

# Connecting the Dots: Tradition and Disruption in Lexicography\*

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

---

**Abstract:** This article botanizes in the history of lexicography trying to connect the dots and get a deeper understanding of what is happening to the discipline in the framework of the Fourth Industrial Revolution. The objective is to suggest possible ways out of the present deadlock. History shows that a sudden change of the technological base, like the one we are now experiencing, suggests a total revolution of the discipline in all its major dimensions. In order to be successful, such a revolution requires a mental break with past traditions and habits. As a matter of example, the article focusses on a series of bilingual writing assistants developed by the Danish company Ordbogen A/S and the new challenges posed to lexicography by these and similar tools. It argues that these challenges cannot be solved by means of traditional user research which is retrospective as it unfolds in the framework of an old paradigm. As an alternative, and without excluding other types of user research, the article recommends disruptive thinking by means of brainstorm, immersion, and contemplation and provides some examples on how to proceed. Finally, it problematizes the incipient competition between human and artificial lexicographers and gives a brief account of a possible future redistribution of tasks.

**Keywords:** FOURTH INDUSTRIAL REVOLUTION, DISRUPTIVE INNOVATION, ARTIFICIAL INTELLIGENCE, BUSINESS MODEL, WRITING ASSISTANTS, ARTICLE STRUCTURE, USER RESEARCH, TEST-DRIVEN DEVELOPMENT, HUMAN LEXICOGRAPHER, ARTIFICIAL LEXICOGRAPHER

**Opsomming: Die lê van verbande: Tradisie en ontwingting in die leksikografie.** In hierdie artikel word die geskiedenis van die leksikografie geanaliseer in 'n poging om verbande te lê en om 'n beter begrip te verkry van wat aan die gebeur is met hierdie dissipline binne die raamwerk van die Vierde Industriële Rewolusie. Die doelwit is die voorstelling van moontlike metodes om van die bestaande dooiepunt te ontkom. Uit die geskiedenis is dit duidelik dat 'n skielike verandering van die tegnologiese basis, soos die verandering wat ons nou ervaar, 'n totale omwenteling van die dissipline in al sy hoofdimensies suggereer. Om suksesvol te kan wees,

---

\* The main ideas of this article were presented at the 24th Annual International Conference of the African Association for Lexicography (AFRILEX), hosted by the Department of Language and Literature Studies, University of Namibia, Windhoek, Namibia, 26–29 June 2019.

vereis so 'n omwenteling 'n breuk met tradisies en gewoontes van die verlede. As voorbeeld fokus die artikel op 'n reeks tweetalige skryfhulpmiddels wat deur die Deense maatskappy Ordbogen A/S ontwikkel is en op die nuwe uitdagings wat hierdie en soortgelyke hulpmiddels vir die leksikografie inhou. Dit voer aan dat hierdie uitdagings nie deur middel van tradisionele gebruikersnavorsing wat op 'n retrospektiewe manier binne die raamwerk van 'n ou paradigma ontvou, opgelos kan word nie. As alternatief, en sonder om ander tipes navorsing uit te sluit, word ontwrigtende denke in hierdie artikel aanbeveel deur te dinkskrum, jou te verdiep en te bespiegel, en enkele voorbeelde van hoe om te werk te gaan, word verskaf. Ten slotte word die aanvanklike wedywering tussen menslike en kunsmatige leksikograwe uiteengesit en 'n kort verslag van 'n moontlike toekomstige herverdeling van take word gegee.

**Slutelwoorde:** VIERDE INDUSTRIËLE REVOLUSIE, ONTWRIGHTENDE INNOVASIE, KUNSMATIGE INTELLIGENSIE, SAKEMODEL, SKRYFHULPMIDDELS, ARTIKELSTRUKTUUR, GEBRUIKERSNAVORSING, TOETSGEDREWE ONTWIKKELING, MENSLIKE LEKSIKOGRAAF, KUNSMATIGE LEKSIKOGRAAF

Creativity is just connecting things. When you ask creative people how they did something, they feel a little guilty because they didn't really *do* it, they just *saw* something. It seemed obvious to them after a while. That's because they were able to connect experiences they've had and synthesize new things. And the reason they were able to do that was that they've had more experiences or they have thought more about their experiences than other people.

Unfortunately, that's too rare a commodity. A lot of people in our industry haven't had very diverse experiences. So they don't have enough dots to connect, and they end up with very linear solutions without a broad perspective on the problem. The broader one's understanding of the human experience, the better design we will have.

(Interview with Steve Jobs in *Wired*, Wolf (1996)).

## 0. Introduction

The history of lexicography encompasses long periods with the accumulation of small and gradual changes within an existing paradigm as well as relatively short periods with abrupt and profound changes within a new paradigm. Today, we are witnessing such a paradigm shift described as a *Cambrian Explosion* by Fuertes-Olivera (2016). The phenomenon is characterized by a turmoil of both old forms that are surviving (printed dictionaries) and new forms that are constantly appearing and disappearing (PDF, CD-ROM, DVD, apps, handheld, web-based dictionaries, etc.). The present turmoil resembles Darwin's "survival of the fittest", especially if other digital reference resources competing with dictionaries are included; see e.g. Frankenberg-García (2018) and Alonso-Ramos and García-Salido (2019). In the final analysis, the turmoil is caused by the introduction and more or less successful application of disruptive technologies which, for their part, are continuously developing and improving.

Current lexicography is developing in the framework of the so-called

*Fourth Industrial Revolution*. According to Schwab (2015), who coined the term, this new phenomenon "is characterized by a fusion of technologies that is blurring the lines between the physical, digital, and biological spheres". The author, who is also the founder and executive chairman of the World Economic Forum, is emphatic that there is no historical precedent to the speed, scope, and complexity of the current breakthroughs:

When compared with previous industrial revolutions, the Fourth is evolving at an exponential rather than a linear pace. Moreover, it is disrupting almost every industry in every country. And the breadth and depth of these changes herald the transformation of entire systems of production, management, and governance. The possibilities of billions of people connected by mobile devices, with unprecedented processing power, storage capacity, and access to knowledge, are unlimited. (Schwab 2015)

As examples of emerging technology breakthroughs that may multiply these possibilities, Schwab lists "artificial intelligence, robotics, the Internet of Things, autonomous vehicles, 3-D printing, nanotechnology, biotechnology, materials science, energy storage, and quantum computing".

Whether one likes it or not, lexicography has to navigate and find its ways in this disruptive explosion of technological innovations. The present complexity of things does not allow lexicographers to hide their heads in the sand. To a large extent, it is a battle of life and death. It is therefore urgent to take action before the discipline gets caught in deep slumber.

This article will look at the turmoil and the crisis that has crept into lexicography during the past years as a result of the new technological breakthroughs. The crisis will be put into a historical perspective in order to get a more profound understanding of its complexity and main characteristics. A short excursion will be made into history with a special focus on the birth of European lexicography 2 500 years ago as well as the long-term consequences of the irruption of printing technology in European lexicography more than 500 years ago.

The current business model which is becoming increasingly obsolete will then be challenged and an alternative model outlined. The objective is to sustain a transformed lexicography that is fully prepared to be part of the new Industrial Revolution. To that end, the article will discuss what it takes to inject new blood into the lexicographical veins. As a matter of example, it will reflect on the current lexicographical challenges posed by the premium development of an integrated tool that provides instantaneous assistance to second-language writing. In this connection, the timely relevance of most current user research will be disputed and new ways of getting closer to the users and their real needs will be recommended.

Finally, the concept of an artificial lexicographer will be introduced as opposed to the traditional human lexicographer and a possible future distribution of tasks and responsibilities between the two of them will be outlined.

## 1. Lexicography in crisis

Although there are still regions, like Southern Africa, that continue living in the happy days of printed dictionaries, the general world tendency is now the propagation of web-based dictionaries. The onlinezation poses new challenges to lexicography. More than ten years ago, Rundell (2007) raised what he called the "hardest question", namely "how to fund all this development":

Electronic versions of MLDs have been around for 15 years or so, but none have yet made any money (and they cost a lot to develop). New revenue models need to emerge, and these could include advertising. (Rundell 2007: 50)

Since then, things have only gone downhill. Most publishers of general dictionaries in Western Europe and North America do still not make money from their digital products. MacMillan, for instance, has completely stopped publishing printed dictionaries; cf. Rundell (2014). Instead, it opted for an ad-financed model for its digital products. This solution, which rather looks like a stopgap, entails another unwanted complication, namely the risk of lexicographical data overload; cf. Gouws and Tarp (2017).

The uncomfortable fact is that most users of online dictionaries expect them to be free. They are not ready to pay for this service unless it offers some highly specialized dictionaries that are indispensable for their jobs or studies. In Europe, the unpleasant result is that publishing houses earning money from their online dictionaries can be counted on the fingers of one hand. The few who prosper are mainly lexicographical newcomers who base themselves on a different business model. Most traditional publishing houses seem to be incapable of adapting to the new market conditions. This tendency was also documented by Simonsen (2017) who conducted research into the Danish market. Many of these publishers of high-quality dictionaries have now been forced to close down their business due to dramatically reduced sales, among them famous ones like Longman, Harrap, and Langenscheidt.

Simultaneously, a large number of free dictionaries of dubious quality flourish on the Internet like mushrooms after the rain. This has led to an awkward paradox. On the one hand, modern information-age users need high-quality dictionaries providing quick and reliable information to solve their complex problems and needs. On the other hand, a growing number of these users opt for free-access dictionaries of dubious quality frequently obtaining inadequate and even incorrect information which only adds to their problems.

In some countries, like Spain, no new big general dictionary, either monolingual or bilingual, has seen the light for more than 15 years. Established publishing houses simply do not have money to fund new projects. This reflects a profound crisis which in a certain sense could be described as a crisis of existence for lexicography as we have known it until now. The consequences of all this are potentially disastrous for a society where communication and information is increasingly important.

Rundell's (2007) big challenge therefore remains: Find the money! It is,

however, important not to forget that the lack of an appropriate business model is only the trigger of the current crisis within lexicography. The fundamental cause is the shift of paradigm and the introduction of new disruptive technologies. Hence, the development of a new lexicographical business model must take its point of departure in a profound knowledge of these innovations and their impact on the millennial discipline.

## 2. A historical vision

Just as in other aspects of life, knowledge of history can prove very useful if one wants to understand the breadth and depth of the current crisis within lexicography. History never repeats itself in a completely identical way, but it nevertheless displays some regularities which, once discovered, can be a great inspiration to understand the present. This is also valid for the relationship between lexicography and technology, a subject that has been treated by many scholars, among them De Schryver (2003), Hanks (2010, 2013), Rundell and Kilgarriff (2011), Nielsen (2013) and Fuertes-Olivera et al. (2018).

Over the years, technology has strongly shaped the development of the five main phases in the practical lexicographical process:

- (1) *determination* of data types to be offered to the future users;
- (2) *retrieval* of raw data from the empirical sources;
- (3) *preparation* of lexicographical data;
- (4) *storing* of data; and
- (5) *presentation* of the lexicographical data in the final product.

Each of these phases has witnessed seismic changes and considerable improvements over the years when new technologies have been introduced. The particular way in which European lexicography was born is inconceivable without the invention of the pen and parchment; cf. Tarp and Gouws (2019). The exponential growth of dictionary output during the past five centuries would not have been possible without the continuous development of the printing and bookbinding technology; cf. Hanks (2010, 2013). The improved quality of lexicographical data during the past fifty years is inseparable from the introduction of digital corpora and the Internet as empirical sources; cf. Hanks (2012), Tarp and Fuertes-Olivera (2016). The lexicographers' present working methods are unthinkable without computers, databases, and e-clouds, etc.

Fuertes-Olivera et al. (2018: 155) have summarized the long-term consequences of the printing technology for lexicography:

Summarily, it can be established that the introduction of the printing technology implied big changes in the production and presentation of the lexicographical product; the empirical basis with the increased use of index cards based on written texts; the design of the dictionary articles with the incorporation of new data categories; the distribution and use of dictionaries; the number of users; the topics treated in dictionaries; and the research areas of scholarly interest. To this

can be added the growing social prestige of lexicographers, some of whom became nationally and internationally famous personalities, as well as the fact that lexicography turned into an increasingly successful business.

The authors conclude that the introduction of the printing technology brought forth an "almost total revolution of the discipline". If we compare all this with the current situation within lexicography, we can see some similarities but also some very interesting differences. There is little doubt that the application of the new disruptive technologies implies an even bigger revolution than the one sparked by the printing technology. As could be expected, it has already led to big changes in the lexicographical product. Fuertes-Olivera et al. (2018: 156), for instance, list "four big transformations" that are going on simultaneously. Among these transformations is the one going from the traditional stand-alone dictionary to a product that is integrated into other information tools, as well as the one going from the dictionary as such to lexicographical data that is handled without appearing in the form of a dictionary. We will later have a closer look at these two phenomena and their possible role as the saviours of lexicography's future. However, we will first look at two other current tendencies.

As we saw above, Fuertes-Olivera et al. (2018) observed that many lexicographers became well-known and famous personalities as a result of printing technology. This was, among other things, due to the fact that many dictionaries were one-man projects, that the printed book format allowed for the authors' names to be put on the cover or front page, and that the dictionaries reached out to a growing number of readers. Today, more or less the opposite is happening. Users who consult online dictionaries will in most cases never see the names of their authors, even if they can be found by means of a link. The brutal fact is that this degrades lexicographers to anonymous data engineers and skilled workers. Obviously, this may affect their self-esteem.

Fuertes-Olivera et al. (2018) also observed that lexicography turned into a successful business after the introduction of printing technology. Today, publishers are struggling to find an appropriate business model and lexicographers are tearing out their hair in frustration when they look for funding for new dictionary projects. The problem in both cases is that an erroneous understanding of the very content of lexicography as a discipline seems to blur the big picture and prevent them from seeing the wood for the trees. A short reflection on the birth of European lexicography will help us to pick up this idea and put it into a future perspective.

### **3. Back to the roots**

Hanks (2013: 507) explains how European lexicography can be traced back to the Classical Greek Period. In the fifth century B.C., it was customary for Greek scribes to insert glosses in manuscript copies in order to explain unusual and obsolete words that appeared in earlier works by Homer and other writers.

Two hundred years later, the glosses were compiled into separate glossaries by scholars at the library in Alexandria (see Figure 1). According to McArthur (1986: 76), historians of lexicography consider that the "origin of the 'dictionary' proper" can be dated back to this practice.



**Figure 1:** Timeline for the birth of European lexicography

Unfortunately, neither Hanks nor McArthur states clearly when exactly lexicography started. Was it in the fifth or the third century B.C.? This is not a rhetorical question as it has huge consequences for the understanding of the discipline, its content, and its product.

In his famous *Dictionary of the English Language*, Johnson (1755) defines a lexicographer as "a writer of dictionaries" and lexicography as the "art and practice of writing dictionaries". These definitions are repeated in a big number of dictionaries and academic works on lexicography. Nobody denies that the compilation of dictionaries is central to practical lexicography. But is it convincing to exclude other activities from the definition? Why were the scribes' glosses not lexicography? Although the etymology of a word does not necessarily define its modern meaning, it may be relevant to know its origin when working on a timeline. The term "lexicography" is originally Greek and means "writing about the lexicon", precisely what the scribes did. They produced lexicographical data (glosses) addressed to difficult words (glotta) that were later compiled into the glossaries that represent prototype dictionaries in the European tradition.

If the scribes' work cannot be considered part of lexicography, then an increasing amount of work made by 21st-century lexicographers cannot be considered lexicography either. The latter are also producing lexicographical data which in many cases do not end up in dictionaries, whether printed or digital. Their data are stored in lexicographical databases and can be used for multiple purposes. When the Danish company Ordbogen A/S in April 2019 published the first version of its Spanish–English *Write Assistant* (Fisker 2019), its infor-

mation engineers retrieved data from six different digital dictionaries, i.e. six different lexicographical databases, in order to serve this tool. The lexicographers who originally prepared these data have probably no idea of what their data are used for and neither would they be able to recognize them in the new environment. This is not the exception, but the beginning of a new era where publishing houses and other companies are increasingly receiving their revenue from using, handling and selling lexicographical data instead of dictionaries.

It goes without saying that this situation does not add to the lexicographers' self-esteem unless the big picture is grasped and the new paradigm understood in all its complexity. Basically, lexicographers have two options in the long run. Either they accept that their profession comprises far more than the compilation of dictionaries, or they will have to prepare a farewell party for lexicography as a millennial cultural practice because their own work is increasingly presented in forms different from the traditional dictionary.

If they chose the first option, they will undoubtedly find it much easier to go for the money and discern a new business model for their discipline.

#### 4. Upstream in the value chain

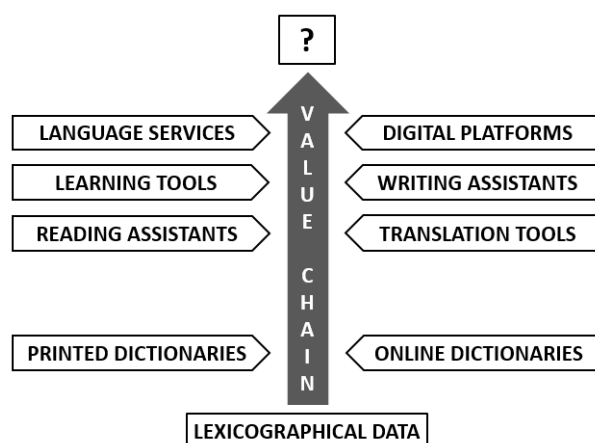
As described in Section 2, lexicography has entered a period of financial unsustainability because its hitherto business model has proved obsolete. Only a handful of publishers earn money from their current flagships, i.e. the online dictionaries. This in spite of the fact that these products have tremendous possibilities of improving the traditional dictionary in terms of quantity, quality, updatability, and accessibility. In this light, it is surprising that only few contributions, and even fewer constructive ones, have been published on the subject in the scholarly literature on lexicography. Among the exceptions are Simonsen (2017) and Fuertes-Olivera (2019). The two authors agree that the only way forward is to go upstream in the value chain. The latter writes:

It is necessary to move upstream in the value chain and develop lexicographic services instead of lexicographic products. (Fuertes-Olivera 2019: 25)

This statement points in the right direction although it lacks some accuracy. First, it is not only a question of developing lexicographic services, but also platforms, tools, etc. Second, a lexicographic service is also a lexicographical product, thus a better term for the latter would be "traditional dictionaries" instead of lexicographic products. And third, it cannot be a question of developing these new things "instead of" dictionaries, but in addition to them. Users will still need dictionaries for a long time ahead. The existence of a societal demand for continuously updated high-quality dictionaries cannot be ignored. Thus, the objective of a new business model should be to find financial resources that *also* allow the publishers to sustain their online dictionaries. The dictionaries could then be used by the publishers to promote their different products as well as resources to which users can be referred from the latter.



All this implies that lexicographers and publishers should focus more on new services, digital tools, platforms, etc., (see Figure 2). There seem to be two ways of using their products. The owners of lexicographical data, i.e. databases, can either commercialize these data, or part of them, allowing external service providers, software developers and other stakeholders to make use of them and integrate them in their products, or they themselves can develop new tools, platforms, and services which users are prepared to pay for. In both cases, they should explore all the possibilities and emerging technology breakthroughs mentioned by Schwab (2015) in his reflections on the Fourth Industrial Revolution. This could be robotics, artificial intelligence, etc.



**Figure 2:** Upstream in the value chain

The move upstream in the value chain cannot be done in a haphazard and amateurish way. The demands of the Fourth Industrial Revolution are enormous, and so are the requirements to future lexicography.

- First of all, it is important to have, or find someone who has, *financial muscles*. The projects will in most cases require considerable investments as the products are expected to be technologically complex and will have to be commercialized on an international scale with many competitors in order to produce revenue.
- Secondly, it is a condition *sine qua non* for success that the work is performed in close *interdisciplinary collaboration* between lexicographers, information engineers, industrial designers, and other relevant specialists. The days with homemade databases are long over. Things have to be done on an industrial scale with the most qualified players.
- Thirdly, a good dose of *creativity* is needed. The collaboration between experts should not be misunderstood as a sum of the different expertise

and skills. An untimely each-of-us-in-our-own-garden attitude would only lead to "linear solutions without a broad perspective" as stated by Steve Jobs at the top of this article. Instead, the collaboration should be carried out as a brainstorm-like interdisciplinary confrontation of ideas that allows the experts to "synthesize new things".

Whatever the visions are, it may prove difficult to concretize a new lexicographical business model without the integration of these three basic requirements. And they will definitely not be the only ones.

## 5. Recent developments

It may seem strange to start a section dedicated to recent developments in lexicography with another visit to the old library in Alexandria. It will nevertheless turn out to be very useful. When the scholars at the library collected the first glossaries more than two thousand years ago, they introduced two important innovations that may help us to throw light on current and future trends in lexicography. The two inventions are the lexicographical article and the dictionary format, respectively.

The first articles were rather simple. Apart from the gloss, they consisted of a "glotta", the Greek for a difficult word, to which the former was addressed (McArthur 1986: 76). The glotta is equivalent to the lemma in modern lexicography whereas the gloss represents the lexicographical data in embryonic form. This is all it takes to constitute a lexicographical article which later became increasingly sophisticated with much more data that were structured in different ways (microstructure).

The second invention was the dictionary format as a collection of lexicographical articles that are structured according to one or another principle by means of the lemmata (macrostructure).

Even though the old glossaries and their articles may seem far from modern standards, they nevertheless represent prototypes of the dictionary format and overall article structure that have survived until our time. During the past two decades, this tradition has creepingly been challenged as a result of the new digital forms of presentation of the lexicographical product.

The first victim to be sacrificed was the dictionary as a collection of articles with a macrostructure. A modern online dictionary consists of a number of articles. The user can explore whether a certain word has been lemmatized and honoured with its own article. However, in most cases, he or she cannot get an overview of all the articles contained in the dictionary. A few years back, many digital dictionaries could be accessed through a separate alphabetic list which allowed the user to embrace all the lemmata treated. This practice has been abandoned in most online dictionaries today. It implies that the dictionary has lost its character of a collection of articles in the sense that the user cannot get a clear sight of this collection as it was the case with the printed dictionary. In addition, it also means that the articles are no longer displayed to the user in

the framework of a specific structure. Any talk of macrostructures in such dictionaries is therefore pure nonsense.

So far, the second victim has only been partially sacrificed. Online dictionaries accommodate articles with lemma, data, and microstructure. But the former static structure has increasingly been replaced by a dynamic structure that adapts to different user needs in different types of consultation. This implies that the amount and organisation of the displayed lexicographical data are fluctuating. And to this should be added that many dictionary articles have been broken up and require clicks, scrolling down and other techniques to be visualized in their totality.

As we will see in the following chapters, this sacrifice of time-honoured lexicographical traditions can be expected to further accelerate in the nearby future. Yet, the various mutations do not change the fact that we can still talk about lexicographical products. If anything, they should rather be viewed as a natural and necessary adaptation to the new digital environment.

## 6. Visions and functions of *Write Assistant*

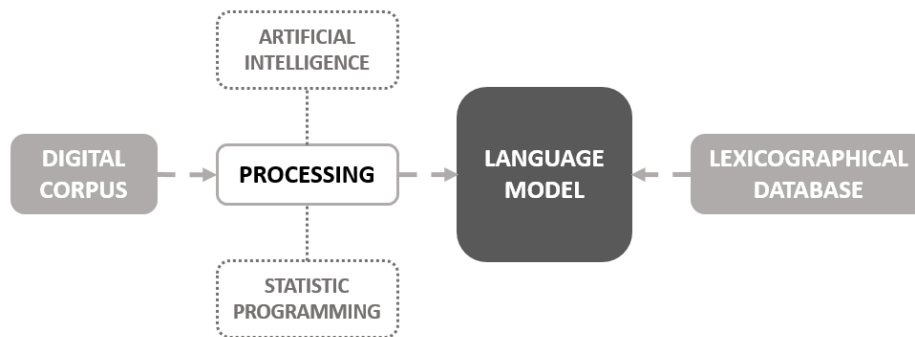
*Write Assistant*, the series of bilingual writing assistants developed by the Danish company Ordbogen A/S, is a good example of both the new possibilities and the new challenges which lexicography experiences today. Its functionality has been extensively explained by Tarp et al. (2017). Hence, in this section we will only discuss some aspects that are either new or relevant to our topic.

*Write Assistant* is designed with only one main function as defined by Function Theory, namely to assist its users when writing in a second language; cf. Tarp (2008). The basic idea was conceived in a remote Swedish farm where five people with different backgrounds gathered during a seven-days brainstorm-like session. As such, it is not born out of traditional user research, but of disruptive thinking.

The underpinning philosophy is based on two important *observations*. The first one is that most writing today is performed on smartphones, tablets, and laptops. Writing skills are mainly used in connection with these devices whereas handwriting is increasingly restricted to our personal use. Dictionary consultation to solve text-production problems is therefore almost exclusively done when writing on these devices. First conclusion: Lexicographical assistance should be available directly on smartphones, tablets, and laptops.

The second observation is that people waste too much time when consulting external sources, even if these sources are available on the mentioned devices. This may affect their writing flow, focus and concentration, and in some cases they may even forget what they were writing about after an excursion to external sources. Second conclusion: Lexicographical and other types of assistance should be integrated in other tools and made available directly in the documents and texts which people write in Word, Outlook, Gmail, PowerPoint, Excell, WhatsApp, Messenger, WeChat, Explorer, Safari, Chrome, Firefox, Facebook, Twitter, and a long etcetera.

*Write Assistant* has therefore been designed as an application that can be downloaded to the user's device and provide instantaneous assistance when the user is writing a text in a foreign language. The tool connects its users to *big data* taken in from two empirical sources, an L2 digital corpus and a lexicographical database. Its driving power is a *language model* that has been trained on the corpus, originally using statistic programming, but now *artificial intelligence* is increasingly being incorporated. As such, *Write Assistant* makes extensive use of lexicographical data that are imported from different sources and even generated in different ways (see Figure 3).

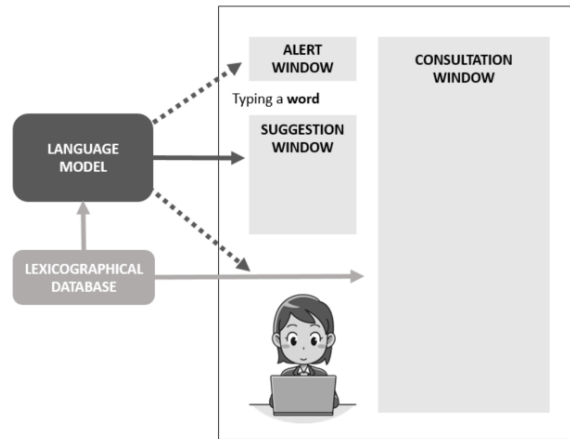


**Figure 3:** The language model and its empirical sources

When fully developed, the assistance will be provided to the users in three different windows popping up in the document on which they are working, i.e. the suggestion, consultation and alert windows, respectively (see Figure 4).

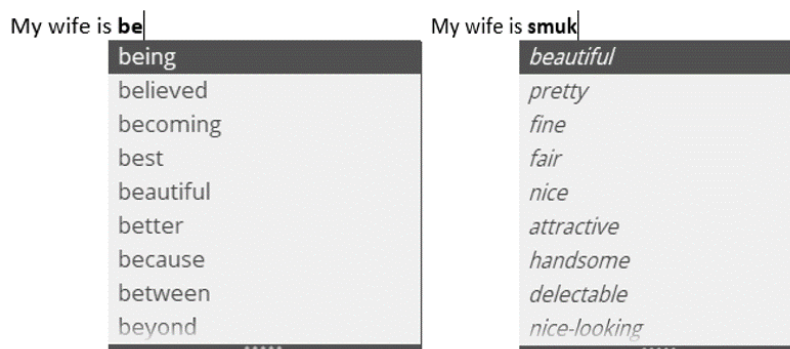
The suggestion window appears as default on the user's screen and offers *L2-word completions* and *next-words* when the writer types one or more letters or a full word in L2. These suggestions are generated automatically by the language model and are ephemeral in the sense that they can only be recreated if the users write exactly the same sequence of words. In addition, the suggestion window also offers *L2 equivalents* to L1 words typed by the users. These equivalents are fed by the lexicographical database and presented in a prioritized, context-aware order by means of the language model.

If the writer does not know which of the suggested words to use, or how to use it, a simple click or touch will activate the consultation window which allows the writer to access lexicographical data such as meaning, inflection, grammar, synonyms, etc. This window is also, in a certain sense, lexicography's window to the future. The third window, which has still to be designed, is foreseen to provide *alerts* that are only activated when the users write a word which the designers of the tool deem to be linguistically or culturally problematic.



**Figure 4:** Write Assistant's three data windows

The different data types provided in the three windows pose some conceptual problems to lexicography. It is evident that the data presented in the consultation and alert windows, as well as the L2 equivalents in the suggestion window, are lexicographical as they have been (or will be) prepared by lexicographers and stored in the database. But what about the L2-word completions and next-words provided in the suggestion window? These data are generated automatically by the language model and, as such, they are not the result of a human lexicographer's work. Can they be considered lexicographical data? Is the intervention of a human lexicographer a prerequisite for data to be considered lexicographical?



**Figure 5:** Suggestion windows with L2-word completions and L2 equivalents

Figure 5 shows the suggestion window with two different types of data. The one to the left appears when the writer types "my wife is **be**..." and contains a list of nine likely L2 *completions* of "be", whereas the one to the right provides nine L2 *equivalents* after the writer has typed the L1 word *smuk*. Apart from this, there is no difference in content between the two windows; both contain nine English words listed in a prioritized, context-aware order. The only difference is how the words are generated. In this perspective, Rundell and Kilgarriff (2011: 278) write:

We envisage a change from the current situation, where the corpus software [...] presents data to the lexicographer in [...] intelligently pre-digested form, to a new paradigm where the software selects what it believes to be relevant data and actually populates the appropriate fields in the dictionary database.

The two authors leave no doubt that the data generated in this way are lexicographical although the process is completely automatic and performed without passing through the human eye of the needle. The only difference between their example and the one discussed above is that *Write Assistant* takes the process even further and offers the automatically created data directly to the end-user.

## 7. Lexicographical challenges

The gradual incorporation of artificial intelligence into *Write Assistant* gives rise to many incognitos. To what extent will deep learning be able to improve the tool? Although success seems to be guaranteed along the main lines, artificial intelligence may not succeed to honour all expectations. Will it, for instance, be able to convert vulgar language into formal language? Will it be optimized to the point where it can convert a first-year student's clumsy text into high-standard academic writing? And if it eventually will be able to do this, how many years will it take until the dream becomes reality?

In spite of these incognitos, in the eye of a lexicographer (there may be other eyes), the lexicographical ingredient represents currently *Write Assistant's* Achilles heel if the vision, as it is, is to turn it into a premium tool. In this perspective, the three different types of window discussed above pose a number of challenges to lexicography, among which can be mentioned:

- Which lexicographical data do writing assistants with the described characteristics require?
- Which are the words requiring linguistic and cultural alerts? And which are the lexicographical data needed to support these alerts?
- Which lexicographical data are required if *Write Assistant*, apart from assisting L2 writing, should also be designed as a learning tool. (It transpires that many learners want to use it with this purpose.)

- How can traditionally prepared lexicographical data interact with data that are automatically generated by means of artificial intelligence?
- How can *Write Assistant* balance *data pushing* and *data pulling* procedures so the users do not lose their responsibility and feel that they have been cornered with the suggestions popping up on their screen? (Within information science, data pushing is defined as a situation where the sender decides on the data to be pushed towards the user, whereas data pulling implies that the receiver can decide what to receive; cf. Gouws (2018). Any decision in this regard may have important consequences for language didactics and learning.)

Many more questions could be asked in this connection. The future lexicographical improvement of *Write Assistant*, for example, will also have big consequences for the lexicographical databases used to sustain it. The experience so far indicates that existing databases are highly deficient and problematic when they are used to feed the tool. Either they do not contain the data required to feed *Write Assistant*; or they do not have them in the necessary quantity; or these data are stored in the database in such a way that they cannot be used properly. As a result, it was necessary to import data from six different digital dictionaries to feed the Spanish–English version of *Write Assistant*, while the German–English version required data from three different sources. Even so, a lot of challenges remain unsolved.

Moreover, once the required data types have been determined and an adequate database designed, another big challenge is posed to lexicography:

- How should the relatively big amount of lexicographical data required to meet the user's consultative needs be presented and structured in the relatively small consultation window without creating data overload with a too long and user-unfriendly access route?

If Shakespeare were still alive, he would have repeated the famous words: "That's the question". In the next section, a possible answer will be discussed.

## 8. The window to the future

The following is not necessarily what is going to happen. The proposal is based on several open and forthright discussions and exchange of ideas with experts from other fields. So far it is exclusively a proposal developed from the side of lexicography, and it only focusses on the consultation window when it is accessed from the suggestion window, i.e. excluding access from a future alert window (see Figure 4). As such, it reflects a possible "laboratory" or working method that can be used in this and similar projects.

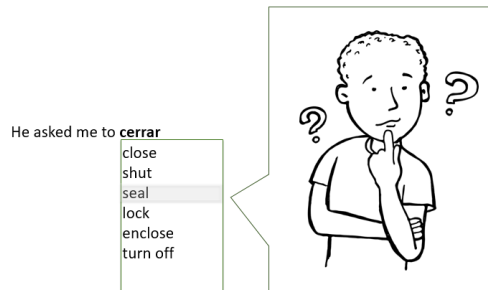


Figure 6: The challenge: How to fill the consultation window?

Figure 6 shows the point of departure. A Spanish user writes "He asked me to" and continues with the Spanish word *cerrar* because he or she is not sure which English word to use. A list of likely English equivalents are immediately furnished in the suggestion window. The writer does not know the meaning of "seal" and therefore activates the consultation window clicking on "seal". Thus, the challenge is how to fill this window with a view of meeting the user's concrete needs. A possible solution can be seen in Figure 6.

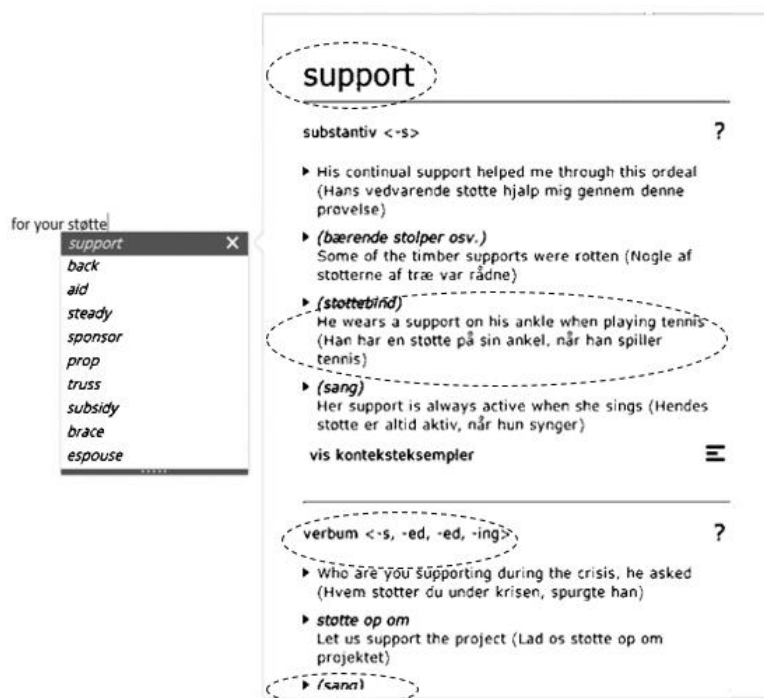


Figure 7: The original Danish–English beta version



Figure 7 is a screenshot from the original beta version of the Danish–English *Write Assistant* where an article from one of Ordbogen's existing online dictionaries has been used to fill the consultation window. This solution has several problems that make it inapt for a tool like *Write Assistant* if we take into consideration that it should be as easy as possible for the user to find what he or she needs. This requires a careful selection of the data to be presented as default. The design of the article in Figure 6 does not help in this regard.

Firstly, the lemma, apart from the big letters, seems to be completely redundant as the user perfectly well knows from which word the article has been accessed. Secondly, the explanations are given in the form of example sentences. This may not be the best solution as it requires a complex and possibly time-consuming mental process to deduce the meaning. Thirdly, the example sentences are provided in both L1 and L2. This is also a space robber that causes the tool to hide several example sentences so they require a further step to be visualized (*vis konteksteksempler*). A better solution would be to furnish an L1 sentence as default with the option to expand it with its English translation. Fourthly, the article offers both part of speech and inflection. These data may also be relevant, but it is probably not the first class of information demanded by the user who, in most cases, is expected to start the consultation process because of comprehension problems. Lastly, the considerable amount of excess data, at least in this consultation phase, occupies space and compels the user to scroll down in order to get more relevant information on meaning.

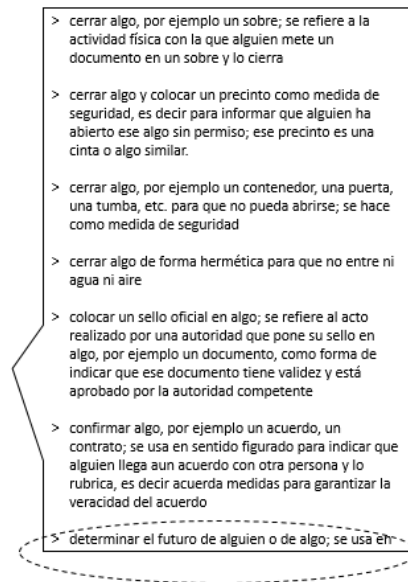
As mentioned above, it became necessary to import data from six different sources in order to feed the first version of the Spanish–English *Write Assistant*, and even so the result was not satisfactory. Ordbogen A/S is, therefore, preparing a premium English–Spanish–English lexicographical database that is currently being made at the International Centre for Lexicography at the University of Valladolid with Pedro Fuertes-Olivera as the main editor; cf. Fuertes-Olivera et al. (2018).

A print of the word "seal" in this database takes up nine A4 pages, containing three lemmas (one verb and two nouns with different inflection paradigms) with a total of 19 senses and a large number of lexicographical data addressed to each sense (see Figure 8). Thus, the big challenge is now to elegantly put all these data into the consultation window without incurring in data overload and forcing the user to take too many steps before getting the needed information.

In the following, an alternative design of the consultation window will be proposed. It is based on the criterion of relevance (cf. Bothma and Tarp 2014). The most relevant data are provided first (as far as it is possible) and the less relevant have to be accessed through further steps. This implies, among other things, that navigation techniques like scrolling down and sweeping to the sides, when they cannot be avoided, are reduced to a minimum. The guiding principles are "less is more" and "simplicity is the ultimate sophistication". Figure 8 shows the initial attempt in this spirit.

- seal** verbo  
seal, seals, sealed, sealing  
<https://howjsay.com/search?word=seal> pronunciación
1. cerrar algo, por ejemplo un sobre; se refiere a la actividad física con la que alguien mete un documento en un sobre y lo cierra  
valencia
    - someone seals something
    - alguien sella algoNota de uso
    - Verbo transitivo, es decir usado con un complemento directo.sinónimo
    - seal up
    - close upequivalente
    - sellar verbocolocaciones
    - she folded it, sealed the envelope, and ran to the postbox
    - lo dobló, cerró el sobre y corrió al buzón de correos
  2. cerrar algo y colocar un precinto como medida de seguridad, es decir para informar que alguien ha abierto ese algo sin permiso; ese precinto es un sello de cera o algo similar  
valencia
    - someone seals something
    - alguien lacra algoNota de uso
    - Verbo transitivo, es decir usado con un complemento directo.sinónimo
    - fasten
    - secureequivalente
    - lacrar verbocolocaciones
    - the civil servant sealed it with the king's seal
    - el funcionario lo lacró con el sello real
  3. cerrar algo, por ejemplo un contenedor, una puerta, una tumba, etc. para que no pueda abrirse; se hace como medida de seguridad  
valencia
    - someone seals something
    - alguien sella algoNota de uso
    - Verbo transitivo, es decir usado con un complemento directo.sinónimo
    - shut
    - close
    - stop upcolocaciones
    - the container was sealed and nobody could open it up again
    - the container was sealed and nobody could open it up againcerrar
    - sealed the wood before applying paint
    - sellaron la madera antes de pintarejemplo
    - She merely filled the containers, sealed them with a cork, and pasted on labels.
    - Simplemente cerró los contenedores, los selló con un corcho y puso las etiquetas.
  4. cerrar algo de forma hermética para que no entre ni agua ni aire  
valencia
    - someone seals something
    - alguien impermeabiliza algoNota de uso
    - Verbo transitivo, es decir usado con un complemento directo.sinónimo

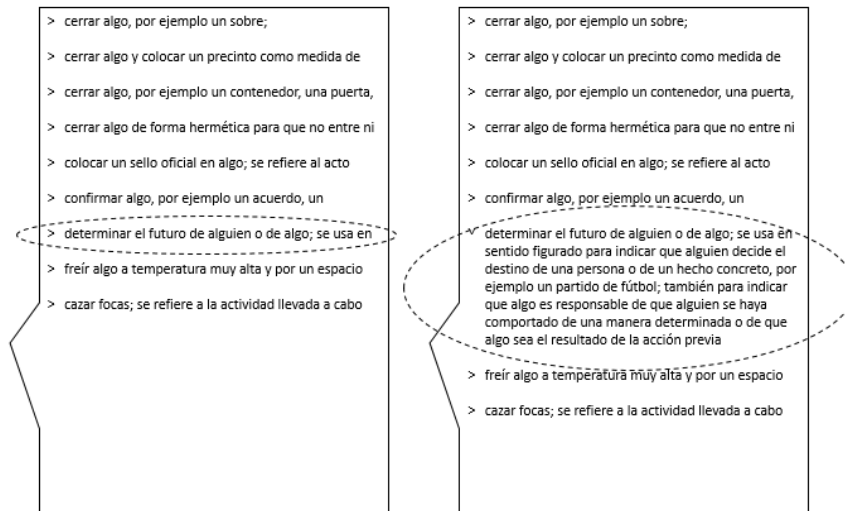
**Figure 8:** Extract from English–Spanish database in the Valladolid-Uva project



**Figure 9:** Design of the suggestion window based on relevance

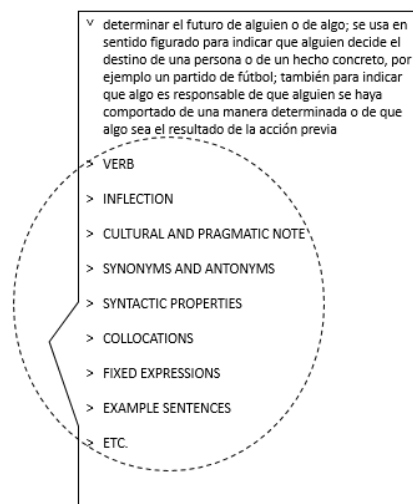
In Figure 9, the definitions of the various senses of the verb "seal" have been imported from the English–Spanish database shown in Figure 8. Here the user gets what is considered most relevant when using *Write Assistant*, namely the meaning of the various senses of "seal". However, the database contains nine senses with the verb "seal" and this implies that the user has to scroll down in order to see the remaining senses. (Some common words in the database have 20, 30 or even more senses). Hence, an alternative solution is shown in the left screenshot in Figure 10. In this case, all the definitions are cut down to only one line allowing for all nine senses of "seal" to be displayed immediately. The idea is that the user in this way can get a preliminary idea of the meaning of each sense, so to say "smell" it. If it "smells" good, it can be expanded with a simple click on ">", whereupon the whole definition will be displayed as shown in the right screenshot in Figure 10.

The solution drafted in Figure 10 creates other problems. If short one-line definitions are not foreseen and prepared for the database, many sentences will be cut in the middle. Although this solution could be recommended as interim, it is nevertheless disturbing and could imply that the user would find it difficult to deduce the preliminary meaning from the abridged definitions to the left in the figure. In the long run, the inclusion of both short and longer definitions into the database will, therefore, have to be planned from scratch as already indicated by Fuertes-Olivera et al. (2018: 159-160).



**Figure 10:** Compact one-line definition and expanded definition

Another inconvenience in Figure 10 is that the expanded definition is placed in the middle of the other one-line definitions. This leaves little room for other data types that could invite the user to a third step in the consultation process. A solution to this problem could be that the remaining one-line definitions are hidden and the expanded one placed at the top of the consultation window as shown in Figure 11.



**Figure 11:** Article with metatexts for further access

The proposal in Figure 11 also includes a list of so-called metatexts that need further elaboration. The idea is that a third click on one of the metatexts allows the user to access additional data that may be needed in a concrete consultation. The order of the metatexts, as well as alternative ways of presenting them, will be discussed in Section 9. As we will see, at this point it does not really matter.

Much more important is it that Figure 11 seems to represent a completely new type of lexicographical "article". First of all, it transpires that all other data, with no exception, are addressed directly to the definition of each sense. What does this mean for the article structure? Has the definition gone through a metamorphosis and converted into a lemma? Or could it be claimed that the real lemma is the word in the suggestion window from where the default article in the consultation window is accessed? But where is then the lemma to the article with L2 words and equivalents listed in the suggestion window? Is it the word — or the letters — typed by the user?

Instead of becoming inquisitive and risking the return of the Cretan Labyrinth, it would be easier to ask the old Greek scribes for an answer. The fact is that the modern "scribes" are now doing something very similar to what their Greek predecessors did before the scholars in Alexandria started compiling glossaries. When they inserted glosses into manuscript copies of old texts, they did it directly in the context where a problem might occur and without the need to invent lemmas and all those frozen article structures that characterized the printed book. They so to say contextualized the lexicographical data; cf. Tarp and Gouws (2019).

The only real difference between the old and the new scribes in this respect is that the latter are doing something much more sophisticated based on disruptive technologies that are light years away from the scribes' pen and parchment.

## 9. Intermezzo on lexicographical user research

For people engaged in the development of *Write Assistant*, it is obvious that the challenges posed by this and similar disruptive tools cannot be solved by means of traditional user research. In 1982, while building the Macintosh computer, a young Steve Jobs was asked whether he thought it was necessary to do some market research to see what customers wanted. His straightforward answer was:

No, because customers don't know what they want until we have shown them. (Isaacson 2011: 143)

This reflection is still relevant almost forty years later. Current lexicographical user research does not make a dent in the universe. It is generally conducted into already published dictionaries or so-called prototypes with no prospect of being produced due to financial constraints. Its results are most often pub-

lished months or even years after the research has been conducted. The real needs of users as they express themselves *before* dictionaries are consulted are largely ignored. It is therefore of little relevance for the design of a completely new tool aspiring to prove its *raison d'être* in the era of the Fourth Industrial Revolution.

Besides, it should not be forgotten that each dictionary has its own personality. It is problematic to generalize from one type of dictionary to another, from one type of user to another, from one language to another, from one culture to another. Although there are general observations that are valid for the consultation of dictionaries in general, concrete instructions for a concrete dictionary can only be based on research into the usage of this particular dictionary. Anything else is shoddy. General instructions based on general observations only have little practical value and may even produce the opposite result when applied to a concrete dictionary.

The users' immediate reaction to innovations does not have to be the supreme court of truth. When the proposal to remove the lemma from the default consultation window (see Figure 6) was presented, it immediately raised a discussion of whether this radical step should be tested among future users. This position was opposed from the side of lexicography. The argument was that 1) most users are conservative and would prefer what they are used to; 2) *Write Assistant's* main target group is the young generation and they will get used to it within a few hours or days; and 3) the removal of excessive data would benefit the overall design of the consultation window. The situation resembles the one that took place when it was decided to eliminate the cursor arrow keys on the Macintosh keyboard in order to force "old-fashioned users to adapt to point-and-click navigation" with the mouse (Isaacson 2011: 138). The result can be seen today where the mouse has become like a pet for many people.

Regrettably, many lexicographers seem intent on wasting their scarce research time on matters which are fast becoming obsolete, instead of applying their minds to the myriad of challenges posed to their discipline by the Fourth Industrial Revolution. The current technological breakthroughs require above all that more time is dedicated to immersion, reflection, deduction, brainstorm, and interdisciplinary confrontation of ideas.

This recommendation by no means denies the need for user research that can contribute to the confirmation, modification, and/or adjustment of new products. In Figure 10, a number of metatexts in random order were introduced. Two questions immediately arise: How should the metatexts be formulated? And in which order should they be presented? The answers can be achieved by means of a different type of user research, namely the use of iterative processes with *test-driven development* (TDD).

It goes more or less like this: A beta version of *Write Assistant* with metatexts included is provided to a small focus group of 10–15 learners. They are then observed using it during a certain time, e.g. an hour, and subsequently interviewed in order to collect their opinions and suggestions. Upon this basis,

a new version with adjusted metatexts is prepared and tested among a bigger group, e.g. 50 learners, who are also observed and interviewed. This leads to a new adjustment that is tested among an even bigger focus group, and so forth.

Simultaneously, a test is conducted to see which metatexts are most frequently activated by the users so they can be arranged in prioritized order according to their relevance. The same kind of iterative processes are used to test other lexicographically relevant aspects, e.g. how *Write Assistant's* users handle the compact one-line definitions shown in Figure 9, or how the additional lexicographical data accessed through the metatexts could be arranged in the best possible way.

This type of (lexicographical) user research is conducted by professional testers supported by a team of information engineers who implement the adjustments straight away. The whole process normally takes a few weeks until a satisfactory result is achieved. As already mentioned, such a test-driven development of the lexicographical product is in every way much more appropriate in the era of the Fourth Industrial Revolution.

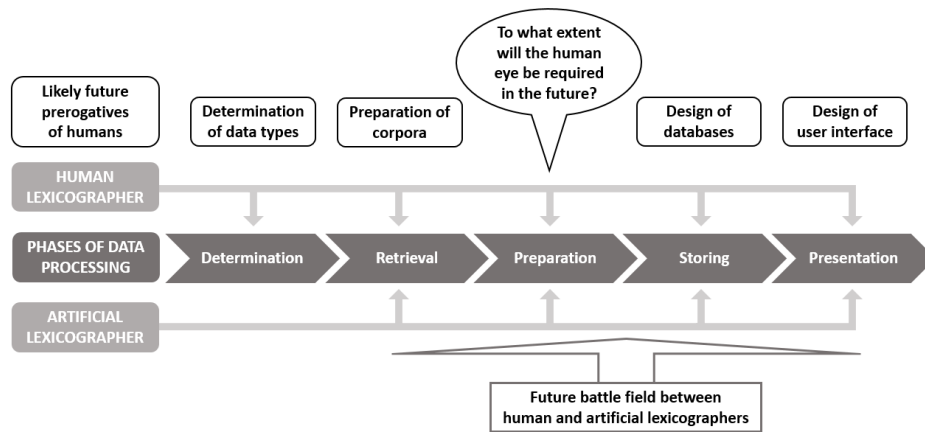
## 10. Perspectives

If we take a panoramic look at the *Write Assistant* experience as well as other developments taking place in present-day lexicography, we will see the contours of a future transformed discipline characterized by the interaction of the traditional human lexicographers with their modern counterparts, the *artificial lexicographers*. This interaction involves at least four of the five main phases of the lexicographical process introduced in Section 2, i.e. the determination, retrieval, preparation, storing, and presentation of lexicographical data.

Figure 12 illustrates how the artificial lexicographer is encroaching on its human counterpart's traditional domain. We already have *example sentences* and other raw data taken automatically from empirical sources and presented to lexicographers for treatment. As discussed in Section 6, we also have lexicographical data taken from empirical sources, processed and presented *directly* to the users *without the interference* of human lexicographers and *without being stored* in the database. In this respect, a new distinction between *lasting* and *ephemeral* lexicographical data is essential.

With the increasing use of artificial intelligence, we will soon have *partially prepared* data presented to human lexicographers for a final touch, and even *fully prepared* lexicographical data stored directly in databases without the interference of human lexicographers.

The trend is unstoppable. There is little doubt that the introduction of artificial intelligence will accelerate it even more. Hence, we can expect that the part of lexicographical data bypassing the human eye will increase over the next years. All this raises the question of the future role of human lexicographers. To what extent will the human eye be required in the future?



**Figure 12:** Interaction between human and artificial lexicographer

Figure 12 gives an overall view of the distribution of tasks between the human and the artificial lexicographer in terms of the five phases mentioned above. Any snapshot of this distribution will soon become obsolete due to the current speed of technological breakthroughs. Even so, it seems likely that the human lexicographer, at least for some time ahead, will continue with a number of prerogatives. These prerogatives include the determination of the data types to be presented to the users, the preparation of corpora, as well as the design of databases and user interfaces which cannot be designed competently without the lexicographical criterium. In addition — and until a software solution capable of deducing meaning from the context is developed — a human lexicographer will still be required to separate meaning into senses, write definitions, select equivalents, attach example sentences and other data to the right sense, write cultural and pragmatic notes as well as other explicit data, for instance, on syntax.

Hence, although future surprises are unavoidable, there is no reason for human lexicographers to despair and be anxious about the future.

At the dawn of the computer age, Dreyfus and Dreyfus (1986) challenged those colleagues who were over-optimistic on behalf of the new technology. They opposed the human mind to the machine with their 5-step model for skills acquisition that was topped by the level of genuine, human expertise characterized by intuition, virtuosity, and effortless performance. Since then, much water has flowed under the bridge "blurring the lines between the physical, digital, and biological spheres" (Schwab 2015). This development, however, should not be seen as an invitation to human lexicographers to retire and externalize their skills and knowledge to the machines. It should rather be seen as a wake-up call to constantly outdoing themselves and proving their worth as an indispensable counterpart to the artificial lexicographers. The newcomers do



not deserve to be met with a frontier wall but should be welcomed with qualified integration for the benefit of their joint users.

### Acknowledgments

The main ideas of this article were hatched during a research stay in May-June 2019 at the Sino-Danish Sindberg Centre for Lexicography, Translation and Business Communication at Guangdong University of Finance, China. Special thanks to its director, prof. Heming Yong, for inviting me and giving me this rare opportunity to dedicate time to reflection, immersion, and contemplation.

Special thanks to prof. Rufus H. Gouws, Stellenbosch University, and prof. Pedro A. Fuertes-Olivera, University of Valladolid, for their comments and constructive input.

Special thanks to the *Write Assistant* team at Ordbogen A/S for always being willing to explain, discuss and exchange ideas.

Special thanks also to the Spanish Ministry of Economy and Competition for financial support (grant FFI2014-52462-P).

### References

- 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): 1-18. <https://academic.oup.com/ijl/advance-article/doi/10.1093/ijl/ecz016/5525336>. (Accessed 17 July 2019.)
- Bothma, T.J.D. and S. Tarp. 2014. Why Relevance Theory Is Relevant for Lexicography. *Lexicographica* 30: 350-378.
- De Schryver, G.-M. 2003. Lexicographers' Dreams in the Electronic-Dictionary Age. *International Journal of Lexicography* 16(2): 143-199.
- Dreyfus, H. and S. Dreyfus. 1986. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York: Free Press.
- Fisker, K. (Ed.). 2019. *Write Assistant. Spanish-English*. Odense: Ordbogen A/S. Available at: [www.writeassistant.com/en/](http://www.writeassistant.com/en/).
- Frankenberg-García, A. 2018. Combining User Needs, Lexicographic Data and Digital Writing Environments. *Language Teaching*: 1-16. <https://doi.org/10.1017/S0261444818000277>. (Accessed 17 July 2019.)
- Fuertes-Olivera, P.A. 2016. A Cambrian Explosion in Lexicography: Some Reflections for Designing and Constructing Specialised Online Dictionaries. *International Journal of Lexicography* 29(2): 226-247.
- Fuertes-Olivera, P.A. 2019. Designing and Making Commercially Driven Integrated Dictionary Portals: the Dicionarios Valladolid-Uva. *Lexicography* 6(1): 21-41. [doi.org/10.1007/s40607-019-00056-8](https://doi.org/10.1007/s40607-019-00056-8). (Accessed 17 July 2019.)
- Fuertes-Olivera, P.A., S. Tarp and P. Sepstrup. 2018. New Insights in the Design and Compilation of Digital Bilingual Lexicographical Products: The Case of the Dicionarios Valladolid-Uva. *Lexikos* 28: 152-176.

- Gouws, R.H.** 2018. 'n Leksikografiese datatrekkingstruktuur vir aanlyn woordeboeke. *Lexikos* 28: 177-195.
- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Hanks, P.** 2010. Lexicography, Printing Technology, and the Spread of Renaissance Culture. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress*, Leeuwarden, 6–10 July 2010: 988-1016. Ljouwert: Fryske Akademy.
- Hanks, P.** 2012. The Corpus Revolution in Lexicography. *International Journal of Lexicography* 25(4): 398-436.
- Hanks, P.** 2013. Lexicography from Earliest Times to the Present. Keith, A. (Ed.). 2013. *The Oxford Handbook of the History of Linguistics*: 503-536. Oxford: Oxford University Press.
- Isaacson, W.** 2011. *Steve Jobs. The Exclusive Biography*. New York: Simon & Schuster.
- Johnson, S.** 1755. *A Dictionary of the English Language*. London: J. & P. Knapton.
- McArthur, T.** 1986. *Worlds of Reference*. Cambridge: Cambridge University Press.
- Nielsen, S.** 2013. The Future of Dictionaries, Dictionaries of the Future. Jackson, H. (Ed.). 2013. *The Bloomsbury Companion to Lexicography*: 355-372. London/New York: Bloomsbury.
- Rundell, M.** 2007. The Dictionary of the Future. Granger, S. (Ed.). 2007. *Optimizing the Role of Language in Technology-enhanced Learning*. Proceedings of the expert workshop organized in Louvain-la-Neuve (Belgium), 4–5 October 2007: 49-51. <https://hal.archives-ouvertes.fr/hal-00197203/document/>. (Accessed 17 July 2019.)
- Rundell, M.** 2014. Macmillan English Dictionary: The End of Print? *Slovenščina 2.0* 2(2): 1-14.
- Rundell, M. and A. Kilgarriff.** 2011. Automating the Creation of Dictionaries: Where Will It All End? Meunier, F., S. de Cock, G. Gilquin and M. Paquot (Eds.). 2011. *A Taste for Corpora. In Honour of Sylviane Granger*: 257-282. Amsterdam/Philadelphia: John Benjamins.
- Schwab, K.** 2015. The Fourth Industrial Revolution: What It Means and How to Respond. *Foreign Affairs*. 12th December, 2015. [www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution](http://www.foreignaffairs.com/articles/2015-12-12/fourth-industrial-revolution). (Accessed 12 July 2019.)
- Simonsen, H.K.** 2017. Lexicography: What is the Business Model? Kosem, I., C. Tiberius, M. Jakubíček, J. Kallas, S. Krek and V. Baisa (Eds.). 2017. *Electronic Lexicography in the 21st Century. Proceedings of the eLex 2017 Conference, 19–21 September 2017, Leiden, the Netherlands*: 395-415. Brno: Lexical Computing CZ, 395415. [elex.link/elex2017/proceedings-download/](http://elex.link/elex2017/proceedings-download/). (Accessed 12 July 2019.)
- Tarp, S.** 2008. *Lexicography in the Borderland between Knowledge and Non-knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Tarp, S. and P.A. Fuertes-Olivera.** 2016. Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-Uva. *Lexikos* 26: 273-296.
- Tarp, S. and R.H. Gouws.** 2019. Lexicographical Contextualization and Personalization: A New Perspective. *Lexikos* 29: 250-268.
- Tarp, S., K. Fisker and P. Sepstrup.** 2017. L2 Writing Assistants and Context-Aware Dictionaries: New Challenges to Lexicography. *Lexikos* 27: 494-521.
- Wolf, G.** 1996. Steve Jobs: The Next Insanely Great Thing. *Wired*, 1st February, 1996. [www.wired.com/1996/02/jobs-2/](http://www.wired.com/1996/02/jobs-2/). (Accessed 17 July 2019.)