows, can make them appear rather unwieldy to users who may need some time to find their way around them. The main problem with the AI-generated rephrasing of whole sentences, however, is that it requires a rather high proficiency level in the language concerned to be able to judge which alternative suggestions are most appropriate in the concrete context, and to understand some of the new words included. As we will briefly discuss in Section 4, this is a serious challenge for *DeepL Write*, *LanguageTool*, and other similar writing assistants that offer rephrasing without further explanation, especially when targeting non-native users.

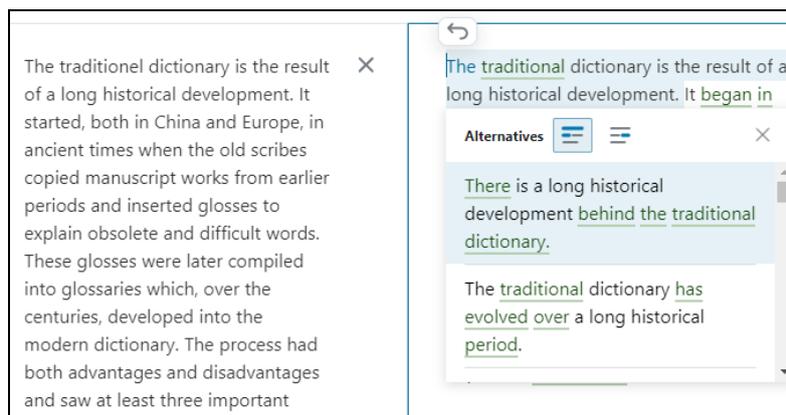


Figure 5: AI-generated rephrasing of sentences in *DeepL Write*

Some writing assistants, like *Grammarly* and *ProWritingAid*, also allow their users to click on some words to get synonyms (see Figure 6). They do not, however, offer any definitions, thus leaving users who do not know the meaning of the suggested alternatives with no recourse. The consultation options are therefore only partial and, as such, particularly problematic for non-native users (this is why this functionality is written in subdued colour in Figure 1). In this way, the most inquisitive writers may feel compelled to consult one of the traditional stand-alone dictionaries (see above), thereby severely disrupting the writing flow. The problem lies neither in cutting-edge technology nor in lexicographical expertise, but in insufficient interdisciplinary collaboration between the two disciplines. This is a tradition inherited from the renowned English thesauri that list synonyms and antonyms, but usually without explaining their meaning, although this could be easily done in the digital environment.

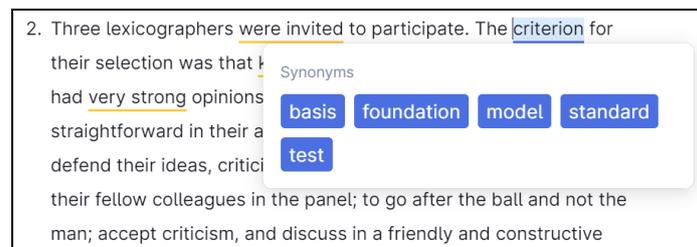


Figure 6: Pop-up window with synonyms suggested by *Grammarly*

This brief review of some of the most interesting and prestigious writing assistants clearly indicates that we are facing huge and rapid changes in the way the written language is produced. In this sense, unprecedented opportunities are opening up to write quickly, correctly and in linguistically varied ways in the desired genre and style. But there are also great risks associated with the new technology if it is not used wisely. If future generations allow themselves to be seduced and become slaves of the new tools, without making a sustained effort to acquire the art of writing, they could end up losing track of what they themselves write.

After discussing the impact of writing assistants on second-language learning, Tarp (2020: 60-61) concludes that "it is above all necessary to *learn to learn by means of technology*" and, therefore, calls for "new didactic methods and convincing arguments in order to motivate language learners". With this in mind, more attention should be paid to the didactic dimension in writing assistants, which existing ones tend to downplay, especially with regard to second-language learning. In addition, a more user-friendly design is necessary to make the writing and correction process as fluid as possible, allowing for not only intentional learning, but also incidental learning. Finally, as we will discuss in Section 4, there is a need for improved user communication, with clear and understandable explanations of problems and suggestions.

3. Description of the project

As mentioned, the overall objective of the current R&D project is to develop a Spanish writing assistant that in many aspects is similar to some of the general writing assistants described above. But there are also some significant differences, inasmuch as the new software aims at being both a writing tool and a learning tool, for both native and non-native users. In addition to a strictly monolingual Spanish version, it will therefore, as a research matter, include bilingual dimensions with four different languages. The first part of the project is carried out exclusively by specialists from Ordbogen A/S. The lexicographers then gradually join in, starting when the writing assistant is trained to detect specific linguistic problems in Spanish, and then continuing when it is taught to communicate with its audience in the respective languages.

The planned writing assistant is based on a set of language models with very different functions, of which the two most important, as already mentioned, are GECToR and BERT. In the following, we will focus on GECToR, which is the one that requires the most lexicographical expertise. This language model is freely available on the internet, but only in a version trained on an English corpus. The computer scientists' first task is therefore to clean it of all English and train it on a Spanish corpus. The training basically consists of special software that breaks the corpus down into its constituent sentences and then automatically introduces between one and five errors into each of them, such as misspelling or misconjugating a word, inserting or deleting a word, and swapping words. In this way, the language model learns to distinguish between right and wrong. Once this training is complete, the language model needs to be fed three different data types to perform optimally: synthetic, semi-synthetic and organic data. This is where lexicography comes in.¹

Synthetic data, a term introduced by computer scientists, are lexicographical data retrieved from a monolingual Spanish database. They are made up of all the lemmata and inflectional forms contained in the database, as well as their respective grammatical categories, i.e. part of speech, gender, number, person, tense and mood, including the more than 60 conjugated forms of Spanish verbs. The aim is to enable the language model to recognise (or dispute) the words and morphological forms that users type and, at the same time, to provide it with a tool for communicating with the lexicographers who prepare the short notes and explanations, as we will see in Section 5. This phase, carried out by a computer scientist, has been completed at the time of writing.

The next phase is to feed the model with *semi-synthetic data*. This is where lexicographers take on a new role. Semi-synthetic data, another computer science term, are sentences framed like the fill-in exercises familiar to many language teachers. It concerns the filling in of words that learners typically confuse, misuse or misconjugate, involving elements such as prepositions, adverbs, pronouns, articles, congruence, tense and mood. However, the main difference with traditional fill-in exercises is that the semi-synthetic data must always be unambiguous, since binary artificial intelligence does not understand ambigu-

ity, which means that there must be only one correct solution. The sentences can either be taken from a corpus or constructed, as they are only for internal use as training material. Figure 7 shows some examples of this material, in this case concerning when — and when not — a verb in the infinitive should be preceded by the preposition *de*. The real challenge, however, is that there is no complete inventory of these types of problem and no one knows how many sentences are needed, so this is a task that is likely to continue to some extent even after the writing assistant is launched. For now, the plan is to prepare 30 correct and 30 incorrect sentences for each problem type, and their final number will probably run into tens of thousands.

Wrong	Right
Sus productos son más fáciles vender este año.	Sus productos son más fáciles de vender este año.
Estos ordenadores son difíciles producir.	Estos ordenadores son difíciles de producir.
Este asunto es fácil aclarar.	Este asunto es fácil de aclarar.
Tengo la intención volver a vivir a Sudáfrica.	Tengo la intención de volver a vivir a Sudáfrica.
La casa es muy bonita y será fácil vender.	La casa es muy bonita y será fácil de vender.
Esta casa es muy bonita y será fácil de venderla.	Esta casa es muy bonita y será fácil venderla.
Resultó imposible de vender la casa.	Resultó imposible vender la casa.
Me es totalmente imposible de pagar ahora.	Me es totalmente imposible pagar ahora.
Es necesario de hacer más investigaciones.	Es necesario hacer más investigaciones.
Nos parece fácil de aclarar el asunto.	Nos parece fácil aclarar el asunto.

Figure 7: Semi-synthetic data used to train the Spanish writing assistant

Organic data are the ones that real users contribute when they start working with the writing assistant. They will therefore only be obtained when the latter has reached a quality that allows it to be tested and then launched. At the time of writing, the writing assistant is available only in its very first version. It can recognise a large number of language structures, but it is also full of bugs that make it useless for its intended users. Further training with semi-synthetic data is therefore required.

In order to fulfil its purpose, a digital writing assistant must not only detect typical errors, but also communicate alternative suggestions to the user in an understandable way, which is a key quality required for such a tool. In addition to providing training material, this is where lexicographers come in with their particular expertise, but in a new role. This is also the main objective of the research project. The idea is to formulate principles for good user communication in the described Spanish writing assistant. Thus, once the quality of the writing assistant has reached a satisfactory level, the next step is to investigate how it can best explain both the problems it detects and the alternative suggestions it automatically generates after appropriate training.

To this end, a critical analysis of existing English writing assistants has identified five different levels of user communication (see Section 4). However, not

all of these levels represent good communication or are relevant to the current project, which will therefore focus on the following two: (1) *Suggestions with a short explanation* and (2) *Suggestions with a supplementary explanation*. As we will see in Section 5, these suggestions must be written from the codes that the language model automatically generates for the respective *problem types*, since they cannot be elaborated for the practically *infinite number of concrete problems* that might arise.

The plan is to write the texts in Spanish. They will then be automatically translated into English, Danish, Italian and Chinese, followed by linguistic revision based on the experience of another recent research project at the Danish company (cf. Tarp 2022b). The reasons for choosing these four languages are that (1) English is the most interesting language in terms of the number of potential users; (2) Danish is relevant to the author's home university and its students; (3) Italian is close to Spanish and may pose specific problems, e.g. with false friends; and (4) Chinese is both linguistically and culturally so distant from Spanish that it may also pose specific challenges.

The next step will be to test the benefits that both native and non-native Spanish speakers derive from the explanatory notes, with a view to possibly modifying them. This will be done using qualitative methods, mainly through observation of individual users, followed by unstructured interviews and analysis of written texts. The initial intention was to answer two main research questions:

- (1) How should the balance between plain language and necessary terminology be struck so that users who cannot be expected to know the grammatical categories can fully benefit from the explanations?
- (2) Do the explanations, even when correctly translated, work for native speakers of all the languages involved, or should they be adapted to each of these languages?

However, a third research question arose during the initial drafting of the explanatory notes (see Section 5):

- (3) To what extent do the automatically generated codes allow lexicographers to write the explanations in the necessary detail so that they are not too superficial, but provide the information that users need to understand the suggested corrections and alternatives?

These three research questions will not be answered conclusively here, as they are beyond the scope of this article. It would require the writing assistant's basic parameters, including the explanations, to have reached a level that would allow it to be tested on real users. However, the answers to these questions are crucial to ensuring good user communication in the final product. They will therefore to some extent inform the discussion in the next two sections and illuminate the direction it takes.

4. User communication in existing writing assistants

This section presents and discusses the results of a critical analysis of various writing assistants that are based on artificial intelligence and lexicographical databases. Most of these tools — such as *Grammarly*, *Ginger*, *LanguageTool*, and *ProWritingAid* — have a premium version, but only the free versions are used here, as the purpose is not to review them but to draw inspiration from them. This also means that writing assistants such as *WhiteSmoke* and *Writefull* are not included, as they are currently only available in premium versions. The analysis is inspired by Norman's (2013) ideas about good communication from machines to humans, and focuses on how these tools communicate with their users and explain the suggestions for improving language and style. In the terminology of this American engineer and cognitive scientist, writing assistants are "every-day things", so their design should allow their users to use them intuitively without further instruction. This implies, among other things, that they should clearly indicate to their users "what actions are possible, what is happening, and what is about to happen" (Norman 2013: 8).

It is worth noting that the writing assistants studied here largely meet this requirement, although, as we shall see, this does not necessarily mean that they are user-friendly in all respects. Their use of *affordances* and *signifiers* that allow users to navigate intuitively in the respective tools seems generally very professional, while there are some challenges related to the *feedback* provided, i.e. how the suggestions are explained or not explained. (See Tarp and Gouws (2020: 6) for a definition of these industrial designer terms.)

According to Norman's human-centred design philosophy, feedback must be given after every action in order to confirm the action and communicate its results. As such, it is an important form of communication that should be both immediate and informative. In the case of writing assistants, this means that they should not only suggest alternative solutions, but also explain why they are suggesting them.



Figure 8: Wrong word in WhatsApp message after automatic correction

There are actually some widely used writing assistants that completely ignore Norman's usability advice. For example, if the autocorrect option is enabled on a smartphone, messages sent too quickly and without enough attention can easily go awry. Figure 8 shows a typical case of a WhatsApp message. A Spanish expatriate, back in her country after two years of Covid shutdown, expresses her joy at finally enjoying a *tinto de verano* (summer red wine). But when she accidentally hits the wrong key, the built-in writing assistant automatically corrects it to the nonsensical *tinto de veranda* (veranda red wine) without her consent. Such an automatic correction without the user's explicit consent is clearly the result of a bad model for the tool's communication with its users.

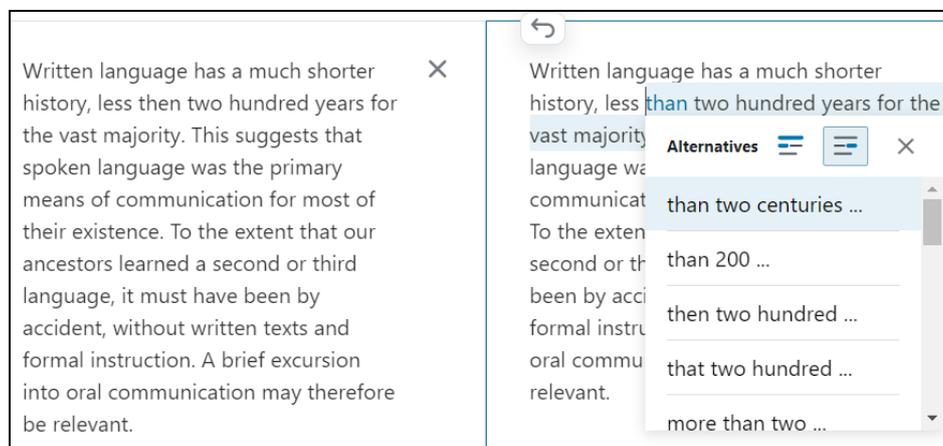


Figure 9: Alternatives to confused word in *DeepL Write*²

DeepL Write is an example of a writing assistant that gives its users a choice between several alternatives, but which, at least in its beta version, does not support these suggestions with explanations to make the decision easier for the users (see Figure 9). This can cause serious problems. The design used makes it impossible for users to see whether the proposed alternative is a correction of a spelling error or a confused word, or a suggestion to use a better alternative, or simply to vary the language. Users will find that the word they originally typed often appears as number two or three in the list of alternatives. And even if the word is clearly misspelt or confused, they will sometimes also find it further down the list, as in Figure 9. This seems to be an example where artificial intelligence has been given too free a rein, and therefore needs to be further domesticated to become user-friendly. As a minimum, one could have expected *DeepL Write* to explain to its users what type of problem it had detected and provided alternatives for. This is at least what other writing assistants do, as we will see below.

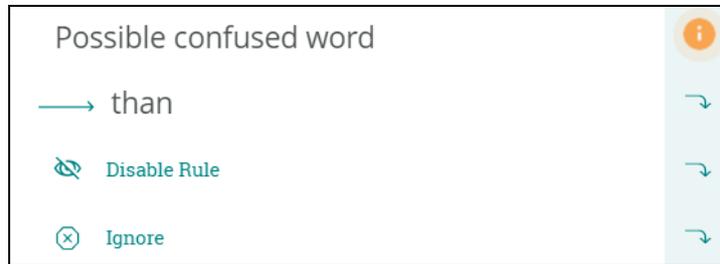


Figure 10: Suggestion with short explanation and reference in *ProWritingAid*

The same text as in Figure 9 was pasted into *Grammarly*, *Ginger*, *LanguageTool* and *ProWritingAid* to see how they dealt with the above challenge. Unsurprisingly, they all offer *than* as an alternative. But beyond that, they describe the problem remarkably differently and with varying degrees of completeness. *ProWritingAid*, for example, describes it as a "possible confused word" (Figure 10), while *Grammarly* classifies it as a spelling problem and a word that "may be used incorrectly" (see Figure 11). In addition, both refer their users to a supplementary explanation (see Figures 13 and 14).

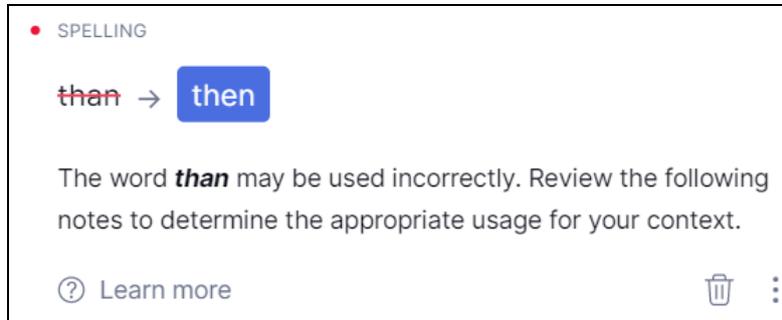


Figure 11: Suggestion with short explanation and reference in *Grammarly*

As can be seen in Figure 12, *LanguageTool* describes *then* as "not likely" and "less likely than" *than* in the specific context. At the same time, it includes two parenthetical definitions of the words as "at that time, later on" and "as in: greater than", respectively. The short definitions themselves are easy to understand, but the condensed definition style known from traditional print dictionaries is likely to be an unnecessary challenge for many users. It simply takes longer time to read this type of text. As such, it hardly represents a model of good user communication.

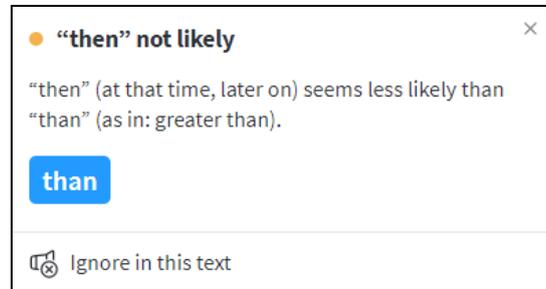


Figure 12: Suggestion window with short explanation in *LanguageTool*

Ginger, for its part, classifies *then* in the concrete context as a "misused word" and explains the incorrect usage as a "confusion between words" with similar orthography or pronunciation (see Figure 13). It also adds two short lexicographical definitions of the words, using grammatical terms such as *conjunction*, *sentence connector* and *comparative*, with which part of the target audience cannot be expected to be familiar. Besides, the whole default text seems too voluminous when it is just a matter of quickly explaining to users why they should replace *then* with *than*. It is a case of *relative data overload*, according to the classification proposed by Gouws and Tarp (2017: 408). A much better solution would therefore be to hide the definitions and allow users to visualize them when needed. As we will see, this is the option chosen by both *Grammarly* and *ProWritingAid*.

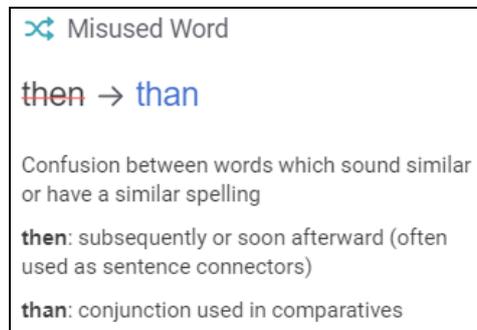


Figure 13: Suggestion window with explanation in *Ginger*

Figure 14 shows the hidden text that *Grammarly* displays when users click on the *Learn more* button in Figure 11. It contains definitions of the two words and examples of their correct and incorrect use. This seems to be all a user needs to get a general idea of the meaning and usage of the two words. It is also worth noting the use of grammatical terms such as *conjunction*, *preposition* and *adverb*. Contrary to Figure 13, the use of such specialized terms makes sense here, as it

can be assumed that the users who click through are also the most motivated and knowledge-seeking. In general, the text in Figure 13 reflects an appropriate way of providing additional explanations in writing assistants. It represents good user communication.

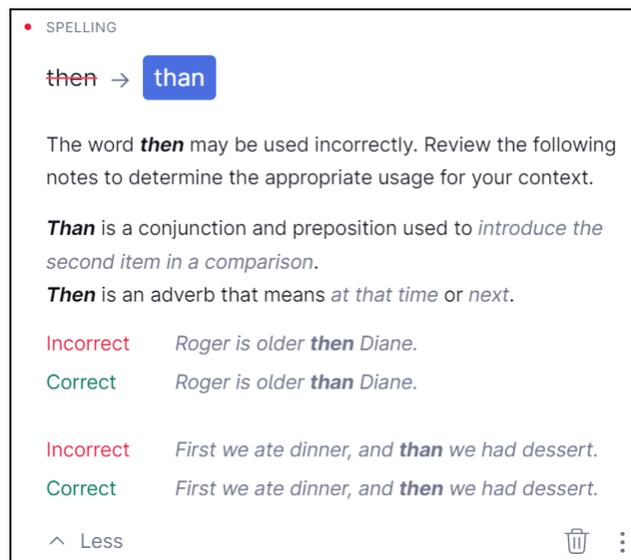


Figure 14: Window with additional explanation in *Grammarly*

Unfortunately, the same cannot be said for *ProWritingAid*. When the users, after being informed that *than* is a "possible confused word", click on the icon for more information (Figure 10), a window pops up with the text shown in Figure 15. This text only deals with the general problem of confused words and does not mention the specific problem of *than* and *then* with a single word. Even if the most inquisitive users then click on *Read Full Article*, no more details about the two words appear, just a long list of other confused word pairs. The users themselves are likely to be left confused and less motivated to perform similar searches. This is clearly not an example of good user-centred communication from the writing assistant to its target audience, but rather the opposite. It is also a bit strange. A year and a half earlier, when users of *ProWritingAid* clicked on the icon for more information, they actually activated a window that explained the use of the two words in much the same way as *Grammarly*, and also included separate links to two definitions, a full article, and even a video. It is difficult to see the logic of replacing an acceptable text with one that is unspecific and too general. The merit of *Grammarly's* supplementary comment in Figure 14 is precisely that it provides a specific explanation for a specific problem, exactly what users need in a concrete writing situation.

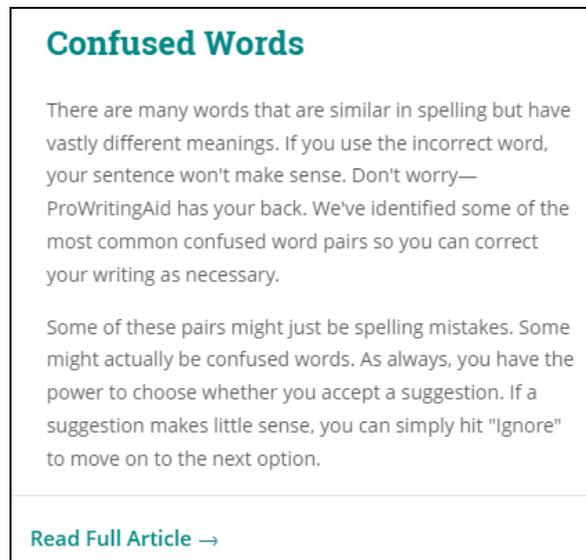


Figure 15: Window with additional explanation in *ProWritingAid*

The screenshots and examples discussed above have been chosen because they are representative of the main conclusions that can be drawn from the analysis of the different writing tools. This does not mean, of course, that there are no other interesting lessons to be learned from them. The first important finding is that the communication in the existing English writing assistants can be divided into five categories according to the degree of elaboration:

- (1) auto-correction without user's acceptance (Figure 8)
- (2) suggestion without explanation (Figure 9)
- (3) suggestion with short explanation (Figures 10, 11, 12)
- (4) suggestion with supplementary explanation (Figures 14, 15)
- (5) extended explanation (after clicking through in Figure 15)

The first two categories clearly indicate poor communication, which invites mistakes and does little to support learning, and should therefore be avoided wherever possible. The fifth category consists of links to both internal and external pages. Such references may be appropriate when explaining a complex grammatical challenge, such as the use of the subjunctive in Spanish. However, it is important that references are used judiciously and do not become references for reference's sake, but are only offered when they are relevant to the use of the particular writing assistant and the associated learning process. The current research project therefore focuses on the third and fourth categories, which are directly related to the new understanding of incidental and intentional learning

in the digital context (cf. Tarp 2022a). These two categories or levels of communication do not in themselves constitute good communication. The latter also depends on a number of other factors, such as the overall design, the language used, the style and terminology, the size of the texts and the use of definitions.

The design of existing writing assistants is generally intuitive and also in line with the recommendation of Tarp and Gouws (2019, 2020) to assist users directly in the context where they have problems, without requiring them to have special reference skills. It is therefore not discussed in detail here. By contrast, language can be a challenge for many users. All the writing assistants described above use English as the sole language of explanation, at least for the time being. This is of course best for native English speakers, but for non-native speakers, including learners who are also part of the target group, it can be very challenging, especially if they are at a beginner or intermediate level. Therefore, if these writing assistants are serious about addressing non-native English speakers, they should consider explaining themselves in the native language of these users.

In terms of style, it is noticeable that most writing assistants take some care in formulating the explanations to support their alternative suggestions. This is reflected in the use of words and expressions such as *may be*, *possible*, *it seems*, *it appears*, etc. This approach is understandable, as the technology is not yet able to identify problems and suggest alternatives with 100% accuracy. For example, in Figure 2, *ProLanguageAid* recommends correcting both *manuscript* and *glosses*, even though these words are perfectly valid in the given context. Still, these reservations can be annoying and are probably not always necessary. So it is refreshing that *Ginger* dares to say things directly, as shown in Figures 13 and 16. Whatever the risk of making mistakes, this points to a more user-friendly product, for which further research and refinement is needed. Since no language model is 100% perfect, the key question in any case is whether a few errors can be accepted and, if so, where the error threshold should be.

Regarding the extent of the explanations and the use of terminology, it has already been argued that the default texts should not be too long and that specialized terms should be avoided as far as possible and preferably used only in the supplementary comments. The latter, as mentioned above, is an aspect of one of the research questions to be answered in the current project.

The *main problem* — and therefore the *main challenge* — that the analysis of the different writing assistants revealed was the lack of definitions for the suggested alternative words and expressions, as users, especially non-native English speakers, cannot always be expected to know the meaning of all of them. The risk is that users simply click on the suggested words and have no idea what they are actually typing. This could have unpleasant consequences for them and certainly does not help them to develop their own writing. It is therefore gratifying that *Ginger*, as an inspiring exception, offers such straightforward definitions, although it could be done more elegantly (see Figure 16).



Figure 16: Explanation and definition of suggested alternatives in *Ginger*

The challenge goes far beyond individual words and phrases that users may be unfamiliar with, as it also involves rephrasing entire sentences. This is a problem similar to the new AI-powered chatbots that students and others are relying on to write essays on topics they are not well versed in, and where they often have no deeper understanding of what the chatbot is writing. With good reason, Chomsky (2023) characterizes the use of these tools as "a way of avoiding learning". In the case of writing assistants, it will probably take some time to develop software that can explain the meaning of the spontaneously rephrased sentences generated by artificial intelligence. This is certainly a challenge that must be met if a further decline in written language is to be avoided. An interim solution could be to develop software that can identify the concrete meaning of individual words in a given context, based on the visionary ideas of Bothma and Gouws (2022). It might also be relevant to use constructive hints to motivate understanding. When writing in a second language, it could even be an option to provide automatic translation into the user's native language. In any case, the whole issue is an urgent topic for research.

5. Some reflections on writing explanations

At the time of writing, the implementation of the above ideas has not yet begun. However, some testing has been done and is discussed here. So far, the GECToR language model has generated over 8,000 internal codes as a prerequisite for writing short explanations to support the alternative suggestions. This number is likely to increase to 30-40,000 as the model is trained further.

Most of the problems currently identified relate to replacing single words, where writing explanations would be relatively straightforward, although time-consuming. However, there are also a large number of cases involving the conjugation of Spanish verbs. In this respect, some of the generated codes are elaborate and easy to handle, while others are very short and uninformative. Below are two examples to illustrate the challenges.

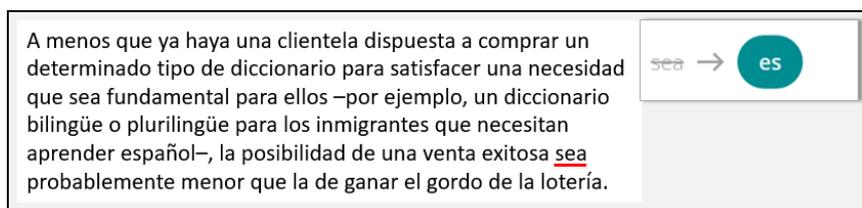


Figure 17: Writing assistant with suggestion for correction

Figure 17 shows a Spanish text pasted into the new writing assistant. One of the verbs was then incorrectly changed from the indicative to the subjunctive and the software immediately reacted by underlining the error and suggesting the correct conjugation. The correction was linked to the following internal code, which the GECToR language model had automatically generated after training:

— \$TRANSFORM_INFLEXION_verb_verb-indicative-present-third-person

The first part in capital letters indicates what the main problem is, i.e. that the inflection needs to be changed. The second part in lower case is based on synthetic data from a lexicographical database and indicates a verb in the third person present indicative as an alternative. Using a specially designed interface, the lexicographer can now write an explanation that links directly to the internal code. This explanation, translated into English in Figure 18, consists of a macro in which the two words in bold are inserted. From now on, whenever there is a typing problem involving the same code, the writing assistant will display this text and insert the two concrete words involved in the particular case.

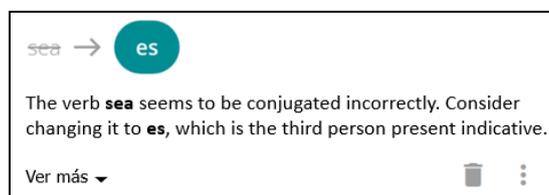


Figure 18: Window with short explanation of the suggestion to change *sea* to *es*

In Figure 19, another Spanish verb was incorrectly changed from the indicative to the subjunctive and immediately marked by the language model, i.e. a problem that is grammatically identical to the one above.

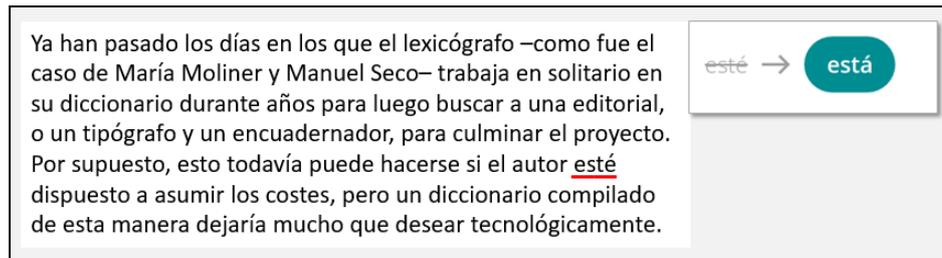


Figure 19: Writing assistant with suggestion for correction

In this case, GECToR associates the problem with the following internal code:

— TRANSFORM_INFLEXION_verb_verb-i

This difference in the treatment of two apparently identical problems is due to the fact that Spanish verbs have three different moods: indicative, subjunctive and imperative. In Figure 17, the inflected form *es* is unique and can only be *indicative, present, singular, third person*. GECToR therefore identifies it unambiguously. By contrast, the inflected form *está* in Figure 19 is ambiguous, as it can be both *indicative, present, singular, third person* and *imperative, singular, second person*. Consequently, GECToR can only identify it as *verb-i* (the first letter in both *indicative* and *imperative*), as it starts parsing from the left.

In fact, it is the same kind of problem that explains why GECToR only indicates *verb* and nothing else for the two incorrectly conjugated words underlined in Figures 17 and 19. The reason is that both *sea* and *esté* are inflected forms used in both the imperative and the subjunctive of the respective verbs (*ser* and *estar*). Therefore, when the language model starts parsing from the left, it cannot go further than *verb* because the two moods begin with different letters.

This problem was not foreseen, but it is logical given the conjugation pattern of Spanish verbs. The challenge is that unambiguity is required to write informative and correct explanations. It remains to be seen how to solve this problem. One possibility is to feed the language model with synthetic data of a syntactic nature, but this implies a redesign of the lexicographical databases used, something already pointed out by Fuertes-Olivera and Tarp (2020) in their discussion of *Write Assistant*. In any case, it will require interdisciplinary collaboration.

As can be seen in Figure 18, the explanation includes the term *indicative*, which in some ways goes against what was previously recommended. The rea-

son for this is that the use of the indicative and the subjunctive is so central to Spanish verbs that it is difficult to say anything meaningful without using these terms. Whether this is a wise idea, or whether another clever solution can be found, will be decided during the tests that make up the next phase of the research project. The aim is to answer the three research questions presented in Section 3, including the one in question here.

6. New role of lexicography

The above discussion shows that lexicographical databases have an important role to play in the new writing assistants, although they need to be adapted to new requirements in some respects. But it also shows that there are a number of new tasks for which lexicographers are well prepared:

- Training of an AI-based language model to support the writing assistant;
- Designing a set of principles for good communication between the writing assistant and its users;
- Preparing a new type of lexicographical database, which is problem-based rather than lemma-based, and contains small texts explaining the alternative suggestions made by the writing assistant.

It could be argued that these three tasks are not lexicography, because they are not what professionals in the discipline usually do. This is true. But it is worth remembering that these *modern glosses* are very similar to the *traditional glosses* that Greek scribes — i.e. the first European lexicographers — inserted into manuscript copies in the Classical Period over two thousand years ago (cf. Hanks 2013). Nor should we forget Rundell's (2012) prediction that the lexicographers of the year 3000 will not be doing exactly the same as their colleagues of today. If the latter do not prepare themselves and move into new areas, they will be left behind by rapid technological development. The design of high-quality writing assistants is a new area that lexicographers will definitely have to cover, although it implies that the centrality of the lemma category in lexicographical work will have to be complemented by the centrality of another data category.

7. Perspectives

Artificial intelligence is here to stay. Lexicographers will have to prepare themselves to interact with it in one way or another, not only in terms of writing assistants, but also in relation to other areas where artificial intelligence is encroaching on their discipline. Exciting challenges lie ahead. There are unprecedented opportunities to provide personalised lexicographical services to users in need. The list is long: writing assistants, reading assistants, learning apps, machine translators, etc. But, as we have seen, there are also major risks that need to be addressed. As written in Spanish in Figure 19, lexicographers should

not hide at home or even in traditional publishing houses. They must proactively seek out the experts who embody the latest technology. They should do this on an equal footing and with confidence, knowing that they have much to contribute to keeping technology in check and creating new techno-lexicographical masterpieces.

This positive, optimistic, and inquisitive look into the future, without fear of making mistakes, is the best way to celebrate Rufus H. Gouws's — to date — 43 years of successful academic work.

Acknowledgments

Thanks are due to the Aarhus University Research Foundation for funding a six-month sabbatical to conduct this research project.

Special thanks are due to Ordbogen A/S for incorporating me into its R&D team and to Programmer and Web Developer Henrik Hoffmann for his constant enthusiasm, inspiration and technical support.

Special thanks are also due to Prof. Pedro Fuertes-Olivera, Valladolid University, for his constructive comments.

Endnotes

1. After the article was submitted, ChatGPT was introduced into the project as a production tool, resulting in both a modified project plan and new tasks for the lexicographers. See Huete-García and Tarp (2024) and Tarp and Nomdedeu-Rull (2024) for further details.
2. While preparing this article (January–February 2023), *DeepL Write* was launched in a beta version that is discussed here. Like similar tools, it is expected to continuously improve with new features and may therefore be modified in some aspects by the time the article is published.

References

A. Writing assistants

ColloCaid: <http://www.collocaid.uk/>

DeepL Write (beta version): <https://www.deepl.com/write>

Ginger: <https://www.gingersoftware.com/>

Grammarly: <https://www.grammarly.com/>

HARenEs: <http://harenes.taln.upf.edu/CakeHARenEs/check>

LanguageTool: <https://languagetool.org>

ProWritingAid: <https://prowritingaid.com/>

WhiteSmoke: <https://www.whitesmoke.com/>

Write Assistant: <https://writeassistant.com/>

Writefull: <https://writefullapp.com/>

B. Other literature

- Alonso-Ramos, M. and M. García-Salido.** 2019. Testing the Use of a Collocation Retrieval Tool Without Prior Training by Learners of Spanish. *International Journal of Lexicography* 32(4): 480-497.
- Bothma, T.J.D. and R.H. Gouws.** 2022. Information Needs and Contextualization in the Consultation Process of Dictionaries that are Linked to e-Texts. *Lexikos* 32(2): 53-81.
- Chomsky, N.** 2023. Noam Chomsky on ChatGPT: It's "Basically High-Tech Plagiarism" and "a Way of Avoiding Learning". *Open Culture*, February 10, 2023
- Frankenberg-García, A., R. Lew, J.C. Roberts, G.P. Rees and N. Sharma.** 2019. Developing a Writing Assistant to Help EAP Writers with Collocations in Real Time. *ReCALL* 31(1): 23-39.
- Fuertes-Olivera, P.A. and S. Tarp.** 2020. A Window to the Future: Proposal for a Lexicography-assisted Writing Assistant. *Lexicographica* 36: 257-286.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Granger, S. and M. Paquot.** 2015. Electronic Lexicography Goes Local: Design and Structures of a Needs-driven Online Academic Writing Aid. *Lexicographica* 31(1): 118-141.
- Grefenstette, G.** 1998. The Future of Linguistics and Lexicographers: Will There Be Lexicographers in the Year 3000? Fontenelle, T., P. Hilgsmann, A. Michiels, A. Moulin and S. Theissen (Eds.). 1998. *Proceedings of the Eighth EURALEX International Congress in Liège, Belgium*: 25-41. Liège: English and Dutch Departments, University of Liège.
- Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Allan, K. (Ed.). 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.
- Huete-García, Á. and S. Tarp.** 2024. Training an AI-based Writing Assistant for Spanish Learners: The Usefulness of Chatbots and the Indispensability of Human-assisted Intelligence. *Lexikos* 34. (to appear)
- Nomdedeu-Rull, A. and S. Tarp.** 2024. *Introducción a la lexicografía en español: funciones y aplicaciones*. London: Routledge. (to appear)
- Norman, D.** 2013. *The Design of Everyday Things*. New York: Basic Books.
- Omelianchuk, K., V. Atrasevych, A. Chernodub and O. Skurzhanskyi.** 2020. GECToR — Grammatical Error Correction: Tag, Not Rewrite. Burstein, J., E. Kochmar, C. Leacock, N. Madhani, I. Pilán, H. Yannakoudakis and T. Zesch (Eds.). 2020. *Proceedings of the 15th Workshop on Innovative Use of NLP for Building Educational Applications*: 163-170. Seattle: Association for Computational Linguistics.
- Rundell, M.** 2012. The Road to Automated Lexicography: An Editor's Viewpoint. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*, 15-30. Oxford: Oxford University Press.
- Tarp, S.** 2019. Connecting the Dots: Tradition and Disruption in Lexicography. *Lexikos* 29: 224-249.
- Tarp, S.** 2020. Integrated Writing Assistants and their Possible Consequences for Foreign-Language Writing and Learning. Bocanegra-Valle, A. (Ed.). 2020. *Applied Linguistics and Knowledge Transfer: Employability, Internationalization and Social Challenges*: 53-76. Bern: Peter Lang.
- Tarp, S.** 2022a. A Lexicographical Perspective to Intentional and Incidental Learning: Approaching an Old Question from a New Angle. *Lexikos* 32(2): 203-222.
- Tarp, S.** 2022b. Turning Bilingual Lexicography Upside Down: Improving Quality and Productivity with New Methods and Technology. *Lexikos* 32: 66-87.

- Tarp, S. and R.H. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.
- Tarp, S. and R.H. Gouws.** 2020. Reference Skills or Human-Centered Design: Towards a New Lexicographical Culture. *Lexikos* 30: 470-498.
- Tarp, S. and R.H. Gouws.** 2023. A Necessary Redefinition of Lexicography in the Digital Age: Glossography, Dictionography and Implications for the Future. *Lexikos* 33: 425-427.
- Tarp, S. and A. Nomdedeu-Rull.** 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. (to appear)
- Tarp, S., K. Fisker and P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.
- Verlinde, S. and G. Peeters.** 2012. Data Access Revisited: The Interactive Language Toolbox. Granger, S. and M. Paquot (Eds.). 2012. *Electronic Lexicography*: 147-162. Oxford: Oxford University Press.