International Journal Of Lexicography, 2023, **36**, 233–259
https://doi.org/10.1093/ijl/ecac020
Advance access publication 5 May 2023
Article

OXFORD

The Definition, Presentation and Automatic Generation of Contextual Data in Lexicography

María José Domínguez Vázquez[1], Rufus H. Gouws[2],

[1]Department of English and German Philology, Faculty of Philology and Institute of Galician Language, University of Santiago de Compostela, Santiago de Compostela, Spain (majo.dominguez@usc.es)
[2]Department of Afrikaans and Dutch, Stellenbosch University, Stellenbosch, South Africa (rhg@sun.ac.za)

Abstract

This paper deals with several aspects of context in lexicography. Section 1 briefly mentions some different approaches to the concept *context* in various fields. Section 2 puts the focus on different uses and perceptions of the concept *context* in lexicography, contrasting it with related concepts, such as *cotext*, *contextualization* and *contextual information*. A more comprehensive discussion also covers different aspects of the occurrence of the concept *context* in dictionary research, with specific reference to central aspects of the so-called inner and outer context. Various portals, dictionaries and dictionary entries will illustrate the above-mentioned approaches. Section 3 approaches the subject from a user perspective. Section 4 addresses the question *How can contextual data be extracted or generated?* To answer this question, some methods and tools for (automatic) acquisition and analysis of contextual data, – in particular of the local contextual data in terms of Faber and León-Araúz (2016) – are introduced. Examples of these are lexical databases or semantic networks, like *WordNet*, and corpora, like *Sketch Engine*, or predictive methods, like *Word2vec* and similar ones. Some advantages and disadvantages of specific data acquisition tools used for the analysis of local contextual data are indicated. This section also contributes to a more detailed discussion of the automatic generation of the so-called local syntactic-semantic context or word environment, specifically of the building of syntactic-semantic argument patterns and their examples.

Keywords: automatic generation, context, cotext, data acquisition tools, dictionary-external context, dictionary-internal context, local contextual data, online dictionary, retrievability, syntactic-semantic context

1. Introduction

Human beings approach data collected in different repositories differently compared to machines. Regarding humans, differences can be observed in terms of the linguistic and lexicographical knowledge they have, as well as their knowledge of the available resources consulted and their use; and, not least, their linguistic awareness, as is the case, for example, with language learners compared to researchers or linguists. What human groups indeed share, is a set of cognitive schemas and cultural patterns linked to different types of situations or micro-scenes (Smith 1991). At a cognitive-linguistic level a network of scenes shapes a scenario (Fillmore 1977: 62). Linked to these are different concepts and meanings

- encoded by different syntactic frames (Heringer 1984; Busse 2012; Hermanns and Holly 2007) or syntactic-semantic scenes (minimal surface structures that evoke different cognitive scenes; Heringer 1984; Ziem 2008). These scenes involve different conceptual slots, which are usually conceptually similar in the languages of neighbouring cultures but may differ in the surface realizations and syntactic-semantic combination rules of the respective languages. Therefore, in a communicative act, a speaker activates different communication strategies and surface realizations in each situation, such as, for example, a lexeme together with its syntactic environment. Thus, it would seem quite clear that the difficulty lies in capturing all this multidimensional information in a lexicographic resource in such a way that the expectations and demands of the user and his/her query are satisfactorily met.

A lexicographic process is not just about capturing this information, but also about providing it to the user *in-actu* (Wiegand 1998: 501). To make the information accessible and comprehensible to the target users of a specific dictionary, various aspects of the context of a given form must be considered, and the lexicographer has to identify and adhere to a clearly defined genuine purpose of a dictionary (cf. Wiegand 1998: 299). When formulating the genuine purpose, the lexicographer must respond to the question: *What do I want my user to be able to do with this dictionary?*. Finding an answer to this question is possible once the lexicographer has determined the lexicographic function or functions of the dictionary. According to Tarp (2008: 81) 'A lexicographic function is the satisfaction of the specific types of lexicographically relevant need that may arise in a specific type of potential user in a specific type of extra-lexicographical situation'.[1] Once the function has been determined the lexicographer can proceed to the selection of the required data and the choice of the appropriate dictionary structures to accommodate the data and give the necessary access to these items. When the function has been determined, the data have been selected and the relevant structures are in place a dictionary can be compiled that can successfully be consulted by the intended target user to solve specific lexicographic problems. The presentation of the required data also implies presenting sufficient contextual data to the user. Lexicographers need to be familiar with different types of contexts and their potential relevance to lexicographic resources.

Context is a multidimensional concept: not only due to its use for referring to different realities, but also because various disciplines approach it in different ways. Whatever authors understand *context* to mean, it is a fact that many research studies about context ultimately aim to describe, represent, or codify meanings or to understand the behaviour of languages. Cognitive science (Mcdonald and Ramscar 2001) and neurobiological approaches (Pereira et al. 2018) stress the importance of different types of contexts and consider contextual co-occurrence as one of the most informative factors in the study of meaning. Many studies in language acquisition have evidenced that the learning of vocabulary units, of lexemes, is closely associated with their syntactic-semantic context (Laufer and Nation 2012; Nation 2005; Treffers-Daller and Rogers 2014). In Linguistics and Computational Linguistics, the availability of lexical and contextual knowledge is clearly essential, and in the field of Natural Language Processing (NLP), and to a lesser extent Natural Language Generation (NLG), lexicons play a central role (Wanner 2013: 1150). The growing confluence between these fields and lexicography reconfirms the status of lexicography as an interdisciplinary field (Horák and Rambousek 2017; Trap-Jensen 2018). On the one hand, there is a close collaboration between these fields in the design and use of techniques for data analysis, extraction and generation (Krek 2019); on the other hand, the availability of lexical and contextual knowledge for developing resources for humans or machines is valued at different levels (Navigli and Ponzeto 2012; Simonsen 2020). In the last few decades, neural language models stand out due to their application in different language processing tasks, especially as they can provide information about probability distributions over sequences of words as well as produce contextual word and sentence embeddings (Peters et al. 2018).

As mentioned above, there are different approaches to context. This article deals with context in lexicography from a multidimensional perspective and it is organized as follows.

Section 2 focuses on a definition of *context* and various aspects of the use of *context* in lexicography, whilst Sections 3 and 4 deal with methods and procedures for consulting, as well as for gathering and producing local contextual data to develop lexicographic resources.

2. Context in lexicography

Within a traditional lexicographic approach to context, Zgusta (1971: 47) distinguishes between verbal and situational context which is determined by the use of a lexical item in a language utterance. This approach is confirmed by Hartmann and James (1998: 29) who regard the verbal context as the part of a text where a word occurs, whereas a situational context is the setting or circumstances with which a word is associated. Burkhanov (1998: 46, 47) distinguishes between a linguistic context which is a given language unit's immediately preceding and/or following linguistic units, a situational context which is the 'communicative situation as the circumstances of a particular speech act', and a pragmatic context which is the correlation of 'formally described linguistic parameters, their groupings within texts, with those situational parameters which are constantly recurrent and relevant to the correct understanding of speech acts and speech events'.

Lexicographers also distinguish between a defining context, an explanatory context and an associative context. A defining context is a context that contains substantial information helping you to understand a concept but does not possess the formal rigor of a terminological definition. It presents the required term as well as other related keywords. An explanatory context provides a brief explanation of a concept, but it is not sufficient to fully understand it. For example, rather than saying what something is, it may say where and how it is used. An associative context contains a minimum amount of information needed to associate a concept with a particular subject field and show that it is used by experts writing in their native language.

In more recent research in lexicography a multidimensional nature of the concept *context* prevails on different levels. It plays a significant role in the microstructure of a dictionary with items enabling coherence between a given word and its immediate linguistic environment, cf. Wiegand et al. (2017:175). It also enables a link between the dictionary and the external world, as well as relationships between a dictionary (and its components) and other lexicographical tools (and their components). In dictionary research *context* is also used to refer to various aspects regarding the activity of using a dictionary as reference source, namely the outer and the inner context of use (Wiegand 1987). The outer context of use refers to the external conditions determining a consultation, e.g., the usage situation and the time and place of use and the length of a consultation. The inner context of use includes the aim of the consultation, the search query, interest in the use of the dictionary, reasons for the consultation, etc., cf. Müller-Spitzer (2013).

Another perception of context in lexicography is the paratextual context, cf. McConchie and Tyrkkö (2018). A further appoach to context is found in the treatment presented in so-called context-sensitive dictionaries. Here context is regarded as the word context or environment. 'Context-sensitive dictionaries perform a linguistic analysis of the word and its context in order to enable and to refine the look-up process by linking the inflected word form to its base form listed in dictionaries, and by narrowing the information to be displayed according to the clues provided by the word context' (Seretan and Wehrli 2013: 1046). Although they are referred to here as dictionaries, they actually are context-sensitive dictionary look up engines and they aim to find the best match between the surrounding word (word context or sentence context) and dictionary information (an example is COMPASS/LOCOLEX, an assistant for text comprehension in a foreign language¸ cf. Seretan and Wehrli 2013: 1048). There also are tools that provide context-sensitive translations where the environment of the input text and the dictionary information are matched. An example of such a tool is MoBiDic/MoBiMouse, cf. Seretan and Wehrli (2013: 1050).

2.1. About context and related concepts

In the most typical use of the term *context* in lexicography it refers to language units within their specific environment. In lexicographic studies (cf. Wiegand 1988; Gouws and Prinsloo 2005: 127; Lettner 2020) there are various terms related to the concept *context*, for example *context*, *cotext*, *contextualization*, *dictionary-internal context, dictionary-external context* and *external dictionary context*. According to our point of view these terms can be used in connection to a) the microstructure of the dictionary – that is the representation of lexical units as items in the dictionary article, complemented by the typical text in which they occur, b) the dictionary and the external world – that is the dictionary-external language use, and c) the relation between the dictionary and other sources – that is additional information beyond the contents of the dictionary. It is also necessary to consider the internal and external context of use, as well as the user *in-actu* in the process of contextualisation. These different uses of *context* as displayed in Figure 1 will be discussed in this paper.

2.1.1. *Context and the dictionary microstructure.*

According to Bolinger (1985: 69):

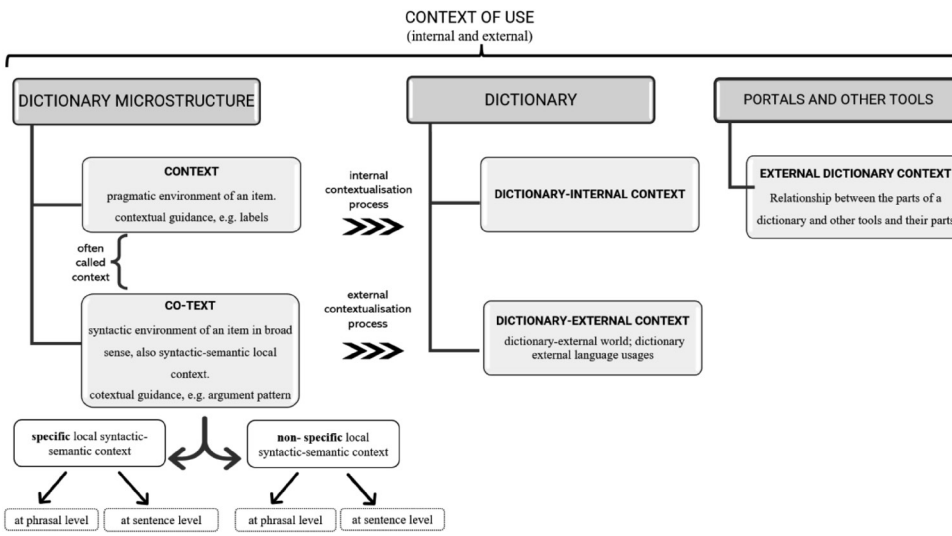


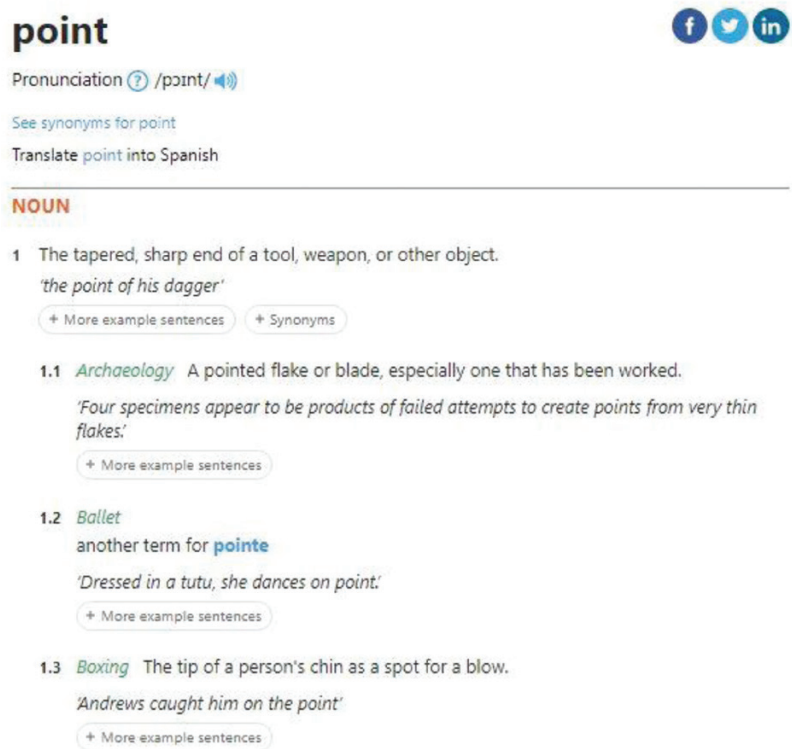
Figure 1. Context at different levels in the lexicographic framework

> 'Lexicography is an unnatural occupation. It consists in tearing words from their mother context and setting them in rows […] to make them fit side by side, in an order determined not by nature but by some obscure Phoenician sailors […] Half of the lexicographer's labor is spent repairing this damage to an infinitude of natural connections that every word in any language contracts with every other word […]'

To enhance the value of a dictionary, an important task of lexicographers is to show something of these 'natural connections that every word in any language contracts with every other word' (Bolinger 1985: 69). This can be done by means of items with a non-lemmatic addressing that complement the items giving paraphrases of meaning and translation equivalents, or items with a lemmatic addressing that complement the lemma. Non-lemmatic addressing occurs where an item in a dictionary article is addressed at an item that is not a lemma, whereas lemmatic addressing is a procedure where an item is addressed at the lemma, (cf. Hausmann and Wiegand 1989: 349; Wiegand and Gouws 2013: 278; Gouws 2015: 164). Complementing items in dictionary articles by means of non-lemmatic or lemmatic addressing

is done in at least two different ways, i.e., on a pragmatic level and on a syntactic level. This results in the distinction made in lexicography between items giving contextual guidance and items giving cotextual guidance. In metalexicography (cf. Wiegand 1988; Gouws and Prinsloo 2005: 127), context is regarded as the pragmatic environment of an item, that is the environment reflecting something of the occurrence of the word in actual dictionary-external language usage, and cotext as its broad syntactic environment.[2] Contextual guidance is typically given by means of lexicographic labels and glosses, whilst cotextual guidance is given by means of example sentences and collocations. In Figure 2, a partial article of the lemma sign **point** (in the former *Lexico.com*; currently no longer accessible), each of the paraphrases of meaning in the subcomments on semantics of the first sense is preceded by a label, the items *archaeology, ballet* and *boxing*. These labels indicate the specific pragmatic environment, the context, of the different sub-senses of the word in actual dictionary-external language use.

The occurrence of different types of contextual information in the dictionary microstructure (see Figure 1) is linked to the relationship between the dictionary and its function(s), so that lexicographers should never underestimate the importance of their dictionaries having to adhere to their identified functions. In this regard Tarp (2000: 198) says that to a greater or


point

Pronunciation /pɔɪnt/

See synonyms for point

Translate point into Spanish

NOUN

1 The tapered, sharp end of a tool, weapon, or other object.

'the point of his dagger'

+ More example sentences + Synonyms

1.1 *Archaeology* A pointed flake or blade, especially one that has been worked.

'Four specimens appear to be products of failed attempts to create points from very thin flakes.'

+ More example sentences

1.2 *Ballet*

another term for **pointe**

'Dressed in a tutu, she dances on point.'

+ More example sentences

1.3 *Boxing* The tip of a person's chin as a spot for a blow.

'Andrews caught him on the point'

+ More example sentences

Figure 2. Article of point in *Lexico.com*

lesser extent, everything in a dictionary is influenced by its functions. The functions must be considered when designing the contents and form of a dictionary. Where either text production or text reception is one of the functions of a bilingual dictionary the lexicographer should not present the translation equivalents as stand-alone and isolated items. Additional items need to be entered to indicate the typical syntactic environment of the word represented by the lemma sign as well as the different translation equivalents. In Figure 3, a partial article from the *Pharos Junior Bilingual School Dictionary* (= Pharos 2018), the comment on semantics contains three

subcomments on semantics with each translation equivalent complemented by an example sentence in both the source and the target language. These example sentences give the syntactic environment, the so-called *cotext* of the lemma, and the different translation equivalents.

In the microstructure of a dictionary the pattern illustrations that display the combination of a word with other words also belong to the *cotext* (or syntactic-semantic context). As indicated in Figure 1, we are distinguishing here between the local context at phrasal level – the immediate environment of a word – and the context at sentence level – the elem-

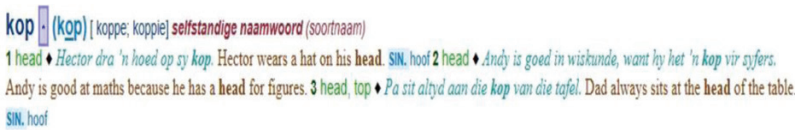

kop · (kop) [koppe; koppie] **selfstandige naamwoord** (soortnaam)
1 head ♦ *Hector dra 'n hoed op sy kop.* Hector wears a hat on his **head**. SIN. hoof 2 head ♦ *Andy is goed in wiskunde, want hy het 'n kop vir syfers.*
Andy is good at maths because he has a **head** for figures. 3 head, top ♦ *Pa sit altyd aan die kop van die tafel.* Dad always sits at the **head** of the table.
SIN. hoof

Figure 3. Article of kop in Pharos.

ents or arguments to ensure at least the sentence grammaticality. In addition, a distinction is made here between specific and non-specific local syntactic-semantic context. This is based on the difference between valency complements (specific) versus circumstants or modifiers (non-specific) (Engel 1996, 2004). It is a matter of a description of the predicate argument structure or combinatory potential of a language unit. In this way we understand in the example *the {refreshing} smell of the {Spanish} oranges* the *of*-phrase as a specific argument of the noun *smell*. The words (given in brackets) *refreshing (*with reference to *smell),* and *Spanish* (with reference to *oranges*) are regarded as non-specific neighbouring words. Likewise, one can also proceed at sentence level, for example: *{Yesterday} Pedro and Juan attended the council meeting {in the assembly hall}.*

In this use of *context,* we are not only talking about syntactic context but also rather of semantic context, because the syntactic structure is linked to the combinatory meaning (cf. Engel 1996, 2004). This syntactic-semantic interface allows the delimitation, understanding or activation of the meaning of a lexical unit and to identify the meaning of a word in a given situation – its syntactic-semantic context. To describe the combinatory meaning, two levels of meaning must be considered (Engel 1996, 2004), namely the relational meaning and the ontological meaning:

(a) The relational meaning determines the semantic relation of semantic roles. For example, the noun *stay* has four semantic roles:
   (1) semantic role 1: 'to be in a certain place for a certain period of time': *the patient's stay*
   (2) semantic role 2: 'place where you stay for a certain period of time': *the stay in hospital*
   (3) semantic role 3: 'a given period of time in which one is in a given place': *the stay from Friday to Sunday*
   (4) semantic role 4: 'Type of stay': *a business stay/a teaching stay/a research stay*.

(b) The ontological meaning is the inherent meaning of the lexical units which can perform a specific semantic role in a pattern structure. In dictionaries this information is evident from the pattern illustrations, e.g., *the stay of someone somewhere,* and the example sentences. Where the lemma represents a polysemous lexical item, the lexicographer has to indicate all the senses – an ontology, but the supporting contextual and cotextual items will show the typical environment in which a specific sense is activated, and the other senses neutralised. When dealing with context, lexicographers acknowledge existing ontologies but also the need to show how and where a specific sense, presented in a subcomment on semantics, applies.

The extent and granularity of the syntactic-semantic information at this level depend on the typology of the lexicographic resource itself. For example, a valency dictionary such as *E-VALBU* presents a much more detailed description of the syntactic-semantic context than traditional

explanatory dictionaries. In the illustration patterns from *E-VALBU*, for example 'jemand/etwas fährt irgendwohin' (**fahren 1**) [someone/something travels somewhere], a pop-up window indicates that *jemand* (someone) is a person in a vehicle [except aircraft] and *etwas* (something) is a vehicle [except aircraft]. In other dictionaries it is shown differently on the user interface, but they often do point it out (cf, Longman: https://www.ldoceonline.com/dictionary/travel). Apart from this, the semantic information linked to syntactic schemes could play an important role in lexicographic resources and should be accounted for in more dictionaries.

2.1.2. Dictionary-internal context, dictionary-external context and external dictionary context

The use of the term *context* in lexicography, referring to both the pragmatic and the syntactic environment, is not restricted to complementing items in dictionary articles. Items like example sentences, collocations and glosses that are presented as microstructural items reflect something of a *dictionary-internal contextualization*. Contrary to this process of internally directed contextualisation lexicographic labels link the item at which it is addressed to the dictionary-external world – a different context. The context given in the dictionary article helps the dictionary user to move from the linguistic expression in the dictionary to the dictionary-external language use. It places the dictionary entry in its context and as a form of lexicographic deixis (cf. Gouws 1988), it anchors the dictionary item in the dictionary-external world. This anchoring could be achieved by the labels *archaeology, ballet* and *boxing* in Figure 2, the partial article from *Lexico.com,* or by means of an item giving a pictorial illustration as seen in Figure 4, the article of the lemma **tuba** in *the Longman* Dictionary of Contemporary English Online.

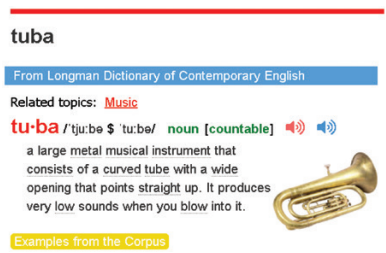


Figure 4. Tuba in the *Longman Dictionary of Contemporary English Online.*

This contextualization plays an important role in satisfying a text reception need of the target user of the dictionary, because words occur in different contexts of use and can convey different senses. The user finding a word in an actual situation of use can move from that dictionary-external context to the dictionary article where the items giving the context will assist him/her to find the required information in the lexicographic treatment of a specific lemma, that is the microstructural items presented in the article from which the user can retrieve the information he/she needs. As an example: when reading a newspaper report on a ballet performance, the reader encounters a sentence like 'during that scene she was dancing on point'. The user who does not know the meaning of this use of the word *point* experiences a problem in an extra-lexicographic environment. He/she then consults a dictionary and moves from an actual situation in a *dictionary-external context* to the article of the lemma **point** in the dictionary where they find a solution to the problem, because this use of the word **point** is explained and labelled as belonging to the field of ballet (Figure 2).

Besides a dictionary-external context as explained in the preceding paragraph, lexicographers also have to make their users aware of another external context – a context provided by reference sources beyond the boundaries of a specific single dictionary. As indicated in Figure 1, this is the so-called *external dictionary context*. One of the results of the transition

from printed to online dictionaries is that the scope of the term *context* exceeds the domain of the individual dictionary articles as lexicographic texts and even goes beyond the dictionary as a text compound (Kammerer and Wiegand 1998: 224). The online environment can compel the lexicographer to negotiate parallel dictionary contexts and other outer contexts. Online dictionaries are often linked to other sources such as e-readers. E-readers, like Kindle, are linked to one or more specific dictionaries. If a user wants to know the meaning of a word found in a text on the e-reader he/she must click on the word and is then guided from the text on the e-reader to an article in the linked dictionary where the specific word is treated. According to Bothma and Gouws (2022) linking establishes a relation between a word in an e-text and the specific lexicographic sources. The word is used in an outer context and to assist the user in an optimal way the linking must be regarded as a contextualizing procedure. Consequently, a parallel context in the software of the e-reader and the dictionary article is needed. This requires an updating of the software of the e-reader so that the context of a specific word in an e-text can be interpreted and linked to the same context in the dictionary article. When clicking on a word in a text read on the e-reader, an outer context, the word and its specific context may be linked to the corresponding word and context in the dictionary. This linking will enable the user to comprehend the specific sense of the word. Increased contextualization procedures are needed on both dictionary-internal and dictionary-external level. This can result in a word occurring in a text included in the e-reader being linked to its specific sense in the dictionary. Here both the pragmatic and the syntactic environments are important and this linking of a smart e-reader to a smart online dictionary brings the comprehensive scope of context in lexicography to the fore.

Online dictionaries often no longer function as stand-alone products but, as search regions, they can be part of a dictionary portal, a search domain. They can be linked to other lexicographic sources in a search universe beyond the dictionary portal (cf. Gouws 2021: 6-7), e.g., the linking to other dictionaries in the DWDS.[3] They can also be linked to non-lexicographic sources in a search universe beyond the dictionary portal (cf. Gouws 2021: 6-7), e.g., in the article of the lemma **computer** in *dict cc* where a click on the *i*, the information icon, opens the pop-up window in Figure 5 that allows the user access to sources like *Google* and *Wikipedia* that are not part of a dictionary portal.

The dictionaries in the portal and the other sources to which the user has access constitute further contexts that might be relevant to the user when consulting a specific dictionary. They can be regarded as external dictionary sources.

3. Retrieving contextual information: a user perspective

A typical user (i.e., without (meta)lexicographic knowledge) can handle resources of different types for obtaining at least some context information from a monolingual and/or bilingual perspective. These sources include:

1. A variety of (electronic) **dictionaries**, with or without multimedia elements, as well as interactive dictionaries like *Visuwords*. Learner's dictionaries can be used to consult different types of local context (semantic, syntactic, pragmatic, etc.; see Section 2.1.1.), although the valency dictionaries (*VALBU 2004*; Herbst et al. 2004; Portlex 2018) and dictionaries of collocations (REDES 2004; *DICE 2004*) seem to be the most useful ones for this purpose.
2. **Corpus analysis tools**, such as *OWID*^{plus}, and a Corpus for language learning, such as *SKELL* (with information about word sketch, similar words and examples), that provide information about the syntagmatic context in a simpler way than general **corpora**.
3. **Machine translation engines** and tools, such as *Google Translate* or *Deepl*. *Deepl*, for example, makes it possible to access paradigmatically related words during

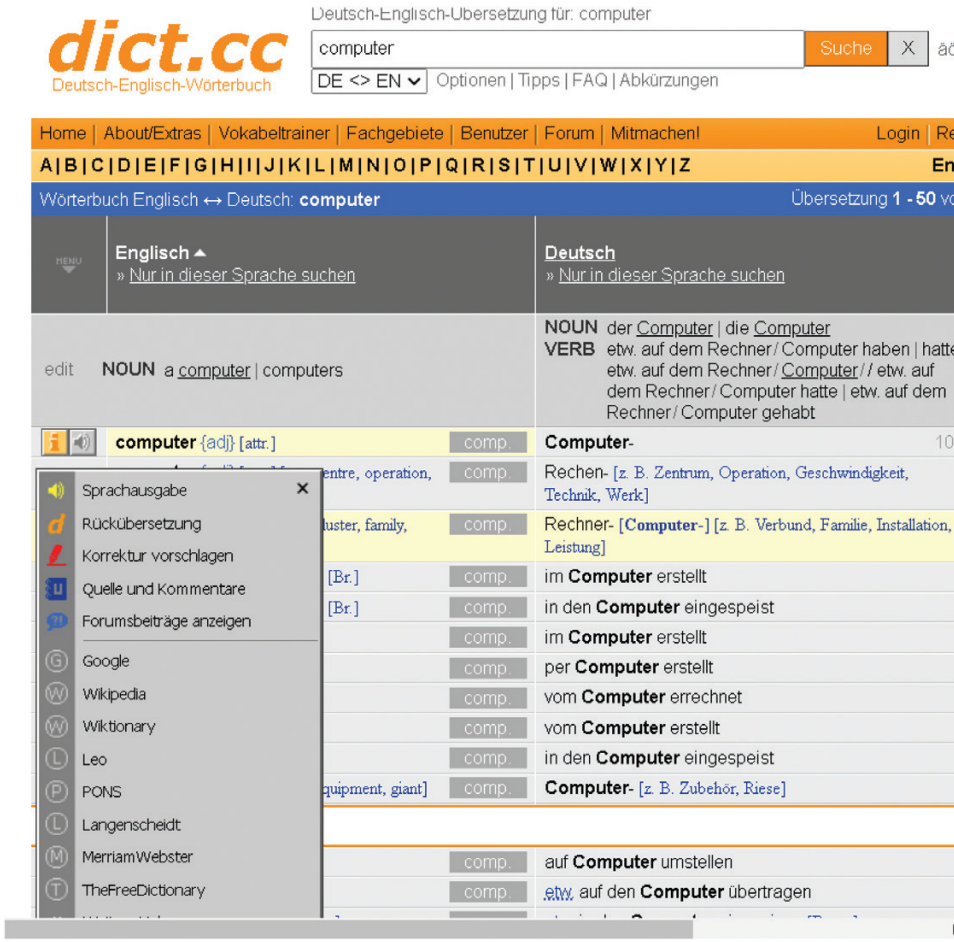


Figure 5. *Computer* in *dict cc*

the process of translation. These words might be more appropriate in the sentence or local context than the ones proposed by the automatic translator, but the user must rely on his/her own intuition and knowledge to select the right one.

4. **Reading and writing assistants.** Reading and writing assistants aim to support comprehension and production tasks. Usually by clicking or touching, the user gets access to more information on a specific word from the integrated dictionaries (for more information, see Gouws and Tarp 2019; Bothma and Prinsloo 2013). An example of a comprehension assistant or reading assistant is *Glosser*. Writing assistants should predict the next word in the sentence – the syntagmatic context. An example of this is the text-editing tool *CollocAid* (Frankenberg et al. 2019), that provides the conventionally used collocations in English texts. In some cases, the assistant is organized to offer equivalents and to provide access to dictionary information about the suggested word. Using more recent tools, it is possible to obtain not only a suggested completion of the next word, but also a selection of possible equivalents in the target language (Fuertes-Olivera and Tarp 2020: 266; Figure 6).

Figure 6. L2-word completions and L2 equivalents from Fuertes-Olivera and Tarp (2020).

5. There are also tools and programs that **automatically generate various text types** (Nallapati et al. 2016; Sordoni et al. 2015), literary sentences (Moreno Jiménez et al. 2020), texts from images and vice versa (Otter et al. 2020) and, in some cases, jokes, poems or stories, almost without input (Roemmele 2016). The **text production robots** can also automatically create a text from a reduced number of words. This means that these robots start from a given context or from a reduced number of words in order to create a text or to complete a message following statistical rules. Some examples are *Inferkit* (Figure 7) or *Saasbook* (Figure 8) (for more information, see Simonsen 2020).

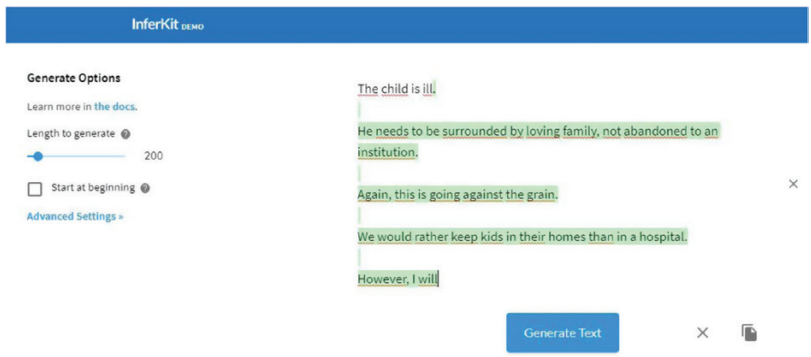


Figure 7. *Inferkit*

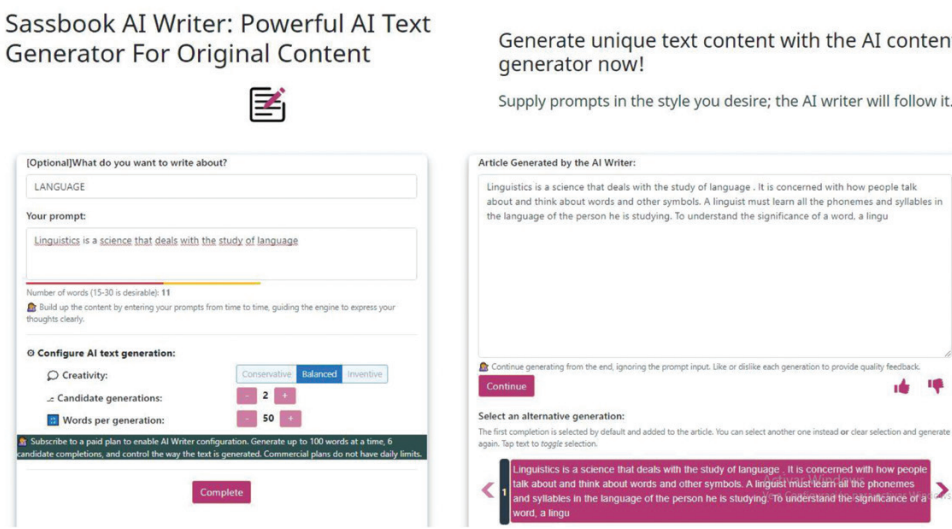


Figure 8. *Saasbook*

The user's approach to these resources and the others mentioned above, such as dictionaries or corpora, is different: in the case of robots, there is neither a concrete user´s query nor a query hypothesis, beyond getting texts generated automatically from a given context. The handling of the dictionary always implies a specific search query (Kühn 1989: 115).

Let us now return to our starting point, namely the use of dictionaries and portals to retrieve relevant information regarding specific subject matter. Dictionaries are used for punctual consultation and there are two participant roles, two agents, that are significant during a dictionary consultation. The active participant is the user looking for a solution to a specific lexicographic problem. This user needs to retrieve information from the data available in a dictionary or tool. This implies that the dictionary that provides the specific data should be regarded as the other role player. In dictionaries and other tools, items reflecting the syntactic and/or pragmatic environment of a given linguistic expression play a significant contextualization role, that is to provide solutions to users' information needs directly in the situation or context where these needs occur (cf. Tarp and Gouws 2019: 250). Consequently, data should be presented in such a way that the intended target user can achieve a successful dictionary consultation process and retrieve the information that satisfies the need that prompted that consultation procedure.[4]

When planning and compiling dictionaries, lexicographers have to negotiate the lexicographic needs and the reference skills of their intended target users. Dictionary articles must contain the contextual items their users need to achieve a successful dictionary consultation. A lexicographer should not rely on the linguistic knowledge and intuition of the dictionary user and should rather err on the side of too much than too little. Consequently, items in dictionaries need to be accompanied by additional items that can enable successful contextualization, according to Tarp and Gouws (2019: 251) 'to provide solutions to users' information needs directly in the situation or context where these needs occur'

A dictionary which successfully deals with this issue is the *Valenzlexikon Deutsch-Spanisch* (cf. Domínguez et al. 2017). As shown in Figure 9, the article provides the meaning

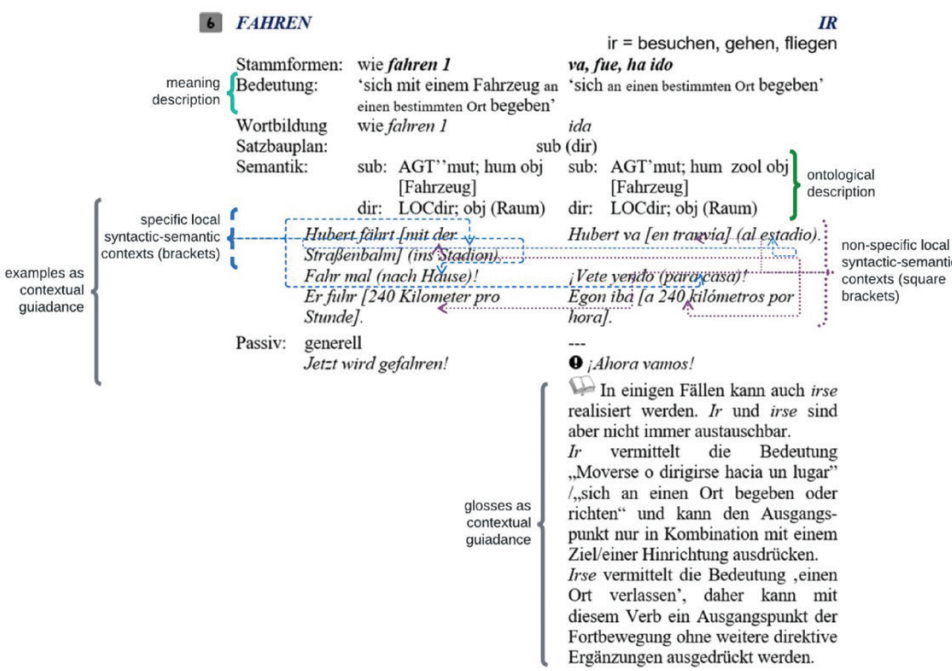


Figure 9. Context information levels in the microstructure

description and the ontological features from a contrastive perspective, as well as examples and glosses as contextual guidance (see 2.2.1.). This article also presents a contrastive way to illustrate the syntactic-semantic local context of the lemmas. The specific and non-specific local context are also indicated in the examples in both languages.

Besides the importance of providing detailed information in the dictionary from a quality standpoint, the users *in-actu* demand this information (cf. Müller-Spitzer et al. 2018; Dominguez et al. 2014). Müller-Spitzer et al. (2018) also demonstrate that users choose the tool according to the number of units they intend to consult: they prefer the dictionary for simple queries (a lexeme), whereas for more complex queries (collocations, phrases and even sentences) they choose machine translation programs, which work much better, if more context is available. This indicates that the dictionaries are not meeting their needs when embarking on complex queries.

4. Capturing and generating contextual data: a lexicographical perspective

In the previous sections we have illustrated the use of *context* in the lexicographic framework at three different levels and have provided examples from different tools. We will now introduce some considerations on the *context at the microstructural level of a dictionary*, that is the level of syntactic-semantic local context (2.1.1.). The main question arises as to how and from what sources lexicographers obtain enough representative, prototypical data for drawing conclusions about linguistic performance. A further question is who systematizes and incorporates these data into different types of dictionaries and platforms, so that these data become comprehensible for the users. Section 4.1. presents methods and tools available for (automatic) extraction of the syntactic-semantic environment, as well as the advantages and disadvantages of some tools for determining the argument pattern (part of the local context) of a lexical unit. In Section 4.2. we will focus on prototypes for generating local contextual data, specifically the syntactic-semantic argument patterns and their examples.

4.1. Overview of sources and methods for data acquisition for lexicographical purposes

Various types of resources are available for observing and extracting local contextual data at microstructural level and finally for supporting the building process of lexicographic resources:

1. **Dictionaries**, especially those which describe the local syntactic-semantic context of different word classes, such as valency dictionaries and dictionaries of collocations (see Section 3).
2. **Corpus tools, corpora and lexical resources**: A wide variety of tools for data gathering and analysis are available. Sketch Engine is an example of a corpus analysis tool; the *BNC*, *CORGA* and *CREA*, among others, are examples of corpora. They offer different methods and possibilities for extracting information according to several parameters. Since many corpora are not annotated with semantic labels (neither argument roles nor ontological features; see Section 2.1.1.), the obtained syntactic argument pattern is not linked to its underlying semantics, and therefore the establishment of the syntactic-semantic interface is more laborious, especially in developing valency-based resources. There are some exceptions, such as the lexical resources *PDEV*, *Verbario*, *PropBank*, *NomBank* and the *Ancora*-Corpora.
3. **Lexical Databases and online lexicons**: WordNet, *FrameNet* or *Ecolexicon* are well-known examples. There are also lexicons available that only describe certain word classes according to their semantic-syntactic linking behaviour, such as *VerbNet* or *VerbAtlas*.

4. It is also possible to retrieve data from different resources, such as **multilingual semantic networks** (for example, the multilingual encyclopaedic dictionary *BabelNet*), or from **collaborative open-source resources** (for example, the multilingual web-based encyclopaedia *Wikipedia*, or *Wiktionary*).
5. **Generators for lexicographical purposes**: Generators are expected to provide meaningful chunks or whole documents (see Section 4.2.), so there is a natural close connection with NLP, NLG and AI, but also with semantics and lexicography (Trap-Jensen 2018; Wanner 2013; Simonsen 2020). Currently different publications and other research outputs deal with the automatic generation of lexicographic content, for example dictionaries (Bardanca 2020; Kabashi 2018), lexicographic articles (Geyken et al. 2017) or dictionary components (the examples in Kosem et al. 2019).

In the last decades, **machine learning strategies** based on neural networks, **statistical language modelling** (such as the N-gram models) and other **neural language models** and **approaches** (such as deep learning models), have revolutionized the field of NLP (Peters et al. 2018; Li et al. 2021) and consequently, directly affected different lexicographical working phases. In particular, we are talking about different predictive methods for (statistical) language modelling and representation, based on different kinds of word embeddings[5] for word prediction. By means of these predictive techniques lexical relations can be represented and encoded. In other words, they enable the capturing of semantic knowledge, as is the case with *Word2vec* (Mikolov et al. 2013) or *fastText* (Bojanowski et al. 2017)[6].

Word2vec (Mikolov et al. 2013) generates contextualized vector representations of a word and draws on a two-level Recurrent Neural Network. When combined with the distributed representation technique CBOW, it tries to predict which word is more appropriate for a specific context (Bardanca 2020), that is, which word frequently fills a slot. It provides the (re)production of the probable distribution of words and hence predicts the contextual framework of a word - understood here in a syntagmatic sense. Data are thus obtained following a semantic-distributional filtering in the sense of Engel (2004: 188) or Firth (1957: 11). Firth has pointed out that 'the meaning of a word is always contextual, and no study of meaning apart from context can be taken seriously' (Firth 1935: 37). With the help of these techniques one can filter out the most frequent local contexts, e.g., in the case of the noun *stay The patient's stay in hospital* as opposed to *The patient's stay in the hotel*. An example of their implementation is shown in Figure 10.

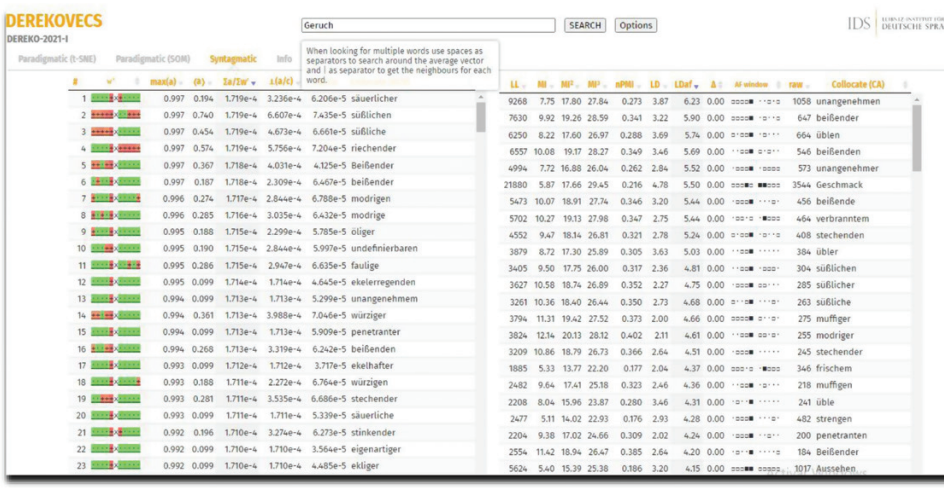


Figure 10. Syntagmatic combinatory potential of *Geruch* from *Derekovecs*.

A further application of *Word2vec* links to Harris's postulate (1954) on the similar meaning of words that occur in the same contexts (understood here in the distributional sense). A possible representation of this approach, applying *Word2vec,* is shown in Figure 11.[7]

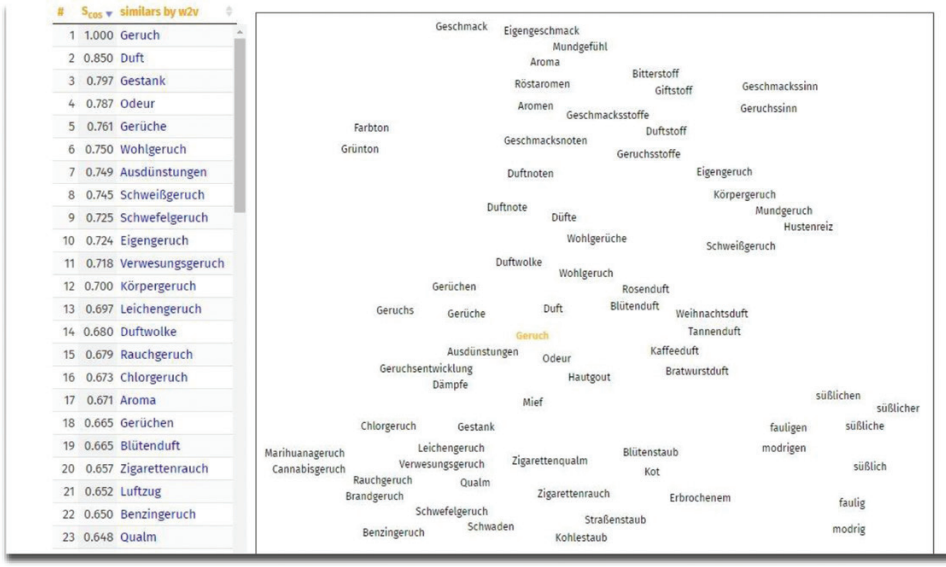


Figure 11. Similar words to *Geruch* from *Derekovecs*

Today there already are neural language models based on word embeddings that aim to describe the sentence context (Peters et al. 2018). The scientific literature highlights the necessity of capturing and modelling the word-external context – the situational context – to ensure the production of coherent utterances (Bender and Koller 2020), as new studies are already doing (Li et al. 2021; Loureiro et al. 2021). How to produce syntagmatic and paradigmatic local context is analysed in Section 4.2.

All the resources mentioned so far already help lexicographers to determine the co-occurrence units on the paradigmatic and syntagmatic axis. As far as corpora in general, it is worth pointing out that they show limitations for establishing syntactic-semantic argument patterns, i.e., the specific and non-specific context of a lexical unit (see Section 2.2.1.). For developing a valency resource, the data required for establishing the local context cannot be based only on statistically filtered results, since one must discriminate between specific and non-specific argument units.[8] Thus, for example, a CQL query in *Sketch Engine* for compounds with the German noun *Frage* (*question*) already yields in the top ten a set of lemmas which does not fulfil the purpose for which the query was originally executed[9]: We are not looking for compounds of *Frage* that may belong to the same semantic field, but we need data about compounds of *Frage* with a first element performing a specific complement. The first elements of compounds cannot be filtered according to semantic roles, as is the case in *Anfängerfrage* (a question of a beginner), *Prüfungsfrage* (an examination question) or *Preisfrage (*a question regarding the price).

Furthermore, some surface realizations do not perform the function of a valency complement, for example: the results from the CQL query for the German noun *Video* in combination with a nominal phrase in genitive cases. The following examples, in this order, show some of the most common occurrences *Video* +*Tag* (video +day), *Video* + *Woche* (video +week), *Video* + *Jahr* (video +year), but none can be considered as a specific argument of the noun *Video*. This implies that, although it is possible to retrieve the patterns and the combinatory potential

of the search term with the *WordSketch* option, the data collected may not always be appropriate to be included in the lexicographic resource we are developing, as is the case here, since we are looking for the specific valency context. It further reveals that statistically measurable data extracted from corpora are not always enough for the linking of meanings, valency argument patterns and surface realizations. The use of word embeddings also does not allow a direct integration of the contextual results in a (valency) dictionary. For example, in the case of adjectives (see Figure 11), the collocational data gathered with *Derekovecs*[10] are of great value, but they do not give any help in delimiting the units according to valency criteria.

All the resources mentioned above have one central feature in common: they show how words are used in actual texts based on corpora. Therefore, on the one hand, we need resources that provide information on the combinatory potential of lexical units and their meanings. On the other hand, collecting data is essential for lexicography, but often the gathered data are not appropriate enough for the resource to be developed. Semi-automatic context generation is another approach that can be explored for this purpose.

4.2. Generating the syntactic-semantic context for lexicographical purposes

Data about the semantic and syntactic local context, as well as the sentence context are key elements for identifying meanings and argument patterns. But until recently there were no resources to generate the predicate argument structure, the combinatory potential of a lexical unit as well as examples. To fill this gap, a set of prototypes – *Xera*, *XeraWord*, *Combinatoria* and *CombiContext* – have been developed. All three simulators describe 60 valency nouns in German, French and Spanish.

The prototype *Xera* provides monolingual argument patterns with examples; *Combinatoria* provides bi-argumental patterns and their examples and the newest generator – *CombiContext* – offers the argument interaction at sentence level. The establishment of argument patterns[11], a semantic description of the arguments (semantic roles and ontological features; see Section 2.2.1.) and the description of the surface realization are common to all generators. From a methodological and theoretical point of view, *CombiContext* draws on the previous generators and predictive methods, such as *fastText* models (see Section 4.1.), and on custom-designed tools, but, obviously, the database has been enriched with verbs and more and new contextual data, in particular adjectives (performing no valency arguments) and adverbs.[12] These prototypes set the stage for the development of interactive, dynamic, user-oriented, and more automatic lexicographic tools, because they automatically produce argument patterns and their examples according to syntactic-semantic pre-selection criteria established by the user. They make it possible to observe which lexical candidates of a semantic class may (or may not) fill a valency slot or be part of the argument pattern within a specific context (see Figure 1), i.e., the specific context. The generators also offer the non-specific context at phrasal and sentence level.

In its design, the main question was how to encode and generate linguistically and computationally the local context of the noun phrase[13], and how to recreate and generate the sentence context[14]. For developing the generators, the *MultiComb* research group follows the proposals of the valency grammar and lexicography of Ulrich Engel (1996, 2004). Consequently, the type of information stored and provided is largely predetermined by the core theory in question. The simulators rely not only on valency and combinatory meaning (Engel 2004), but also on semantic approaches, such as lexical functions (Mel'čuk 2015) and on a new prototype concept (Domínguez 2021). The methodological approach also combines corpus analysis, co-occurrence databases and semantic networks such as *WordNet*, natural language processing tasks (information retrieval and extraction, semantic similarity measures using predictive methods such as *Word2vec*), automatic generation as well as automatic translation of lexical packages (Domínguez et al. 2021). Applying NLP and NLG procedures, often handling self-designed tools (Domínguez et al. 2019), the simulators can map abstract structures onto linguistic surface structures.[15]

As mentioned before, two types of contexts (see 2.1.1.) can be automatically generated in the framework of this prototypes: the specific local syntactic-semantic context and the non-specific local syntactic-semantic context at phrasal level and sentence level (see Figure 15). Once the user has applied her/his search filter, a list is displayed with all the possible combinations that align with the already selected filter. The following is an example for the German noun *Aufenthalt* with its specific local context from the tool *Combinatoria* (see also Figure 12):

(5) *Der Aufenthalt des Patienten im Krankenhaus*
[The patient's stay in hospital]
(6) *Der Aufenthalt der Gastarbeiter in Frankreich*
[The stay of the guest workers in France]
(7) *Der Aufenthalt der Familie in Aachen*
[The family's stay in Aachen]
(8) *Der Aufenhalt in Berlin von Freitag bis Sonntag*
[The stay in Berlin from Friday to Sunday]
(9) *Der Forschungsaufenhalt der Dozentin*
[The research stay of the lecturer]

In the user interface the specific combination of valency arguments at phrasal level can be selected following an ontological filter. The user chooses first the information that should appear in the first position after the valency noun in the nominal phrase and then the information for the second position– in Figure 12 {human profession educational} as the element appearing first and {buildings educational} as the second element.

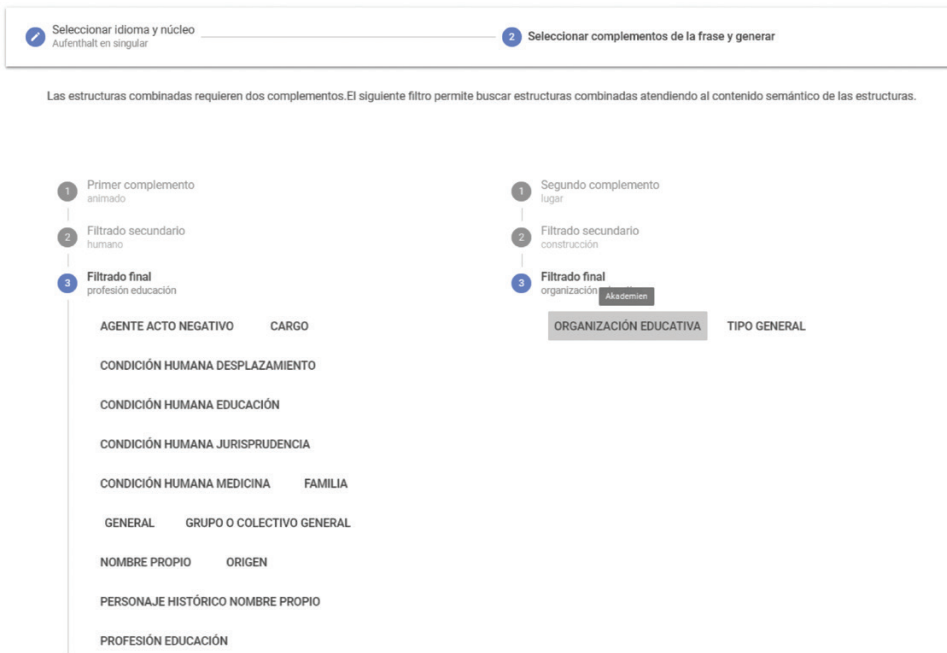


Figure 12. Selection of bi-argumental pattern in *Combinatoria*

Then the tool displays a bi-argumental combination. By clicking on it a list of randomly generated examples appears (Figure 13), that meets the research criteria and can be exported in JSON and CSV formats.

ejemplo	complemento1	complemento2
der Aufenthalt des Dozenten an den Akademien	animado humano profesión educación	lugar construcción organización educativa

frases generadas
Der Aufenthalt des Gastprofessors an der Vorschule
Der Aufenthalt der Gastprofessoren an den Universitäten
Der Aufenthalt des Gastprofessores an den Privatschulen
Der Aufenthalt des Gastprofessors an der Sprachschule
Der Aufenthalt des Englischlehrers an den Mittelschulen
Der Aufenthalt des Dozenten an den Oberschulen
Der Aufenthalt der Erzieherin an den Hochschulen
Der Aufenthalt der Englischlehrerinnen an der Universität
Der Aufenthalt des Dozenten an der Privatschule
Der Aufenthalt der Erzieherinnen an den Hochschulen
Der Aufenthalt des Akademikers an der Vorschule
Der Aufenthalt der Englischlehrerin an der Privatschule
Der Aufenthalt der Akademikerinnen an den Privatschulen
Der Aufenthalt der Akademikerin an der Gastuniversität
Der Aufenthalt der Akademiker an der Musikschule
Der Aufenthalt der Erzieherin an den Oberschulen
Der Aufenthalt des Gastprofessores an den Hochschulen
Der Aufenthalt des Gastprofessores an den Mittelschulen
Der Aufenthalt der Englischlehrerin an den Grundschulen

Figure 13. Randomly generated examples in *Combinatoria*

Currently *Combinatoria* offers more than 20 600 bi-argumental syntactic structures with their examples, which implies an average of 343 combined syntactic-semantic structures per noun. The examples for the syntactic-semantic structures can be generated *ad libitum*. This is because different semantic classes are associated with each syntactic argument and because each semantic class consists of several lexical candidates. An example of this is shown in Table 1, that provides information about some semantic classes and their lexical units connected to the argument ND: *Location* (see also Figure 14).

To achieve this specific variability of context, the prototyping phase also is essential: As it is known, the predicate conditions the semantic roles, their interactions, and the ontological meaning (see Section 2.1.1.). Therefore, it is necessary to get and prototype a list of adequate, prototypical lexical candidates (Domínguez 2021) or a specific lexicon that fits into the concrete valency slot.[16] This phase is carried out on two levels:

(a) After debugging, the data from Sketch Engine prototypical lexical candidates should be identified and classified as specific slot fillers. Their description relies on a custom-tailored ontology[17] (Figure 14).
(b) The establishment of specific semantic classes for filling a specific valency slot, as it is shown in Table 1. In this way, the semantic-ontological combinatory preferences and restrictions of each valency argument and the different surface realizations can be captured.

Table 1. Example for the syntactic-semantic argument *Location* and some semantic classes.

German noun: *Aufenthalt*	semantic classes	example
[determiner + {adjective} + noun + in + argument: `location´]	{proper names of countries}	*Deutschland* [Germany]
	{cities}	*Berlin* [Berlin]
	{territories}	*Schlesien* [Silesia]
[determiner + {adjective} + noun + auf + argument: `location´]	{proper names of island}	*Kuba* [Cuba]
	{landscape}	*der Berg* [the mountain]
[determiner + {adjective} + noun + an + determiner in dative + argument: `location´]	{landscape aquatic}	*die Ostsee* [the Baltic Sea]
	{buildings educational}	*die Universität* [the University]
	{territory part}	*die Grenze* [the border]
[determiner + {adjective} + noun + in + determiner in dative + argument: `location´]	{proper names of countries}	*die Schweiz* [Switzerland]
	{buildings jurisprudence}	*das Gefängnis* [the prison]
	{buildings medicine}	*das Krankenhaus* [the hospital]

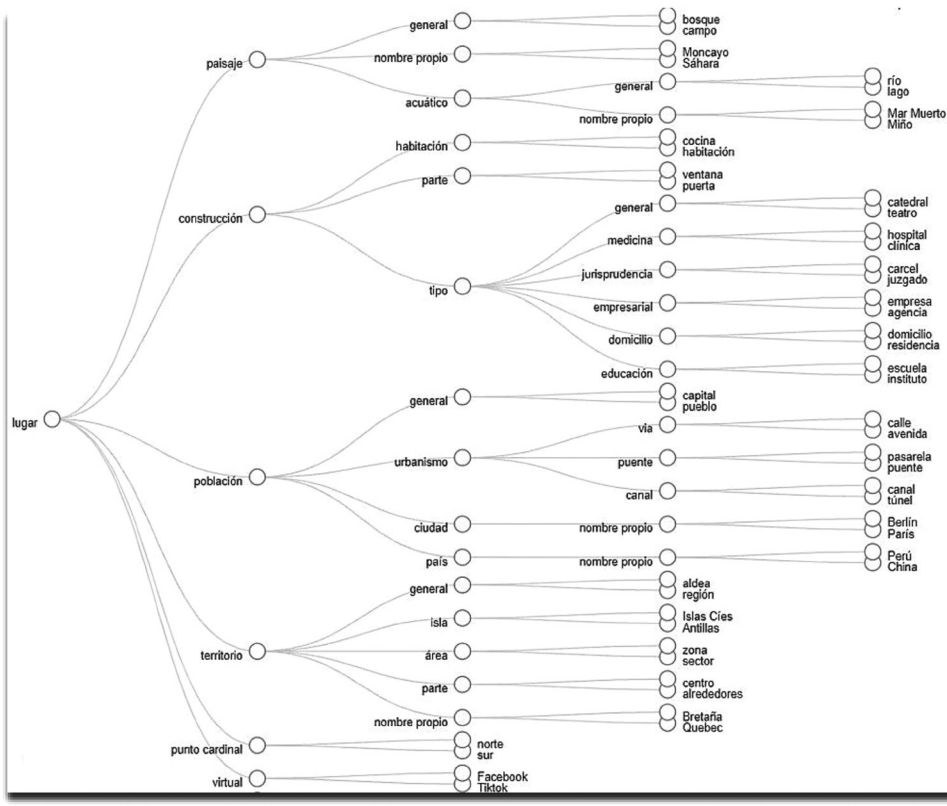


Figure 14. Partial screenshot of the Ontología léxica

The possibility of generating data *ad libitum* also involves the methodological procedure of lexical data expansion and translation. The prototypes resort to the lexical content collected in the different *WordNet* ontologies. In this way it is possible to extract a collection of lexical candidates on the paradigmatic axis, which share their semantic characteristics with the lexical-semantic prototypes that we take as a starting point. As an example: In the case of the semantic class {buildings accommodation} in the mono-argumental structure [determiner + {adjective} + Aufenthalt + in+ determiner in dative + argument: `location´] this expansion procedure enables the extraction of further lexical candidates, e.g., *der Aufenthalt im Hotel* | *in der Gaststätte* | *in der Pension*| *in dem Ferienhaus* | *in der Jugendherberge* | *in dem Luxushotel* | *in dem Gasthaus*. This methodological phase is essential, because the output variability is a quality factor for evaluating language generators (Hashimoto et al. 2019). It is also important from other points of view: semantic variability of a word's contexts positively influences the learning of the word (Brendan et al. 2016), language learners value examples positively (Dziemianko 2012) and it is acknowledged that more examples help learning (Frankenberg 2015).

Since human communication does not usually take place in phrases, but in the form of sentences (or enunciations), the new generator *CombiContext* recreates the most frequent sentence contexts in which the valency nouns usually appear.[18] This is shown in Table 2.

Table 2. Verbal structures

copulative structure	*Der Aufenthalt der Truppen an der Grenze ist notwendig.* [The stay of the troops at the border is necessary]
intransitive structure	*Nach einen zweimonatigen Aufenthalt in Deutschland kehrt er zurück.* [He returns after a two-month stay in Germany]
transitive structure	*Peter plant einen Aufenthalt von März bis Juli in den USA.* [Peter plans to stay in the USA from March to July]
prepositional structure	*Er berichtet von dem Forschungsaufenhalt in Deutschland.* [He reports on the research stay in Germany]

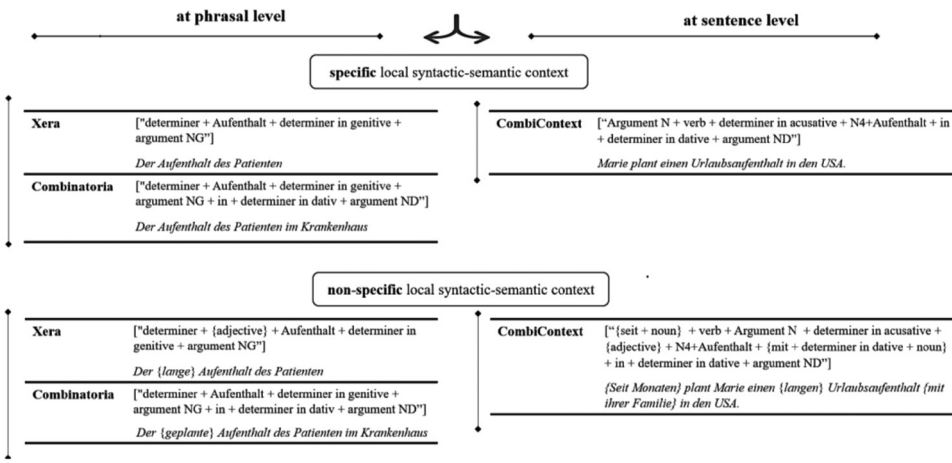


Figure 15. Examples of context information in the prototypes.

Currently *CombiContext* offers approximately 90 000 sentence pattern structures of this nature. The examples and argument combinations can again be generated *ad libitum*, as previously explained for the nominal phrase.

Figure 15 also shows that the generation levels of the wider syntagmatic and paradigmatic context - the non-specific syntactic-semantic context at phrasal and sentence level- is also possible with these tools.

The non-specific local context at phrasal level includes for example non-specific adjectives, that are not essential to ensure the grammaticality of the generated phrase, they are thus welcome for humanizing the automatically generated results. For example:

(10) *der {geplante | notwendige | erste | regelmässige...} Aufenthalt der Dozentin in Deutschland*
[the {planned/ necessary/ first/ regular...} stay of the lecturer in Germany]

At sentence level the generators enable the presentation of non-specific context, such as the following information in brackets:

(11) *{Seit Monaten/Jahren/März} plant sie einen Urlaubsaufenthalt {mit ihrer Familie/mit ihren Freunden/mit der Schule} in den USA.*

[She plans {since months/since years/since March} a holiday *stay* {with her family/her friends/the school} in the USA.]

In summary, the data produced by the generators can have many applications, e.g., these tools could be used in the teaching of foreign languages because the specific context and distribution of words are displayed in the user interface. The user can therefore retrieve information on structural patterns (within information on complements and their case government), but also on the semantic classes and lexical candidates that can fit into a valency slot. In addition, these tools could also be used as independent and stand-alone valency dictionaries or be integrated into learning apps. The syntactic-semantically annotated data as well as examples could also be integrated into other platforms.

5. Conclusion

This article reviews different approaches to the concept of *context* in the field of lexicography. Firstly, this concept has been defined from a metalexicographical perspective, pointing out differences with other concepts such as *cotext*, *contextualization*, *dictionary-internal* and *-external context*. Its presentation in different types of resources has also been addressed. Secondly, an overview of different tools for context data acquisition, namely the local syntactic-semantic context, has been provided. Finally, the paper focuses on new tools created for the automatic generation of syntactic-semantic argument patterns – the local environment of words - and their examples. The tools presented for the local context generation as well as sentence generation are based on a new methodological approach that can be applied for studying other languages and designing new resource types. A good example of further applications is the usability of the paradigmatic lexical packages, that become more granular, turning into the so called *classes d'objets* (Gross 2008). Because they are being codified using open standards, they can be integrated and re-used in other tools. Some explorative research has yielded promising results concerning the application of the generator's data combined with lexical functions for disambiguating meanings (López Iglesias 2020).

In the field of language learning the semantic-conceptual organization of the vocabulary in *Combinatoria* or the formal access in *Xera* and *CombiContext*, as well as the variety of examples enable a way to systematically acquire the combinatory potential of a word and to avoid errors, which are common amongst learners in a production situation (cf. Nied 2014; Müller-Spitzer et al. 2018; Gao and Liu 2020). This cannot be underestimated for developing future studies on the local context.

Acknowledgements

This contribution has been developed within the framework of the research project FFI2017-82454-P financed by Ministerio de Ciencia e Innovación, Agencia (Spain), Estatal de Investigación (Spain) and Fondos de la Unión Europea 'Una manera de hacer Europa' and the Humboldt Research Group of the University of Santiago de Compostela (Xunta de Galicia: ED431B 2020/12).

Notes

1 The main types of lexicographic functions are the communication functions (text reception, text production and translation) and the cognitive function.

2 In this paper the way in which the terms *context* and *cotext* is used in lexicography is acknowledged, but this terminological distinction will not always be maintained. The term *context* is often used in a more comprehensive way to include both context and cotext, as identified in metalexicography.

3 For example: *Deutsches Wörterbuch* (¹DWB), *Deutsches Wörterbuch, Neubearbeitung* (²DWB), *Wörterbuch der deutschen Gegenwartssprache* (WDG), etc.
4 Therefore, many studies underline that the users of lexicographic tools demand an individualized and fast feedback to their query - simple and fast queries, as well as individualized and concrete data acquisition (Spohr 2011; see also Villa Vigogni Thesis, Nr. 11).
5 Word embedding is a technique within natural language processing that allocates a vector to each word. Semantic information is stored in the vectors, and it is possible, to associate or not associate a vector (word) with other vectors by means of the grammatical context. By applying this technique, we can obtain information about the semantic similarity of words. For example:model.wv.most_similar_cosmul (positive=['woman', 'king'], negative=['man']):[('prince', 0.9849334359169006), ('queen', 0.9684107899665833), ('princess', 0.9518582820892334), ('bishop', 0.9380313158035278), ('duke', 0.9368391633033752), ('duchess', 0.9353090524673462), ('victoria', 0.920809805393219), ('mary', 0.9180552363395691), ('mayor', 0.912704348564148), ('prussia', 0.9100297093391418)]Retrieved from: https://www.depends-on-the-definition.com/guide-to-word-vectors-with-gensim-and-keras/
6 *FastText* allows the representation of words not found in the training data and by optimizing *Word2Vec's* algorithm to work on subword elements.
7 In this case one can detect the similarity of the German words *Geruch* (1.000), *Duft* (0.850), *Gestank* (0.797) and *Odeur* (0.787).
8 The first ones are necessary to ensure the grammaticality of the message. The non-specific ones are relevant from a communicative and informative point of view, but not from the grammatical accuracy/correction. This is the well-known classification of complements versus circumstants (or modifiers at phrasal level).
9 This is the case with compounds such as *Anfrage* (Frequency: 311543), *Nachfrage* (201651), *Umfrage* (136169) or *infrage* (31876). For further details, see the CQL query:
10 *Derekovecs* is explained as follows on the website of the Leibniz Institut für Deutsche Sprache: 'DeReKoVecs serves for investigating and comparing of measurements, dimension reduction procedures, visualizations etc., to track down detailed paradigmatic and syntagmatic relations between words based on their use in very large corpora such as the German Reference Corpus DeReKo (Kupietz et al. 2010)'. (From: http://corpora.ids-mannheim.de/openlab/derekovecs/)
11 For this purpose, the simulators rely firstly on the multilingual valency dictionary Portlex (2018) and on frequency and combinatory data from *Sketch Engine*. After data compiling, the specific valency argument within their relational meaning (Engel 2004) and their surface realizations are analyzed and annotated. Some data are debugged according to valency criteria (see some difficulties presented in 3.2.).
12 For achieving sentence-capable lexical packages, an open-source corpus, namely *Wikipedia*, was parsed, and a text chunker and a PoS (Part of Speech) tagger were developed. At the current stage, all data were obtained from Wikipedia text-only dumps (Domínguez et al. 2021).
13 i.e. the active valency or the slots that a noun head opens.
14 i.e. the passive valency of the noun, its relation to other higher units in the dependency hierarchy.
15 For a detailed description of the methodological approaches and the tools used in the development of the language generators see Domínguez et al. (2019) and Domínguez et al. (2021).
16 This is no less relevant, since we know that there are restrictions and preferences. For example, with the verb *essen* (to eat) the subject performing the agent role can be a human, but never an animal. On the contrary, in the case of the verb *fressen* (to eat) in the non-metaphorical sense the agent slot is to be filled by an animal.
17 For the development of the ontology the starting point was the inventory of ontological features of valency grammar and lexicography (Engel 2004) and the WordNet ontologies, of which the synsets are associated with semantic-cognitive features. For a detailed description the categorial features of the valency inventory turned out to be insufficient, because they only allow the identification of general categories or classes (e.g., {human}, {object}, etc.). The granularity of the ontology depends on the need to describe in greater or lesser detail the ontological units that can fill a specific valency slot.
18 It concerns the description of the passive noun valency.

References

1. Dictionaries and tools

AnCora. Multilingual and Multilevel Annotated Corpora. Accessed on 16 November 2022. http://clic.ub.edu/corpus/es/ancora

BabelNet. Multilingual encyclopedic dictionary. Accessed on 16 November 2022. https://babelnet.org/
BNC. The British National Corpus, version 3. Accessed on 16 November 2022. http://www.natcorp.ox.ac.uk/
CombiContext. Prototipo online para la generación automática de contextos frasales y oraciones de la frase nominal en alemán, español y francés. Accessed on 16 November 2022. http://portlex.usc.gal/combinatoria/verbal
Combinatoria. Prototipo online para la generación biargumental de la frase nominal en alemán, español y francés. Accessed on 16 November 2022. http://portlex.usc.gal/combinatoria/usuario
CORGA. Corpus de referencia do galego actual. Accessed on 16 November 2022. CORGA (cirp.gal)
CREA. Corpus de Referencia del Español Actual. Accessed on 16 November 2022. https://www.rae.es/banco-de-datos/crea
Derekovecs. Accessed on 16 November 2022. http://corpora.ids-mannheim.de/openlab/derekovecs/
dict.cc. Deutsch-Englisch Wörterbuch. Accessed on 16 November 2022 https://www.dict.cc/
DiCE. Diccionario de colocaciones del español. Accessed on 16 November 2022. http://www.dicesp.com/paginas
DWDS. Digitales Wörterbuch der deutschen Sprache. Accessed on 16 November 2022. https://www.dwds.de/
EcoLexikon. Accessed on 16 November 2022. http://ecolexicon.ugr.es/en/index.htm
E- VALBU. Das elektronische Valenzwörterbuch deutscher Verben. Accessed on 16 November 2022 https://grammis.ids-mannheim.de/verbvalenz
Flexiona. Accessed on 16 November 2022. http://portlex.usc.gal/develop/flexiona.php
FrameNet. Accessed on 16 November 2022. https://framenet.icsi.berkeley.edu/fndrupal/
FreeLing's dictionaries. Accessed on 16 November 2022. http://nlp.lsi.upc.edu/freeling/node/1
Glosser. Accessed on 16 November 2022. http://www.let.rug.nl/alfa/glosser/Glosser/
Lematiza. Accessed on 16 November 2022. http://portlex.usc.gal/develop/lematiza/
Lexico.com. Accessed on 16 November 2022. https://www.lexico.com/
Longman. Longman Dictionary of Contemporary English Online. Accessed on 16 November 2022. https://www.ldoceonline.com/
Multilingual and Multilevel Annotated Corpora. Accessed on 16 November 2022. http://clic.ub.edu/corpus/es/ancora
NomBank. Accessed on 16 November 2022. https://nlp.cs.nyu.edu/meyers/NomBank.html
Ontología léxica. Accessed on 16 November 2022. http://portlex.usc.gal/ontologia/
OWIDplus. Accessed on 16 November 2022 https://www.owid.de/plus/live-2021/
PDEV. Pattern Dictionary of English Verbs. Accessed on 16 November 2022. https://pdev.org.uk/
Pharos. 2018. Junior Bilingual School Dictionary. Cape Town: Pharos.
Portlex. 2018. Portlex. Diccionario multilingüe de la valencia del nombre. Accessed on 16 November 2022. http://portlex.usc.gal/diccionario/
PropBank. Proposition Bank. Accessed on 16 November 2022. http://verbs.colorado.edu/~mpalmer/projects/ace.html
Prototipo online para la generación automática monoargumental de la frase nominal en alemán, español y francés. Accessed on 16 November 2022. http://portlex.usc.gal/combinatoria/usuario
REDES. 2004. Bosque, I. (dir) 2004. REDES. Diccionario combinatorio del español contemporáneo. Madrid: SM. Accessed on 16 November 2022.
Sketch Engine. Accessed on 16 November 2022. https://www.sketchengine.eu
SKELL. Sketch Engine for Language Learning. Accessed on 16 November 2022. https://www.sketchengine.eu/skell
TraduWord. Accessed on 16 November 2022. https://ilg.usc.gal/gl/proxectos/interoperabilidade-de-recursos-e-producion-automatica-de-linguaxe-natural
V. de Ruiter. 2004. VALBU-Valenzwörterbuch deutscher Verben. Tübingen: Narr.
Verbario. Semántica de los verbos en contexto. Accessed on 16 November 2022. http://www.tecling.com/verbario/
VerbAtlas. Accessed on 16 November 2022. http://verbatlas.org
VerbNet. A Computational Lexical Resource for Verbs. Accessed on 16 November 2022. https://verbs.colorado.edu/~mpalmer/projects/verbnet.html
Villa Vigoni Theses. 2018. Dictionaries for the future - the future of dictionaries: Challenges for lexicography in a digital society [15 Villa Vigoni Theses on Lexicography]. Accessed on 22 April 2022. https://www.emlex.phil.fau.eu/files/2019/03/Villa-Vigoni-Theses-2018-English.pdf
Visuwords. Visual Dictionary. Visual Thesaurus. Interactive Lexicon. Accessed on 16 November 2022. https://visuwords.com/

Wikipedia. Accessed on 16 November 2022. https://www.wikipedia.org
Wiktionary. Accessed on 16 November 2022. https://www.wiktionary.org
WordNet. A Lexical Database for English. Accessed on 16 November 2022. https://wordnet.princeton.edu
Xera. Prototipo online para la generación automática monoargumental de la frase nominal en alemán, español y francés. Accessed on 16 November 2022. http://portlex.usc.gal/combinatoria/usuario
XeraWord. Prototipo de xeración automática da argumentación da frase nominal en galego e portugués. Accessed on 16 November 2022. http://ilg.usc.es/xeraword/en/

B. Other literature

Bardanca Outeiriño, D. 2020. *Automatic generation of dictionaries*. M.A. Thesis, Universidade de Santiago de Compostela.
Bender E. M. and A. Koller. 2020. 'Climbing towards NLU: On meaning, form, and understanding in the age of data' In Jurafsky, D., J. Chai, N. Schluter and J. Tetreault (eds), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, Online, 5-10 July 2020. Association for Computational Linguistics, 5185-5198.
Bojanowski, P., E. Grave, A. Joulin and T. Mikolov. 2017. 'Enriching word vectors with subword information'. *Transactions of the Association for Computational Linguistics* 5: 135-146.
Bolinger, D. 1985. 'Defining the Undefinable' In Ilson, R. (ed), *Dictionaries. Lexicography and language learning*. Oxford: Pergamon Press, 69-73.
Bothma, T., and D. J. Prinsloo. 2013. 'Automated dictionary consultation for text reception. A critical evaluation of lexicographic guidance in linked kindle e-dictionaries'. *Lexicographica: International annual for lexicography* 29: 165-198.
Bothma, T. 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.
Brendan, T. J., M. Dye and M. N. Jones. 2016. 'The influence of contextual diversity on word learning'. *Psychon Bull Rev* 23.4: 1214-1220.
Burkhanov, I. 1998. *Lexicography A Dictionary of Basic Lexicography.* Rzeszów: Wydawnictwon Wyzszej Szkoly Pedagogiccnej.
Busse, D. 2012. *Frame Semantik - Ein Kompendium.* Berlin, Boston: De Gruyter.
Domínguez Vázquez, M. J. 2021. 'Zur Darstellung eines mehrstufigen Prototypbegriffs in der multilingualen automatischen Sprachgenerierung: vom Korpus über word embeddings bis hin zum automatischen Wörterbuch'. *Lexikos* 31: 1-31.
Domínguez Vázquez, M. J., F. Mollica and M. Nied Curcio. 2014. 'Simplex-Verben im Italienischen und Spanischen vs. Präfix- und Partikelverben im Deutschen. Eine Untersuchung zum Gebrauch von online-Wörterbüchern bei der Übersetzung'. In Domínguez, M. J., F. Mollica and M. Nied (eds), *Zweisprachige Lexikographie zwischen Translation und Didaktik*. Berlin: de Gruyter, 191-231.
Domínguez Vázquez, M.J., U. Engel and G. Paredes Suárez. 2017. *Neue Wege zur Verbvalenz. Deutsch-spanisches Valenzlexikon,* Volume 2: Frankfurt: Peter Lang.
Domínguez Vázquez, M. J., M. A. Solla Portela and C. Valcárcel Riveiro. 2019. 'Resources interoperability: Exploiting lexicographic data to automatically generate dictionary examples' In Kosem, I. and T. Zingano Kuhn (eds), *Proceedings of the VI. eLex conference. Electronic lexicography in the 21st century: Smart Lexicography, eLex 2019, Sintra, Portugal, October 1-3, 2019*. Brno: Lexical Computing CZ s.r.o, 51-71.
Domínguez Vázquez, M. J., D. Bardanca Outeiriño and A. Simões. 2021. 'Automatic Lexicographic Content Creation for Lexicographers' In Kosem, I., M. Cukr, M. Jakubíček, H. Kallas, S. Krek and C. Tiberius (eds), *Electronic lexicography in the 21st century: post-editing lexicography. Proceedings of the eLex 2021 conference, eLex 2021, virtual, July 5-7, 2021*. Brno: Lexical Computing CZ s.r.o., 269-287.
Dziemianko, A. 2012. *Noun and Verb Codes in English Monolingual Dictionaries for Foreign Learners: A Study of Usefulness in the Polish Context*. Poznan: Adam Mickiewicz University Press.
Engel, U. 1996. 'Semantische Relatoren. Ein Entwurf für künftige Valenzwörterbücher' In Weber, N. (ed), *Semantik, Lexikographie und Computeranwendung*. Tübingen: Niemeyer, 223-236.
Engel, U. 2004. *Deutsche Grammatik-Neubearbeitung*. München: Iudicium.
Faber, P. and P. León-Araúz. 2016. 'Specialized Knowledge Representation and the Parameterization of Context'. *Frontiers in Psychology* 7: 1-20. Accessed on 16 November 2022. https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00196/full

Fillmore, C. J. 1977. 'Scenes-and-Frames Semantics', In Zampolli, A. (ed), *Linguistic Structures Processing*. Amsterdam, New York and Oxford: North-Holland, 55-81.
Firth, J.R. 1935. 'The technique of semantics'. *Transactions of the Philological Society* 34: 36-73.
Firth, J. R. 1957. 'A synopsis of linguistic theory 1930-1955', *Studies in Linguistic Analysis*, (special Volume of the Philological Society). Oxford: Blackwell, 1-31.
Frankenberg-Garcia, A. 2015. 'Dictionaries and Encoding Examples to Support Language Production'. *International Journal of Lexicography* 28.4: 490-512.
Frankenberg-Garcia, A., R. Lew, J. Roberts, G. 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.
Gao, J. and L. Haitao. 2020. 'Valency Dictionaries and Chinese Vocabulary Acquisition for Foreign Learners'. *Lexikos* 30: 111-142.
Geyken, A., F. Wiegand and K. M. Würzner. 2017. 'On-the-fly Generation of Dictionary Articles for the DWDS Website' In Kosem, I., C. Tiberius, M. Jakubíček, J. Kallas, S. Krek and V. Baisa (eds.), *Electronic lexicography in the 21st century: Proceedings of eLex 2017 conference, Leiden, the Netherlands, 19–21 September 2017*. Brno: Lexical Computing CZ s.r.o., 560-570.
Gouws, R. H. 1988. 'Die gebruik van etikette as leksikografiese hulpmiddel'. *SA Tydskrif vir Taalkunde Geleentheidsuitgawe* 6: 1-56.
Gouws, R. H. 2015. 'Expanding the notion of addressing relations'. *Lexicography* 1.2: 159-184.
Gouws, R. H. 2021. 'Expanding the use of corpora in the lexicographic process of online dictionaries' In Piosik, M., J. Taborek and M. Woźnicka (eds), *Korpora in der Lexikographie und Phraseologie*. Berlin: De Gruyter, 1-20.
Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography.* Stellenbosch: African SUN MeDIA.
Gouws, R.H. and S. Tarp. 2019. 'Lexicographical Contextualization and Personalization: A New Perspective'. *Lexikos* 29: 250-268.
Gross, G. 2008. *Les classes d'objets*. Paris: Presses de l'Ecole normale supérieure.
Harris, Z. 1954. 'Distributional structure'. *Word* 10.23: 146-162.
Hartmann, R.R. K. and G. James. 1998. *Dictionary of Lexicography*. London: Routledge.
Hashimoto, T.B., H. Zhang and P. Liang. 2019. 'Unifying human and statistical evaluation for natural language generation' In Burstein, J., C. Doran and T. Solorio (eds), *Proceedings of the 2019 Conference of the North American Association for Computational Linguistics: Human Language Technologies, Minneapolis, USA, June 2-7, 2019*. Minneapolis: Association for Computational Linguistics, 1689-1701.
Hausmann, F.J. and H.E. Wiegand. 1989. Component Parts and Structures of General Monolingual Dictionaries. In Hausmann, F.J., O. Reichmann, H.E. Wiegand and L. Zgusta (eds.), *Wörterbücher. Dictionaries. Dictionnaires*. Berlin: De Gruyter, 329-360.
Herbst, T., Heath D., Roe I. F. and D. Götz. 2004. *A Valency Dictionary of English: a corpus-based analysis of the complementation patterns of English verbs, nouns and adjectives*. Berlin: De Gruyter.
Heringer, H. J. 1984. 'Neues von der Verbszene' In Stickel, G. (ed), *Pragmatik in der Grammatik: Jahrbuch 1983 des Instituts für deutsche Sprache*. Düsseldorf: Schwann, 34-64.
Hermanns, F. and W. Holly. 2007. *Linguistische Hermeneutik. Theorie und Praxis des Verstehens und Interpretierens*. Berlin: De Gruyter.
Horák, A. and A. Rambousek. 2017. 'Lexicography and natural language processing' In Fuertes-Olivera, P. A. (ed), *The Routledge Handbook of Lexicography*. Nueva York: Routledge, 179-196.
Kabashi, B. 2018. 'A lexicon of Albanian for natural language processing'. In Čibej, J., V. Gorjanc, I. Kosem and S. Krek (eds), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts, EURALEX 2018, Ljubljana, July 17-21*. Ljubljana: Ljubljana University Press, 855-862.
Kammerer, M. and H. E. Wiegand. 1998. 'Über die textuelle Rahmenstrutkur von Printwörterbüchern. Präzisierungen und weiterführende Überlegungen'. *Lexicographica* 14: 224-238.
Kosem, I., K. Koppel, T. Zingano Kuhn, J. Michelfeit and C. Tiberius. 2019. 'Identification and automatic extraction of good dictionary examples: the case(s) of GDEX'. *International Journal of Lexicography* 32.2: 119-137.
Krek, S. 2019. 'Natural Language Processing and Automatic Knowledge Extraction for Lexicography'. *International Journal of Lexicography* 32.2: 115-118.

Kühn, P. 1989: 'Typologie der Wörterbücher nach Benutzungsmöglichkeiten' In Hausmann, F. J., G. Ungeheuer, H. Steger, H. E. Wiegand and A. Burkhardt (eds), *Wörterbücher: ein Internationales Handbuch zur Lexikographie*, Vol. 1. Berlin, New York: De Gruyter, 111-127.

Laufer, B., and P. Nation. 2012. 'Vocabulary' In Gass, S. M. and A. Mackey (eds), *The Routledge Handbook of Second Language Acquisition*. London, New York: Routledge, 163-176.

Lettner, K. 2020. *Zur Theorie des lexikographischen Beispiels*. Berlin: De Gruyter.

Li, B. Z., M. Nye and J. Andreas. 2021. 'Implicit Representations of Meaning in Neural Language Models' In Zong, C., F. Xia, W. Li and R. Navigli (eds), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, ACL-IJCNLP 2021, virtual, August 1-6, 2021, Vol.1. Long Papers*. Online: Association for Computational Linguistics, 1813-1827.

López Iglesias, N. 2020. *Analysing nominal phrase contexts for the automatic extraction of linguistic and lexicographic data*. M.A. Thesis, Universidade do Minho.

Loureiro, D., M. J. Alípio, and J. Camacho-Collados. 2021. 'LMMS Reloaded: Transformer-based Sense Embeddings for Disambiguation and Beyond'. [Preprint]. *Artificial Intelligence Journal*, 1-84. Accessed on 16 November 2022. https://arxiv.org/pdf/2105.12449.pdf

McConchie, R. and J. Tyrkkö. 2018. *Historical Dictionaries in their Paratextual Context*. Berlin: De Gruyter.

McDonald, S. and M. Ramscar. 2001. 'Testing the distributional hypothesis: The influence of context on judgments of semantic similarity' In Moore, J. D. and K. Stenning (eds), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society, CogSci 2001, Edinburgh, Scotland, August 1-4, 2001*. London: LEA, 611-616.

Mel'čuk, I. 2015. *Semantics. From meaning to text*, Vol 3. Amsterdam/Philadelphia: John Benjamins.

Mikolov, T., K. Chen, G. Corrado and J. Dean. 2013. 'Efficient estimation of word representations in vector space' In Bengio, Y. and Y. LeCun (eds), *Proceeding of the International Conference on Learning Representations Workshop Track, ICLR 2013, Scottsdale, USA, May 2-4, 2013*. Arizona: Conference Track Proceedings, 1-12.

Müller-Spitzer, C. 2013. 'Contexts of dictionary use' In: Kosem, I., P. Gantar, S. Krek, M. Langemets, M. Tuulik (eds.), *Electronic lexicography in the 21st century: thinking outside the paper. Proceedings of the eLex 2013 conference, Tallinn, Estonia, 17-19 October 2013*. Ljubljana/Tallinn: Institute for Applied Slovene Studies/Eesti Keele Instituut, 1-15.

Müller-Spitzer, C., M. Nied Curcio, M. J. Domínguez Vázquez, I. ilva Dias and S. Wolfer. 2018. 'Rechercheprаxis bei der Verbesserung von Interferenzfehlern aus dem Italienischen, Portugiesischen und Spanischen: Eine explorative Beobachtungsstudie mit DaF-Lernenden'. *Lexicographica* 34.1: 157-182.

Moreno Jiménez, L. G., J. M. Torres-Moreno, R. S. Wedemann and E. SanJuan. 2020. 'Generación automática de frases literarias'. *Linguamática* 12.1: 15-30.

Nallapati, R., Z. Bowen, C. dos Santos, Ç. Gulçehre and B. Xiang. 2016. 'Abstractive text summarization using sequence-to-sequence RNNs and beyond' In Riezler, S. and Y. Goldberg (eds), *Proceedings of the 20th SIGNLL conference on computational natural language learning, SIGNLL 2016, Berlin, Germany, August 11-12, 2016*. Berlin: Association for Computational Linguistics, 280-290.

Nation, P. 2005. 'Teaching and learning vocabulary' In Hinkel, E. (ed) *Handbook of research in second language teaching and learning*. New York: Routledge, 581-595.

Navigli, R. and S. P. Ponzeto. 2012. 'BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network'. *Artificial Intelligence* 193: 217-250.

Nied, M. 2014. 'Die Benutzung von Smartphones im Fremdsprachenerwerb und -unterricht' In Abel, A., C. Vettori and N. Ralli (eds), *Proceedings of the XVI EURALEX International Congress: The User in Focus, EURALEX 2014, July 15-19, 2014. Bolzano*. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism, 263-280.

Otter, D. W., J. R. Medina and J. K. Kalita. 2020. 'A Survey of the Usages of Deep Learning for Natural Language Processing'. *IEEE Transactions on Neural Networks and Learning Systems* 32.2: 604-624.

Pereira, F., B. Lou, B. Pritchett, S. Ritter, S. H. Gershman, N. Kanwisher, M. Botvinick and E. Fedorenko. 2018. 'Toward a universal decoder of linguistic meaning from brain activation'. *Nature communications* 9: 1-13.

Peters, M., M. Neumann, M. Iyyer, M.Gardner, C. Clark, K. Lee and L. Zettlemoyer. 2018. 'Deep contextualized word representations' In Walker, M., H. Ji and A. Stent (eds), *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human*

Language Technologies, NAACL 2018, New Orleans, USA, June 1-6, 2018, Vol.1. Long Papers. New Orleans, Louisiana: Association for Computational Linguistics, 2227-2237.

Roemmele, M. 2016. 'Writing stories with help from recurrent neural networks'. In Magerko, B., J. C. Bahamón, M. Buro, R. Damiano, J. Mazeika, S. Ontañón, J. Robertson, J. Ryan and K. Siu (eds), *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, AAAI-16, Phoenix, USA, February 12-17, 2016*. Phoenix, USA: AAAI Press, 4311-4312.

Seretan, V., and E. Wehrli. 2013. 'Context-sensitive look-up in electronic dictionaries' In Gouws, R. H., U. Heid, W. Schweickard and H. E. Wiegand (eds), *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin, Boston: De Gruyter, 1046-1053.

Simonsen, H. K. 2020. 'Augmented Writing Needs Lexicography: A Symbiotic Relationship? In Gavriilidou, Z., M. Mitsiaki and A. Fliatouras (eds), *Proceedings of XIX EURALEX Congress: Lexicography for Inclusion, EURALEX 2020, Alexandropoulis, Greece, September 8-12, 2020, Vol.1*. Ljubljana: Ljubljana University Press, 509-514.

Smith, C. 1991. *The Parameter of Aspect*. Dordrect: Kluwer Academic Publishers.

Sordoni, A., M. Galley, M. Auli, C. Brockett, Y. Ji, M. Mitchell, J. Nie, J. Gao and B. Dolan. 2015. 'A neural network approach to context-sensitive generation of conversational responses' In Mihalcea, R., J. Chai, and A. Sarkar (eds), *Human Language Technologies: The 2015 Annual Conference of the North American Chapter of the ACL*. Denver: Association for Computational Linguistics, 196-205.

Spohr, D. 2011. 'A multi-layer architecture for 'Pluri-Monofunctional' dictionaries' In Fuertes-Olivera, P. A. and H. Bergenholtz (eds), *e-Lexicography: The Internet, Digital Initiatives and Lexicography*. London: Bloomsbury Academic, 103-120.

Tarp, S. 2000. 'Theoretical challenges to practical specialised lexicography'. *Lexikos* 10: 189-208.

Tarp, S. 2008. *Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learner's lexicography*. Tübingen: Niemeyer.

Tarp, S. and R. H. Gouws. 2019. 'Lexicographical Contextualization and Personalization: A New Perspective'. *Lexikos* 29: 250-268.

Trap-Jensen, L. 2018. 'Lexicography between NLP and Linguistics: Aspects of Theory and Practice' In Čibej, J., V. Gorjanc, I. Kosem and S. Krek (eds), *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, 17-21 July, Ljubljana. Ljubljana: Ljubljana University Press, 25-37.

Treffers-Daller, J. and V. Rogers. 2014. 'Grammatical Patterns and the L2 lexicon' In Milton, J. and T. Fitzpatrick (eds), *Dimensions of Vocabulary Knowledge*. London: Palgrave, 106-122.

Wanner, L. 2013. 'Electronic dictionaries for Natural Language Generation (NLG)', In Gouws, R. H., U. Heid, W. Schweickard and H. E. Wiegand (eds), *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin, Boston: De Gruyter, 1150-1161.

Wiegand, H. E. 1987. 'Zur handlungstheoretischen Grundlegung der Wörterbuchbenutzungsforschung'. *Lexicographica* 3: 178-227.

Wiegand, H. E. 1988. 'Wörterbuchartikel als Text' In Harras, G. (ed), *Das Wörterbuch. Artikel und Verweisstrukturen. Jahrbuch des Instituts für deutsche Sprache 1987*. Düsseldorf: Schwann, 30-120.

Wiegand, H.E. 1998. *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. Berlin: De Gruyter.

Wiegand, H. E. and R. H. Gouws. 2013. 'Addressing and addressing structures in printed dictionaries' In Gouws, R. H., U. Heid, W. Schweickard and H. E. Wiegand (eds), *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin, Boston: De Gruyter, 273-314.

Wiegand, H. E., R.H. Gouws, M. Kammerer, M. Mann and W. Wolski. 2017. *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research*, Volume 3: Berlin: De Gruyter.

Zgusta, L. 1971. *Manual of Lexicography*. Berlin, New York: De Gruyter Mouton.

Ziem, A. 2008. *Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz*. Berlin, New York: De Gruyter.