

Systemising corpus-based definitions in second language lexicography*

Summary

This study explores the application of a semantic analysis methodology to the creation of definitions of learners' dictionaries. Defining words is a delicate art, traditionally left to the general writing skills of the lexicographer, introspection, or even intuition. In the past decades, many efforts have been guided to systemise the technique, particularly in the pedagogical dictionaries movement. In this paper, we try to demonstrate that there is still room for improvement, and that a systematic corpus analysis can be applied to build better explanations for the meaning of words. We explain the theoretical background chosen for our study, the associated methodology and five specific strategies which can direct the lexicographer through his/her task. In order to give a concrete example, we show our work with a Spanish dictionary project for foreign learners which is currently under development and has a core of very frequent Spanish verbs already available on the Internet.

Key words: *corpus linguistics, full-sentence definition, lexico-syntactic patterns, semantic analysis*

Sistematização de definições baseadas em corpus em lexicografia de segundas línguas

Resumo

Este estudo explora a aplicação de uma metodologia de análise semântica para a criação de definições destinadas aos dicionários de aprendizagem. Definir palavras é uma arte complicada, que tradicionalmente foi legada às capacidades gerais de redação do lexicógrafo, à introspecção ou mesmo à intuição. Nas últimas décadas, muitos esforços foram realizados para sistematizar a técnica, nomeadamente com a eclosão dos dicionários didáticos. Neste artigo, tenta-se demonstrar que ainda há lugar para a melhora, e que a análise sistemática de corpus pode aplicar-se à construção de melhores explicações para o significado das palavras. Explicam-se os antecedentes teóricos selecionados para o nosso estudo, a metodologia associada e cinco estratégias específicas que podem guiar o lexicógrafo no seu trabalho. Para proporcionar um exemplo concreto, mostramos a nossa experiência com um projeto de dicionário para estudantes de espanhol como língua estrangeira, atualmente em desenvolvimento mas que mostra já um núcleo de verbos altamente frequentes do espanhol disponíveis on-line.

Palavras-chave: *análise semântica, definição fraseológica, linguística de corpus, padrões léxico-sintáticos*

1. Introduction

It is well established in lexicographical tradition of monolingual pedagogical dictionaries that these dictionaries must be tools for learning a language, and not only resources to solve concrete doubts in a quick, practical way during the process of decoding or encoding (Bogaards, 2010; Atkins and Rundell, 2008: 406–411). This position has been defended from theoretical as well as methodological points of view, and many dictionaries of different languages have been edited following this general consensus. As an example, the *Collins Cobuild English Language Dictionary* (Sinclair 1987), the *Oxford Advanced Learners Dictionary of Current English* (Hornby and Cowie 1963), and other pedagogical dictionaries of English can be included in this category.

* This paper is part of the project [ANONYMIZED].

Specifically, there is a fundamental aspect which seems to make monolingual dictionaries useful tools for learners: they include information of the grammar-lexicon interface in a way it does not seem to cover neither grammar nor traditional dictionaries. In this sense, electronic dictionaries can improve substantially the proposals of paper dictionaries (DeSchryver 2003). For example, in Figure 1 it is shown one meaning for *jouer* ‘to play’ in *Dictionnaire d’apprentissage du français langue étrangère ou seconde*, DAFLES (Verlinde and Selva 2006). Full-sentence definition provides information about meaning (...*pratique un sport où l’on utilise une balle, un ballon*), syntax (*une personne joue au foot...* is a transitive pattern the definition shows) and collocations (*jouer au foot, au basket, au rugby...*). Examples are rich in syntactical information: in the first one, transitive pattern of the definition is repeated (*jouer au foot*), and the second one provides the alternative passive pattern with *se* particle (*le foot se joue aussi en salle*).

Figure 1. Meaning 3a of the verb *jouer* in DAFLES.

As can be observed in the previous example, definitions are one of the main parts of the dictionary entry in which grammar-lexicon interface can be shown (Bosque 2006: 47–48). Classically considered as explanations of the meanings of words (Johnson 1755), they have become also patterns of syntactic and semantic behaviour of the word in context. In concrete, full-sentence definitions, firstly used in the *Cobuild* dictionary (Sinclair 1987), as DAFLES’ in Figure 1 are those which include the *definiendum* in the definition: *lorsque un personne joue au foot...* There are also rhetorical resources to introduce the explanations, such as *lorsque* in the previous example or *If a person plays a sport...*, commonly used in the *Cobuild* dictionary. Another typical beginning for a full-sentence definition of verbs are the so called *when-definitions* (Lew and Dziemianko 2006; Adamska-Salaciak 2012): *When somebody plays a sport...* These formulae pretend to be a more natural way to explain what is the meaning of a word, and, as said already, they allow to show the syntactic pattern of the verb meaning.

Full-sentence definitions have been defended and studied from different perspectives. Hanks (1987) explains that these kind of definitions permit to provide a more precise information related to syntax and collocations. Harvey and Yuill (1997), in a user study, concluded that full-definitions were helpful for encoding tasks. As Rundell (2006) points out, full-sentence definitions, also called folk definitions, reflect the natural way in which a teacher describes what a word means, in order to both make the definitions accessible to a learner of the language, as they sound natural to the user, and to show meaning in context. At the same time, they are not exempt of controversy: the same author urges to be prudent with their use, because the quantity of semantic and syntactic information they provide can diminish its clarity. Full-sentence definitions are larger and more complex indeed than the

classical ones, and this could be the reason, Rundell (2006) explains, why they are not a generalised lexicographic practice.

The present paper starts from the work engaged by lexicographers and linguists we briefly showed in the previous lines. We consider that there has not been yet a specific proposal about how full-sentence definitions are to be built, despite the fact that they have been used and considered useful for learners in different ways. The debate about folk definitions is, nevertheless, part of the long discussion about how to explain word meanings, which has been from Aristotle to the present. The art of defining a word in a clear, rigorous way is fine and delicate, and traditionally has been addressed with the help of the general writing skills of the lexicographer, introspection or even intuition. In the specific context of language learning, two basic problems must be tackled: a) establishing the criteria for adapting the semantic and syntactic information extracted from the corpus, and b) establishing the criteria for explaining this information in a didactic way, in order to make all these data more comprehensible to a foreign user.

In this study, we will focus on verb definitions in a pedagogical dictionary, specifically, an online dictionary of Spanish for foreign learners, the *Diccionario de aprendizaje del español como lengua extranjera*, DAELE (Arias-Badia et al., in press). We deal with verbs as verbs facilitate the observation of the grammar-lexicon interface. In relation to the methodology, we adopt Corpus Pattern Analysis, CPA (Hanks 2004a) as the technique for corpus analysis, because this methodology was specially created for lexical analysis and for lexicographical purposes. Thus, in the following pages, we argue why a systematic corpus analysis is needed in front of the traditional introspective model (Section 2) and we make a brief introduction to CPA and its theoretical background (Section 3). Section 4 is devoted to describe our proposal which connects CPA with the full-sentence definitions of a Spanish dictionary for foreign students. Finally, in Section 5 we draft some conclusions and lines for future work.

2. Definitions in corpus-driven dictionaries vs classical dictionaries

In this section, we show some examples of classical, non corpus-based definitions, in order to justify why learners' dictionaries could benefit from corpus analysis procedure and from more systematic definitions. Two key questions are strongly connected: on the one hand, the need of having a system to make corpus driven dictionaries, and, on the other hand, the possibility of offering more rigorous grammatical information to the user, systematically linked to the semantic information offered by the dictionary.

As already said, traditional dictionaries were (and in most cases still are) based on introspection and previous dictionaries (Sinclair 1991: 37–41), and traditional lexicographer's task was to try to discover the inherent meanings of words. However, corpus data have shown that this system is not capable of producing a satisfactory description of the normal patterns of use of words. To show it in a simple way, traditional lexicographers ask themselves 'What does *arrive* mean?' or 'Which are the different meanings of *arrive*?', in contrast with corpus-driven questions about meanings in language, such as 'What does *arrive* mean *in this context*?' (see Section 3 for a review of the theoretical postulates underlying this change of perception). Since the 'corpus revolution' and its application to lexicography in the first corpus-based dictionary, *Cobuild* (Sinclair 1987), the use of a corpus for building dictionaries became a *sine qua non* condition in the state-of-the-art. Even if the task can be done without this type of analysis, corpus analysis has proven to improve lexicographical work in many different ways. See, for example,

these meanings of different entries in *Diccionario Salamanca de la lengua española* (Gutiérrez Cuadrado 1996)²:

estallar³ v. intr. [...] **5** Manifestar < una persona > [una emoción o un sentimiento] de repente y con fuerza: *El muchacho estalló en sollozos. / Parecía que iba a estallar de emoción cuando le dieron el premio.*

sorprender v. tr. **1** Causar < una persona o una cosa > sorpresa [a una persona]: *Me sorprende con esa pregunta.*

coser v. tr. [...] **5** Causar < una persona > [muchas heridas] [a otra persona]: *Yo vi el cadáver en el depósito, y lo habían cosido a balazos.*

In the first case, two different patterns of the verb *estallar* ‘to burst’ have been included in the same meaning: *estallar + en + noun* and *estallar + de + noun*. If we look up in a corpus⁴, two different groups of concordances, according to the complement, can be observed: a) *estallar + en + abucheos, gritos, aplausos, blasfemias, cánticos, carcajadas, llanto...*; b) *estallar + de + alegría, gratitud, furia, ira...*

Group a) consists basically of external expressions of feelings which manifest themselves through noise (such as cries, shrieks, or guffaws); group b) are intense feelings like joy or anger. Thus, it would not be accurate, in the most frequent use of these two patterns, to combine these two groups, such as **estallar de llanto* or **estallar en alegría*. If both structures are combined in the same meaning in a dictionary, would not be possible for a learner to predict how to use them.

In the case of the verb *sorprender* ‘to surprise’, if we look into the IULA50 corpus (see note 2), 167 concordances denoting meaning 1 of *Salamanca*, but approximately 10% of them are complemented by a clause, as in the following sentence: *Me sorprende que la detención de Isabel Pantoja sea objeto de controversia política* (‘It surprises me that the arrest of Isabel Pantoja is an object of political controversy.’). Thus, a very common use of this meaning is not indicated in the entry, so the learner can have doubts about the correct use of this structure.

Finally, the fifth meaning of the verb *coser* (‘to riddle’) is defined as ‘to cause injuries’. If we look up this word in the corpus, we find the following group of complements: *coser + a + balazos, codazos, patadas, porrazos, puñaladas...* (‘bullet wounds, elbows, kicks, blows, knife wounds...’). That is, the action is not only restricted to ‘injuries’, but to many types of aggressions a person can inflict on another, for example with elbows, legs, or

² Despite the lack of precision found in some of the entries, we have considered this dictionary because it is one of the most complete, rigorous traditional Spanish dictionaries for foreign learners currently available.

The English version of the examples is intended to be literal word-for-word translations of the Spanish entries. However, it is not always possible. For instance, in the case of *coser*, the literal equivalent in English would be ‘to sew’. But for sense 5, it is not correct to give ‘to sew’ as an equivalent. In English, the verb used in this case is ‘to riddle’. It must be highlighted, however, that ‘to riddle’ activates a slightly different conventional metaphor than the one activated by the Spanish verb *coser*.

³ Translations of the three entries are (respectively): **burst** v. intr. [...] **5** To manifest < somebody > [an emotion or feeling] suddenly and strongly: *The boy burst into tears. / It seemed that she was going to burst with joy when they gave her the prize.* **surprise** v. tr. **1** <Of a person or thing> to cause [somebody] to feel surprise *The question surprised me.* **riddle** v.intr. [...] **5** To cause [a lot of injuries] [to another person]: *I saw the corpse in the morgue, and it was riddled with bullets.*

⁴ For this analysis, we use the IULA50 corpus, consisting of 50 million words of press articles linked with the Spanish CPA project (AUTHOR 2012: 185-186).

truncheons. Furthermore, all complements are in plural (it is not possible to say ...*lo habían cosido a *balazo*), and this is not illustrated in the entry.

To summarize this section, we conclude that many aspects of the real usage may be omitted from a traditional dictionary entry, and these omissions could cause confusion among students. A corpus-driven approach offers clues to make the dictionary closer to the users' needs and makes possible to offer data connected with normal and real use of a given word.

3. Theoretical and methodological framework: the Theory of Norms and Exploitations and Corpus Pattern Analysis

In the previous section, we showed that corpus analysis methodology can be used for improving the quality and quantity of the information offered to the learner in a dictionary. The present section is devoted to make a general presentation of CPA, the specific methodology for corpus analysis we choose for the present proposal. CPA is theoretically supported by the Theory of Norms and Exploitations – henceforth, TNE – (Hanks 2004b, 2013). We will briefly present the main postulates of TNE and how they are connected to CPA. We will also show some samples of practical work with CPA in English and Spanish.

3.1. TNE, a theory for explaining how meaning is created through words

TNE is a lexically based and corpus-driven approach, whose main objective is to describe how speakers use words to make meanings. In this theory, it is postulated that, on the one hand, words are used in normal lexico-syntactic patterns: in TNE, a pattern is defined as ‘a semantically motivated and recurrent piece of phraseology’ (Ježek and Hanks 2010: 8). Each normal pattern is associated with a unique meaning. On the other hand, norms may also be ‘exploited’ for rhetorical or other effect. As Hanks (2013: 211–215) states ‘an exploitation is a dynamic mechanism in language to create new meanings ad hoc and to say old things in new ways’. Anomalous collocations, ellipsis, creative linguistic metaphors and similes, as well as other creative figures of speech are examples of exploitations. In relation to dictionaries, lexicographers have a duty to describe norms, but to ignore exploitations, though the dividing line between a normal use of a word and exploitations of that norm may be fuzzy (Hanks 2013: 16).

In order to briefly exemplify how this double system of norms and exploitations works, we consider the intransitive verb *to arrive*. In its most frequent use, it has the grammatical structure *subject + arrive + at + complement*, but this is not sufficient for establishing the difference between ‘He arrived at the house’ and ‘Jane and I quickly arrived at joint decisions about the project’, two sentences that are syntactically identical but semantically different. Thus, it is the semantics of the structure valency of the verb and not the verb in isolation, which gives evidence of the specific meaning for these instances of *arrive*. The two patterns exemplified above can be formalised as follows⁵:

Pattern 1	[[Human Vehicle]] arrive [NO OBJ] {(at [[Location]])}
<i>Implicature</i>	[[Human Vehicle]] comes to [[Location]] after a journey.
Pattern 2	[[Human Institution]] arrive [NO OBJ] {at [[Concept = Considered Opinion]]}
<i>Implicature</i>	[[Human Institution]] adopts [[Concept = Considered Opinion]] after a process of long and careful thought and/or discussion.

⁵ See the *PDEV* (Hanks, in progress) for the whole analysis of this verb. The *PDEV* is available at <http://www.pdev.org.uk/#browse?q=&f=C> (last access: 10/11/2014).

As what regards to CPA's annotation, semantic types are written in double squared brackets, complements between curly brackets, one of them between parenthesis meaning its is optional. The [NO OBJ] indication blocks the possibility of a direct or indirect object complementing the verb. In italics, there is the *implicature*, a paraphrasis or explanation of the conventional meaning connected to the pattern.

TNE has its origin in the work of a large number of authors who focused on the study of the lexical unit and its connection with the context in which it is used. It is especially relevant the theory of Sinclair (1991, 1999), who was in turn influenced by the work of Firth (1957) and Halliday (1976). The pioneer work of Hornby (1954) in dictionaries for foreign learners was also fundamental, and Hunston and Francis (2000) made also contributions to the study of grammar patterns. Following these authors, the semantics of words is linked to the context. Sinclair (1999) argues that words have their meanings in context and not in isolation, and by 'context' we mean not only the syntactic structure, but also collocations.

Nevertheless, in all the previous approaches there is still not a formalised methodology of mapping meaning onto use, in spite the established theoretical basis. In TNE, the grammar pattern is populated with semantic and statistical information about collocates. While syntax is analysed according to clause roles, using the SPOCA model (subject, predicate, object, complement, adverbial), semantics require using semantic types. Semantic types are intrinsic attributes of a noun; they represent cognitive concepts such as Human, Institution, Vehicle, Event, etc. They can be seen as hypernyms to define more specific words. Thus, in the previous example of the verb *to arrive*, collocates in subject position of pattern 1 could be *ambulance, guest, train, messenger, visitor, plane* or *convoy*, among others, such that the associated semantic types are [[Human | Vehicle]]. For the sake of coherence, the semantic types are hierarchically organised in a bottom-up shallow ontology of basic concepts built from corpus⁶. The semantic types are complemented by lexical sets and contextual roles. Lexical sets are groups of collocates occupying an argument position, and they are used to complement the semantic type when the latter is too general to characterize the intended meaning. Contextual roles, in turn, are more specific concepts belonging to a semantic type and are assigned by context. For instance, [[Human]] is a semantic type, and 'Professional', 'Footballer', or 'Judge' can be roles. In our example, the semantic type [[Concept]] of pattern 2 is specified by the role 'Considered opinion'. So, if we focus only on patterns 1 and 2, the verb *arrive* means something similar to 'to come to a place after a journey' if the complement is a location (pattern 1), and 'to adopt an opinion after a process' if the complement is a concept or opinion (pattern 2). Furthermore, only persons and vehicles can be normal subjects of pattern 1, whereas only persons and institutions can be normal subjects of pattern 2. It is not possible to know in advance what *arrive* or any other word means, without taking into consideration its context of occurrence. In this way, let's imagine we make again the general (and common) above-mentioned question: 'What does *to arrive* mean?'. Taking into account a contextual analysis, a possible answer would be, according to Firth (1957: 11): 'It depends on the company [the word] keeps'. The context activates one of the various meanings that only virtually exist in the verb. Thus, a word requires the presence of other words if it is to mean something – 'many, if not most,

⁶ The ontology used in CPA project (see section 3.2) is in progress as all PDEV project (Hanks, in progress). The current version is available at <http://www.pdev.org.uk/#onto> (last access: 4/10/2014). For more information on the ontology see Ježek and Hanks (2010).

meanings require the presence of more than one word for their normal realization' (Sinclair 1999: 133).

Some native speakers may point out that pattern 2 can be classified as a conventional metaphor based on the more literal use in pattern 1. This may well be true, but metaphorical status is irrelevant to the reader's or listener's task of decoding the meaning of an utterance. In this sense, the picture is not complete without the concept of 'exploitation', a mechanism for creating unusual meanings for a particular context when the word does not convey the exact meaning the speaker wishes to express. In the sentence 'The plot had arrived at Beirut', the noun *plot* is being treated as if it was a moving vehicle. With rigorous respect for corpus data, this sentence does not fit with pattern 1, but the notion that a plot is something that moves is not frequent enough to be considered a separate pattern. According to the TNE this is an exploitation, a metaphorical, creative modification of a established pattern (exploitations are explained in detail in Hanks 2013: 211–250).

3.2. Systemising corpus analysis of lexico-syntactic patterns with CPA

CPA (Hanks 2004a, 2010) is the procedure for analysing normal patterns of usage of words in context. It is based on the TNE postulates, and establishes the formula to corpus analysis and pattern extraction. The result of an analysis applying CPA is the one shown in the previous section as an example (the verb *to arrive*).

CPA is inspired mainly by lexicographical needs, but in fact represents an innovative way of doing corpus analysis that could be used for natural language processing (Hanks and Pustejovsky, 2005); for instance, for word sense disambiguation (El Maarouf et al. 2013). It has also been applied to terminology (AUTHOR, 2009; AUTHOR, 2013) or pedagogical lexicography (AUTHOR, 2012). It is still currently a manual system, though it is supported by computational tools. There are already some preliminary attempts to automatise certain parts of the task (AUTHOR ET AL, 2012), but this is still work in progress. Finally, CPA is the basis the *Pattern Dictionary of English Verbs*, PDEV (Hanks, in progress), in which the main analysis of CPA patterns is being developed.

4. Applying CPA to pedagogical full-sentence definitions

In this section, we describe a proposal for adapting CPA patterns to full-sentence definitions of Spanish verbs, in the context of the already mentioned DAELE project. As we stated in previous sections, an appropriate definition for learners must take care of the following aspects:

- a) Correspond with real usage, that is, strictly follow corpus data.
- b) Offer information about semantics of the word not in isolation but connected to other lexical units, apart from collocations.
- c) Show information about how the word can be used in terms of most frequent syntactic structures.
- d) Finally, offer all these components in a clear and comprehensive way, in order to make the information easy to understand for a nonexpert user.

We first make a brief presentation of the pilot project of DAELE. Secondly, we illustrate the application of CPA for the building-up of full-sentence definitions.

4.1. A pilot online dictionary of Spanish for foreign learners

As already mentioned in Section 3.2, DAELE is a project for building a new dictionary for Spanish learners which, being monolingual, is conceived for intermediate or advanced

levels. The project is currently in its first stages of development, and the first grammatical category being treated is the verb, as it is a fundamental part of the sentence and one of the most difficult categories of the dictionary in terms of grammar complexity. DAELE is based on the work developed fundamentally by British pedagogical lexicography, and it is trying to apply the Sinclairian conception of dictionaries, above all in his major dictionary project, *Cobuild*. We adopt the conception of corpus as the origin not only of examples, but of the whole analysis of entries; full sentence definitions are also used, according to the principles set out in Hanks (1987: 116–136). Every definition is supported by examples of real usage. A description of various aspects of the dictionary are issued in Battaner (2010), AUTHOR ET AL (2011), AUTHOR (2012: 244–245) and Arias-Badia et al. (2014), among others. In DAELE's website (<http://www.daele.eu/>, last access: 4/10/2014) there are currently around 350 high frequency Spanish verbs. From these verbs, a core of 60 verbs were analysed with CPA and adapted to the dictionary following the methodology we are describing in this paper. A sample of these verbs can be consulted in a preliminar version of Spanish CPA database (AUTHOR, 2012: 179–242): <http://melot.upf.edu/dsele/scpa> (last access: 4/10/2014). The web format is fundamental to offer information about grammar and collocations, because it allows to provide extended explanations and a large amount of data. These data can be connected through hyperlinks creating a net of semantic and syntactic features. Nevertheless, in web applications it is also necessary to be concise and to devote attention to the user's specific needs (Atkins and Rundell 2008: 20–24). But, at the same time, space limitations, one of the biggest difficulties in all lexicographical traditions, is no longer a problem, and it is now possible to organise the information with labels that can be either exposed or hidden by the user.

Figure 2 shows the DAELE entry of the verb *costar* 'to cost' as an example.

costar (verbo)

Conjugar

1 TENER COSTE

a

- transitivo** Un producto, servicio o proyecto cuesta una cantidad de dinero a una persona cuando tiene como precio dicha cantidad, que la persona ha de pagar para obtenerlo:
 - Entre semana, la entrada al festival costará 3 euros. (IULA50)
 - El curso de 12 horas costará 170 euros. (IULA50)
 - El Gobierno prevé que la reforma costará al Estado 5.000 millones de euros. (IULA50)
 - [absoluto] El gasto en pensiones y sanidad le va a costar cada vez más caro al Estado. (IULA50)

combi (adv.) barato, caro, demasiado, mucho, poco, nada...

nota Se utiliza seguido de diversas expresiones coloquiales que indican que la cosa mencionada cuesta mucho, es muy cara: *costar la torta un pan, un ojo de la cara, un Potosí, un riñón, una fortuna, una pasta, un huevo* (malsonante)...

b

- transitivo** Una acción o proceso cuesta tiempo, dedicación o esfuerzo a alguien cuando requiere usar o gastar dicho tiempo, dedicación o esfuerzo para realizarlo o desarrollarlo:
 - Dice que, en esta vida, lo que no cuesta esfuerzo no tiene ningún valor. (SWC)
 - [subordinada suj., infinitivo] Levantar esta industria me ha costado mucho tiempo. (IULA50)
 - [subordinada suj., infinitivo] Tal vez te cueste trabajo comprender el poema y tengas que releerlo. (IULA50)

nota El OI suele ser un pronombre personal.

c

- transitivo** Una acción o hecho adverso o equivocado cuesta algo de valor no material a alguien cuando conlleva la pérdida de esa cosa valiosa:
 - Su postura le costó la derrota electoral en dos ocasiones. (IULA50)
 - Los atentados han costado la vida a muchas personas. (IULA50)
 - La libertad ha costado mucha sangre. (IULA50)
 - La mala gestión acabará por costar votos. (IULA50)
 - [subordinada, infinitivo] El exceso de velocidad le va a costar perder el carné de conducir por un mes. (IULA50)
 - [subordinada suj., infinitivo] Ir en moto sin casco cuesta tres puntos del carné de conducir. (IULA50)

nota El OI suele ser un pronombre personal.

2 SER DIFÍCIL

- intransitivo** Una acción, actitud o decisión cuesta a alguien cuando le resulta difícil o le supone un esfuerzo:
 - Tengo bastante facilidad para escribir, no me cuesta. (SWC)
 - [subordinada suj., infinitivo] Dijo que le costaba creer la noticia. (IULA50)
 - [subordinada suj., infinitivo] Cada vez le cuesta más subir y bajar escaleras. (IULA50)
 - [subordinada suj., infinitivo] Cuesta comprender por qué fueron derrotados estos tenistas. (IULA50)

nota El OI suele ser un pronombre personal.

costar caro (locución verbal)

- Una acción realizada por una persona le cuesta cara a esta cuando conlleva consecuencias negativas para ella, como un problema, un castigo o una venganza:
 - Los errores nos pueden costar caros. (SWC)
 - Aquella actitud podía costarle cara. (SWC)

costar Dios y ayuda (locución verbal)

- Una acción cuesta Dios y ayuda a alguien cuando le resulta muy difícil de realizar o le requiere un gran esfuerzo:
 - [subordinada suj., infinitivo] A mi novia le ha costado Dios y ayuda encontrar trabajo. (SWC)
 - [subordinada suj., infinitivo] El pueblo se encuentra situado sobre un alto, y me ha costado Dios y ayuda llegar. (SWC)

cueste lo que cueste (locución adverbial)

- Alguien dice que va realizar una acción cueste lo que cueste cuando debe conseguirse realizarla poniendo toda la voluntad en ella y empleando el tiempo, esfuerzo o dinero necesarios:
 - Hay que ganar cueste lo que cueste. (IULA50)
 - Estaba dispuesto a llegar hasta el fondo de este escándalo, costara lo que costara. (web)

nota Se utiliza solamente con los tiempos verbales *cueste lo que cueste* y *costara lo que costara* (o *costase lo que costase*).

FAMILIA *coste* (sust.) **FAMILIA** *costear* (verb.) **FAMILIA** *costo* (sust.) **FAMILIA** *costoso* (adj.)

Figure 2. The entry *costar* ‘to cost’ in DAELE, expanded version.

This entry, as a result of intensive corpus research, has two meanings, one labelled as ‘tener como coste’ (‘to have as cost’) and the other as ‘ser difícil’ (‘to be difficult’). In the first case, this wide meaning is divided into two more specific uses, *a* and *b*. The second meaning is constituted by only one use. The difference between the two uses of meaning 1 is that *a* is devoted to products or other things that have a price, and *b* describes actions or processes that must take place or that happen by spending time or effort. There are also notes (headed by the word ‘nota’), specific notes for examples (in square brackets), collocates (headed by the label ‘combi’), and information about word family (at the bottom).

4.2. Proposal for the application of CPA to full-sentence definitions of a learners dictionary

Regarding the application of CPA to definitions of a Spanish learners’ dictionary, in section 1 we already explained two key aspects to take into account: criteria of adaptation and pedagogical goal. In this section, we will show the procedure and illustrate how we proceed with some verbs.

The methodology involves the following three basic steps:

- Firstly, we manually analyse a random sample of sentences from the corpus.
- Secondly, we create the patterns in an online database associated with the corpus.
- Finally, we create the definition adapting the patterns to the characteristics of a learners' dictionary.

Figure 3 shows a schema of the whole process. As step a) is the pure corpus analysis, in the next section it is left behind in order to focus on the connection of steps b) and c).

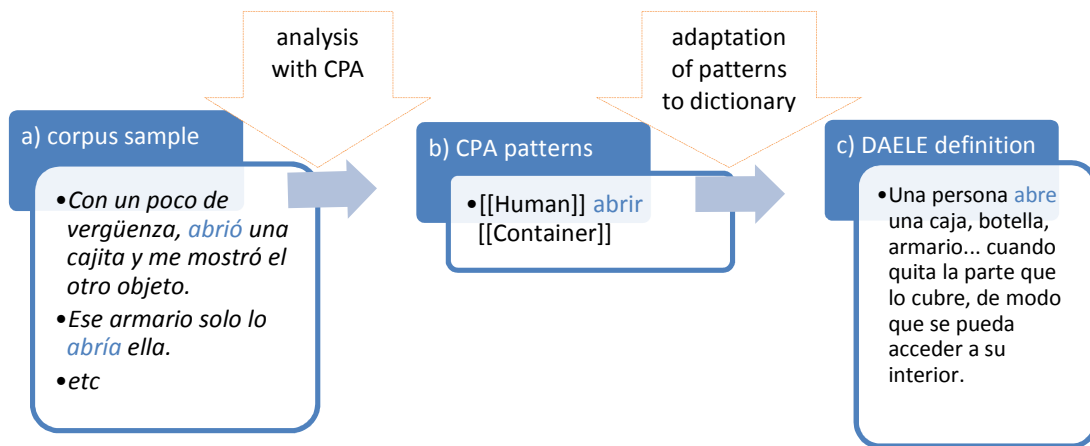


Figure 3. The process of corpus analysis, pattern extraction and adaptation to the dictionary.

4.2.1. *CPA-DAELE connection.* The process to convert a CPA pattern (step b) into a full-sentence definition (step c) has different implications: it means to convert a highly encoded information made for being understood only by specialists (being linguists, language teachers or computer scientists) into a pedagogical explanation for non-native students. To sum up, it means to make the process we synthesise in Figure 4⁷.

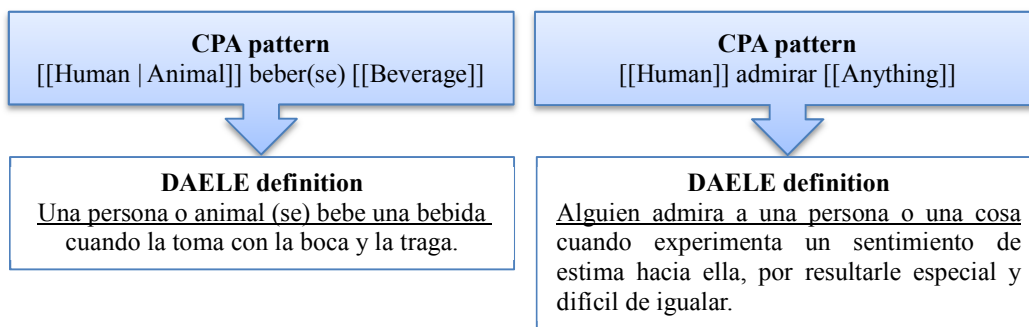


Figure 4. Two examples of CPA patterns of the verbs *beber* ‘to drink’ and *admirar* ‘to admire’, and their corresponding definitions for DAELE. We underline the part of the definition corresponding to the pattern of usage, and the rest corresponds to the explanation of the pattern.

⁷ For the whole analysis and the lexicographical proposal of the verbs shown as examples in this section, see AUTHOR (2012). The samples can be found at <http://melot.upf.edu/dsele> and <http://melot.upf.edu/dsele/scpa> (last access: 29/09/2014).

In this figure, first part of the definition (underlined) corresponds to the CPA pattern, which contains the semantic and syntactic information about the context in which these specific meanings of the verbs are used. The part of the definition after the *definiendum* is the explanation of the meaning, which corresponds to the conventional meaning or implicature. The following strategies can be assumed to adapt syntactico-semantic patterns such as CPA ones into full-sentence definitions for a learners' dictionary such as DAELE.

a) *Convert CPA semantic types into basic vocabulary.* The first criteria of adaptation is to change semantic labels used on CPA ontology into basic vocabulary. As explained in section 3.1, in terms of semantics, CPA patterns are created mainly with semantic types (concepts) inter-connected in a shallow ontology. They are used to characterise the semantics of verb arguments. For example, semantic types are [[Human]], [[Artifact]], that is, all the things created by human beings; [[Process]], all things which happen spontaneously or without human intervention; [[Emotion]], all feelings, etc. In order to adapt these labels, the most obvious step is to keep the same noun, when it is clear enough for the non expert user. For example, [[Process]] or [[Illness]] are directly adapted to *proceso* 'process' and *enfermedad* 'illness'. Nevertheless, in many cases some partial change is needed. In the case of [[Human]], for instance, it is converted to *persona* 'person' or *alguien* 'somebody', because these are the most common options in dictionaries to refer to humans, and are familiar to users. Another example is the more general semantic type [[Physical Object]], which in CPA ontology refers to 'anything with physical nature', such as a cup, a chair but also a building or a planet. In our case, [[Physical Object]] is normally adapted to *cosa* 'thing' or *objeto* 'object', because it is the most common, natural word to refer to these objects, without further specification. Also, in any natural language, such as English or Spanish, *objects* are prototypically *physical*. Figure 5 shows an example of strategy a).

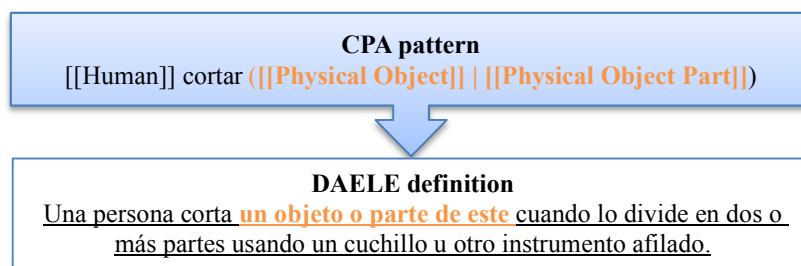


Figure 5. A pattern of the verb *cortar* 'to cut' and the corresponding definition in DAELE.

b) *Selection of a lexical set to delimit a semantic type.* In some cases, semantic types mentioned in a) can be less informative for a user due to its generality. For example, it is more informative to define *open* with *Somebody opens a box, bottle, can... when...* rather than with *Somebody opens a container when...* Two reasons may explain this fact: firstly, that vocabulary units such as *box*, *bottle* or *can* are more frequent than *container*, and secondly, related to the previous one, that they are more illustrative and informative. This happens more often in arguments in direct rather than in subject position. Another more dramatic example for the same Spanish verb *abrir* 'to open' is its use with the meaning of

‘to make an injury’, but in this case it is restricted mainly to *head* and *brow*. It is not possible to create a clear definition with a strategy such as *Alguien abre una parte dura y redondeada del cuerpo a otra persona cuando...* (‘Somebody causes an injury in a hard, rounded body part to another person when...’). It is more clear and simple to say *Alguien le abre la cabeza o la ceja a otra persona cuando...* (‘Somebody causes an injury in other person’s head or brow when...’). This strategy allows to include other aspects of usage meaning, such as the expletive (redundant) pronoun *le* (*Alguien le abre la cabeza...*). In sum, lexical sets are group of words which populates the semantic valency structure of the verb and are united by a semantic type. They usually do not include all the lexical items that could potentially be included on it, that is, there are open sets. In definitions, this can be solved with ellipsis (...) or the abbreviation *etc.* We use frequency criteria to decide which lexical units to select from the set. To obtain frequency and salient data, we use Word Sketch tool in Sketch Engine (Kilgarriff et al. 2004). Figure 6 shows different examples of this option.

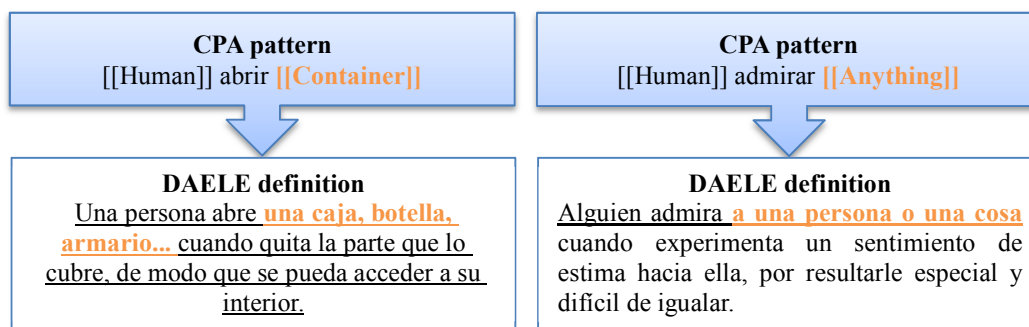


Figure 6. Patterns of the verb *abrir* ‘to open’ and *admirar* ‘to admire’ and its corresponding definitions in DAELE.

Strategy b) is also used in dictionaries such as DAFLES (see the example of *jouer* in Section 1), and pretends to reduce the verbosity and complexity of the full-sentence definitions observed by Rundell (2006). In Figure 6, two examples of process b) can be found. In *abrir* ‘to open’, the direct object [[Container]] is converted into an open lexical set (*caja, botella, armario...* ‘box, bottle, closet...’), and the label [[Anything]] is converted into *persona o cosa* ‘somebody or something’, the most representative components of the set.

c) *Combining semantic types with lexical sets.* A combination of a) and b) is also used in many cases, when it is considered that a general semantic label is explicative enough by itself, but some example of a lexical item may be of help. For instance, in the case of *casar* ‘to marriage’, the lexical set *cura, juez*, etc (‘priest, judge’) can be restricted with *...or another competent authority*, an adaptation of the semantic type [[Human = Civil or Religious Authority]] (Figure 7).

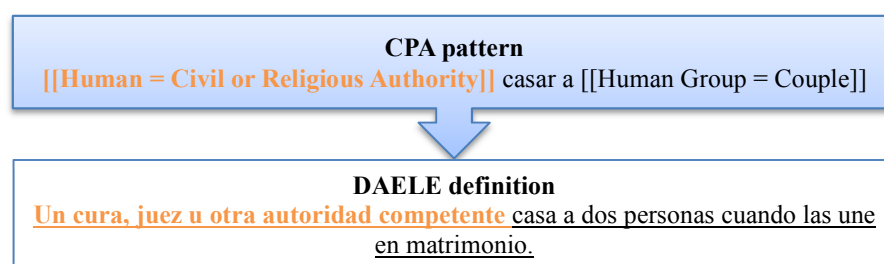


Figure 7. A pattern of the verb *casar(se)* ‘to marriage’ and its corresponding definition for DAELE.

In the example, *autoridad competente* (‘competent authority’) would be the equivalent of the semantic type [[Human = Civil or Religious Authority]], complemented by two lexical items, *cura* and *juez* (‘priest’ and ‘judge’).

d) *Making the syntactic structure explicit.* Finally, there are some patterns that are only used in a specific syntactic structure. In this case, this structure is made explicit in the definition, instead of ‘hiding’ it in a more general pattern. For example, it is relatively common that some senses are activated only with clauses in direct object position, and this is reflected in the definition. In Spanish CPA, clauses are represented by the semantic type [[Eventuality]], which alludes to actions or processes. See Figure 8 for an example with the verb *imaginar(se)* ‘to imagine’.

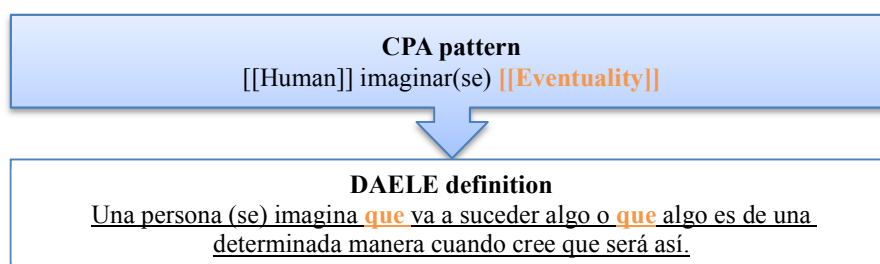


Figure 8. A pattern of the verb *imaginar(se)* ‘to imagine and its corresponding definition for DAELE.

4.2.2. *‘Detaching’ meanings from the corpus to the dictionary: an example with the verb desprender/se.* This section is devoted to explain in detail the process shown in the previous section with the verb *desprender(se)* ‘to detach, to give off’ as an example.

English CPA as well as Spanish CPA use a random sample of 250 concordances as a minimum, in the case of a verb such as *desprender(se)*. Highly frequent and polysemous verbs need, however, larger samples. For creating the sample, and also for labelling each concordance with its respective number of pattern, all versions of CPA for each language are using a modified version of the Sketch Engine. In the case of Spanish, a journalistic corpus of 50 million words is used (IULA50, see note 2).

Table I shows the CPA patterns (and implicatures) derived from the analysis of the IULA50 random sample, and the definitions of DAELE created from these patterns.

For *desprender/se*, the following patterns were noticed in the IULA50 corpus:

CPA pattern	DAELE definition	Corpus example
Implicature		
1 (a) [[Human Event]] desprender [[Physical Object 1 Physical Object Part]] {(de [[Physical Object 2]])} [[Human Event]] <i>detaches</i> [[Physical Object 1 Physical Object Part]] from [[Physical Object 2]]	1 (i) Algo o alguien desprende una cosa de otra a la que estaba unida, pegada o en la que estaba sostenida cuando la separa de modo que deje de estar en contacto con ella.	El viento ha desprendido láminas metálicas de los tejados.
1 (b) [[Physical Object 1 Physical Object Part]] desprenderse {(de [[Physical Object 2]])} [[Physical Object 1 Physical Object Part]] <i>becomes detached</i> from [[Physical Object 2]]	1 (ii) Algo se desprende de un sitio cuando cae y queda suelto o separado de él.	Tuvo que apartar unas cuantas piedras que se habían desprendido de la montaña.
2 [[Human Institution]] desprenderse {de [[Entity = Possession]]} [[Human Institution]] <i>parts with or disposes of</i> [[Entity = Possession]]	2 Alguien se desprende de algo valioso cuando lo entrega voluntariamente, muchas veces en un gesto de solidaridad.	Para apoyar la causa, el pintor se desprendió de una de sus obras.
3 [[Physical Object]] desprender [[Stuff]] [[Physical Object]] <i>gives off small particles of</i> [[Stuff]]	3 (a) Un objeto desprende un olor, gas, sustancia o radiación cuando lo emite o produce, haciendo que salga de él al exterior.	Al reaccionar con el agua, el compuesto desprende iones.
4 (a) [[Anything]] desprender [[Emotion]] [[Anything]] <i>causes an</i> [[Emotion]]	3 (b i) Algo o alguien desprende una sensación o sentimiento cuando lo muestra y lo hace perceptible.	Vino a mi encuentro con actitud severa, que desprendía una clara preocupación.
4 (b) [[Emotion]] desprenderse {de [[Anything]]} [[Emotion]] <i>is caused by</i> [[Anything]]	3 (b ii) Una sensación o sentimiento se desprende de algo o alguien cuando es mostrado o exteriorizado por esta persona o cosa, de modo que sea perceptible.	De sus palabras se desprendió la sensación de que el acuerdo financiero es posible.
4 [[Idea = Conclusion]] desprenderse [[de Information]] [[Idea = Conclusion]] <i>can be deduced from</i> [[Information]]	4 Una conclusión se desprende de una información o datos cuando se deduce o es consecuencia de estos.	Del análisis comparativo se desprende que la enfermedad aparece muy pronto en ese sector.

Table I. Correspondence between CPA patterns and the meanings shown in DAELE.

Prototypically *desprender* means (shown in pattern 1 a) that an agent like a human or an event (for example, the wind) detaches a thing or part of a thing from another object. This prototypical meaning is infrequent in the corpus, and, on the contrary, pattern 1 (b) is fairly

frequent: it is used to express that a thing or a part of it detaches from another object without the intervention of any agent (inchoative structure). The rest of the patterns are figurative meanings derived from the first two: pattern 2 is used when a person or institution ceases to possess something else, generally by donating it. Pattern 3 expresses that a physical object gives off smaller parts of itself away. Pattern 4 is, as pattern 1, another case of causative-inchoative alternation, and both denote the situation in which something or somebody causes certain emotion in people. Finally, pattern 5 is used for ideas that are derived from some piece of information.

With respect to the definitions, all the effort must be made in order to make them easy and quickly understandable. As explained in the previous section, it seems impossible to create the implicature exactly as it is created in CPA: some semantic types, such as [[Event]] (*evento* in Spanish), refer to very broad concepts which may not be sufficient for clarifying the use of the verb to a learner of the language.

For this reason, in some cases, a lexical set is used in the definition instead of the corresponding semantic type (strategy b in Section 4.2.1). The words in our lists are chosen as being typical members of the relevant lexical set. For example, in pattern 3 the semantic type [[Stuff]] ‘*materia*’ is used to describe the complement, but this is underrestrictive for a learner. In the definition, therefore, we help him or her with the list *olor, gas, sustancia o radiación* (‘smell, gas, substance, or radiation’): at the beginning we indicate the most frequent options (smell and gas), and at the end the less frequent ones. Furthermore, the opposite situation must also be avoided: there are semantic types that are so specific that restrict the meaning of the word too much. In this case, more general nouns such as *algo* (‘something’) or *cosa* (‘thing’) are used in the definition.

In sum, when a CPA pattern is used as a basis for an entry in the DAELE, a balance between generalization and specification is required in order to clarify the meaning to the user and adapt the process to his or her needs.

DAELE entry built from the patterns we illustrated in Table I is shown in Figure 9.

desprender/se (verbo)

Conjugar

1 SEPARAR/SE

- transitivo** Algo o alguien desprende una cosa **de** otra a la que estaba unida, pegada o en la que estaba sostenida cuando la separa de modo que deje de estar en contacto con ella:
 - El viento ha desprendido láminas metálicas **de** los tejados. (IULA50)
 - La joven desprendió un brillante **de** una de sus gargantillas y lo arrojó al canal. (DAV)
- desprenderse (de)** pronominal (*cambio de OD a sujeto*) Algo **se** desprende **de** un sitio cuando cae y queda suelto o separado de él:
 - Tuvo que apartar unas cuantas piedras que **se** habían desprendido **de** la montaña. (IULA50)
 - La plataforma **se** desprendió y causó heridas a tres obreros. (IULA50)

2 ENTREGAR VOLUNTARIAMENTE

- desprenderse (de)** pronominal Alguien **se** desprende **de** algo valioso cuando lo entrega voluntariamente, muchas veces en un gesto de solidaridad:
 - Se** ha desprendido **del** célebre lienzo para que cuelgue en una galería neoyorquina. (IULA50)
 - Para apoyar la causa, el pintor **se** desprendió **de** una de sus obras. (IULA50)

3 EMITIR

a

- transitivo** Un objeto desprende un olor, gas, sustancia o radiación cuando lo emite o produce, haciendo que salga de él al exterior:
 - Esta planta desprende un olor estimulante y aromático. (SWC)
 - Cuando un cometa se acerca al Sol, desprende gas y polvo formando la espectacular cola. (IULA50)

b

- transitivo** Algo o alguien desprende una sensación o sentimiento cuando lo muestra y lo hace perceptible:
 - Tus ojos desprenden vitalidad. (SWC)
 - Vino a mi encuentro con actitud severa, que desprendía una clara preocupación. (IULA50)
- desprenderse pronominal (cambio de OD a sujeto)** Una sensación o sentimiento **se** desprende **de** algo o alguien cuando es mostrado o exteriorizado por esta persona o cosa, de modo que sea perceptible:
 - De** sus palabras **se** desprendió la sensación de que el acuerdo financiero es posible. (IULA50)
 - De** l césped húmedo, **de** los árboles **se** desprendía un ambiente de tristeza indefinible. (DAV)

4 MOSTRAR/PERCIBIRSE

- desprenderse (de)** pronominal Una conclusión **se** desprende **de** una información o datos cuando se deduce o es consecuencia de estos:
 - De** lo anterior **se** desprenden algunas conclusiones. (SWC)
 - ¿Qué es lo que puede desprenderse **de** todo esto? (IULA50)

Figure 9. Entry of the verb *desprender(se)* in DAELE.

5. Conclusions and future work

In this paper, we have shown a proposal for systematising definitions for learners's dictionaries. Using CPA for building lexicographical entries is highly time-consuming. In addition, the database and other tools still need to be improved to become more efficient not only in terms of the time invested, but also in the quality of the resulting data. However, CPA provides a fine-graded analysis of language in use, and it can be considered a systematisation and extension of Sinclair's ideas. As pointed out in Section 1, it is a reasonable assumption that dictionaries must be built from a corpus, but corpus analysis must be supported by a system which guarantees coherence of the work of one lexicographer but also (and this seems even more important) the work of every component of a lexicographical team.

Spanish CPA as well as DAELE are currently ongoing projects, and many tasks are still left for future work. Apart from some obvious steps, such as increasing the number of verbs and types of verbs to be analysed or testing the same methodology for other languages, the automatization of the process is our main concern for the near future. CPA is very time-consuming, so, if it is confirmed as a proper methodology for dictionary making, this must be followed by proposals for making the work easier, faster and more precise. In this sense, our work is following two ways: a) The semiautomatic creation of verb patterns, that is, that the process of analysing the corpus and creating the patterns is executed partially by automatic procedures (see AUTHOR 2012, for a part of the process of automatization); b) The automatization of some parts of the creation of definitions: having the patterns as the ones showed in Figure 4, it would be relatively easy to implement templates to automatically generate natural definitions from the patterns by translating CPA's notation, facilitating the definition writing process.

REFERENCES

- Adamska-Salaciak, Arleta. 2012. "Dictionary definitions: problems and solutions". *Studia Linguistica Universitatis Iagellonicae Cracoviensis 129 supplementum*. 323–339.
- AUTHOR. 2009.
- AUTHOR. 2013.
- AUTHOR. 2012.
- AUTHOR ET AL. 2011.
- Arias-Badia, Blanca, Elisenda Bernal and Araceli Alonso. 2014. "An Online Spanish learners' dictionary: The DAELE project". *Slovenščina 2.0* 2, 53-71.
- Atkins, S. and Michael Rundell. 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Battaner, Paz. 2010. "El uso de las etiquetas semánticas en los artículos lexicográficos de verbos en el DAELE". *Quaderns de Filologia. Estudis Lingüístics* 15, 139–158.
- Battaner, Paz. (dir.). In progress. *Diccionario de aprendizaje del español como lengua extranjera* (verbs) [DAELE]. URL: <http://www.daele.eu> (last access: 4/10/2014).
- Bogaards, Paul. 2010. "Dictionaries and Second Language Acquisition". In: Anne Dykstra and Tanneke Schoonheim (eds.): *Proceedings of the XIV Euralex International Congress*, Ljouwert (Netherlands): Fryske Akademy, 99–123.
- Bosque, Ignacio. 2006. "Una nota sobre la relevancia de la información sintáctica en el diccionario". In: Elisenda Bernal and Janet DeCesaris (eds.): *Palabra por palabra. Estudios ofrecidos a Paz Battaner*, Barcelona: IULA-UPF, 47–53.

- De Schryver, Gilles Maurice. 2003. "Lexicographers' Dreams in the Electronic Age". *International Journal of Lexicography* 16 (2), 143–199.
- El Maarouf, Ismail, Víték Baisa, Jane Bradbury and Patrick Hanks. 2014. "Disambiguating Verbs by Collocation: Corpus Lexicography meets Natural Language Processing". In: Nicoletta Calzolari et al. (eds.): *Proceedings of LREC'14*, European Language Resources Association, 1001–1006.
- Firth, John Rupert. 1957. *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.
- Gutiérrez Cuadrado, Juan (dir.). 1996. *Diccionario Salamanca de la lengua española*. Madrid: Santillana.
- Halliday, Michael. 1976. *System and Function in Language* (ed. by G. Kress). Oxford: Oxford University Press.
- Hanks, Patrick. 1987. "Definitions and Explanations". In: John Sinclair (ed.): *Looking Up: Account of the Cobuild Project in Lexical Computing (Collins Cobuild dictionaries)*, London: Collins, 116–136.
- 2004a. "Corpus Pattern Analysis". In: Geoffrey Williams and Sandra Vessier (eds.): *Proceedings of the Eleventh EURALEX International Congress, Euralex 2004*, Lorient: Université de Bretagne-Sud, 87–97.
- 2004b. "The Syntagmatics of Metaphor and Idiom". *International Journal of Lexicography* 17 (3), 245–274.
- 2010. "How People Use Words to Make Meanings". In: B. Sharp and M. Zock (eds.): *Proceedings of the 7th International Workshop on Natural Language Processing and Cognitive Science, NLPCS 2010*. In conjunction with ICEIS 2010, Funchal, Madeira, Portugal.
- 2013. *Lexical Analysis: Norms and Exploitations*. Cambridge, MA: The MIT Press.
- Hanks, Patrick (ed.). In progress. *Pattern Dictionary of English Verbs [PDEV]*. URL: <http://www.pdev.org.uk> (last access: 4/10/2014).
- Hanks, Patrick and James Pustejovsky. 2005. "A Pattern Dictionary for Natural Language Processing". *Revue Française de Linguistique Appliquée* 10 (2), 63–82.
- Harvey, Keith and Deborah Yuill, 1997: "A study of the use of a monolingual pedagogical dictionary by learners of English engaged in writing". *Applied Linguistics* 18 (3), 253–278.
- Hornby, Albert Sidney. 1954. *A guide to Patterns and Usage in English*. Oxford: Oxford University Press.
- Hornby, Albert Sidney and Anthony Cowie (eds.). 1963: *Oxford Advanced Learners Dictionary of Current English*, 2nd ed. Oxford: Oxford University Press. [OALDCE]
- Hunston, Susan and Gilles Francis. 2000. *Pattern Grammar*. Amsterdam: John Benjamins.
- Ježek, Elisabetta and Patrick Hanks. 2010. "What lexical sets tell us about conceptual categories". *Lexis: E-journal in English Lexicology*, 4: *Corpus Linguistics and the Lexicon*, 7–22.
- Johnson, Samuel. 1755. *Preface to a Dictionary of the English Language*. (Edited by Jack Lynch). URL: <http://andromeda.rutgers.edu/~jlynch/Texts/preface.html> (last access: 10/11/2014).
- Kilgarriff, Adam, Pavel Rychly, Pavel Smrz and David Tugwell. 2004. "The Sketch Engine". In: Geoffrey Williams and Sandra Vessier (eds.): *Proceedings of the Eleventh EURALEX International Congress, Euralex 2004*, Lorient: Université de Bretagne-Sud, 105–116.

- Lew, Robert and Anne Dziemianko. 2006. "A new type of folk-inspired definition in English monolingual learners' dictionaries and its usefulness for conveying syntactic information". *International Journal of Lexicography* 19 (3), 225–242.
- Rundell, Michael. 2006. "More than one way to skin a cat: why full-sentence definitions have not been universally adopted". In: Elisa Corino, Carla Marello Cristina (eds.): *Atti del XII Congresso di Lessicografia, Torino, 6–9 settembre 2006*. Torino: Edizioni dell'Orso, 323–338. [Repr. in Fontenelle T. (ed.) 2008. *Practical lexicography: A reader*, Oxford: Oxford University Press, 197–209.
- Sinclair, John (dir.). 1987. *Collins Cobuild English Language Dictionary*. Glasgow: Harper-Collins. [*Cobuild*]
- Sinclair John (ed.). 1987. *Looking Up: An Account of the Cobuild Project in Lexical Computing*. London: Harper-Collins.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- 1999. "The Lexical Item". In: Ernest Weigand (ed.): *Contrastive Lexical Semantics*. Amsterdam: John Benjamins, 1–24.
- Verlinde, Serge and Thierry Selva. 2006. *Dictionnaire d'apprentissage du français langue étrangère ou seconde*. Katholieke Universiteit Leuven. URL: <http://www.kuleuven.be/ilt/frans/> (last access: 18/11/2014). [DAFLES]