

Corpus Pattern Analysis in determining specialised uses of verbal lexical units¹

ARACELI ALONSO CAMPO
Université de Bretagne-Sud
araceli.alonso@univ-ubs.fr

IRENE RENAU ARAQUE
Universitat Pompeu Fabra
irene.renau@upf.edu

Araceli Alonso Campo és doctora

en Lingüística Aplicada per la Universitat Pompeu Fabra. Ha estat professora associada a la Facultat de Traducció i Interpretació de la Universitat Pompeu Fabra i a la Facultat de Formació del Professorat de la Universitat de Barcelona. Actualment és *Attaché Temporaire d'Enseignement et de Recherche* a la Université de Bretagne-Sud de Lorient on imparteix docència de llengua espanyola i terminologia. Els seus principals interessos de recerca són la lingüística de corpus, la lexicografia i terminologia, la traducció i la didàctica de la llengua.



Irene Renau Araque és doctora

en Lingüística Aplicada per la Universitat Pompeu Fabra. Com a investigadora del grup Infolex d'aquest mateix institut, ha participat en quatre projectes d'investigació competitiu sobre lexicologia i lexicografia, que són les seves principals línies de recerca, a les quals s'ha de sumar la didàctica de llengües. És professora associada de Llengua Castellana al Departament de Traducció i Ciències del Llenguatge de la Universitat Pompeu Fabra.



Abstract

We present a methodology for the detection of specialised uses of verbs that starts from Corpus Pattern Analysis, CPA (Hanks 2004a, 2004b), a procedure for combining syntax and semantics in order to describe lexical meaning in context. Our hypothesis is that this methodology can be applied in terminology to distinguish specialised uses from the general ones.

KEYWORDS: Corpus Pattern Analysis; environment; lexicographical representation; Theory of Norms and Exploitations; specialised uses; verbal lexical units

Resum

'Corpus Pattern Analysis' per a la determinació dels usos especialitzats de les unitats lèxiques

Presentem una metodologia per a la detecció de verbs especialitzats que parteix del Corpus Pattern Analysis o CPA (Hanks 2004a, 2004b), un procediment d'anàlisi de corpus que proposa la combinació de la sintaxi i la semàntica per donar compte del significat del lèxic en context. La nostra hipòtesi és que aquesta metodologia es pot aplicar en terminologia per distingir els usos especialitzats de verbs dels no especialitzats.

PARAULES CLAU: Corpus Pattern Analysis; medi ambient; representació lexicogràfica; teoria de normes i explotacions; usos especialitzats; unitats lèxiques verbals

TERMINÀLIA 7 (2013): 26-33 · DOI: 10.2436/20.2503.01.49
Data de recepció: 01/07/2012. Data d'acceptació: 08/04/2013
ISSN: 2013-6692 (impresa); 2013-6706 (electrònica) · <http://terminalia.iec.cat>

1 Introduction

Studies of the lexicon based on the perspective that the use of lexical units must be fundamentally determined by its context have inspired several theories from different linguistic approaches devoted to the analysis of words in contexts — e.g. Harris 1954, Austin 1955, Firth 1957, Halliday 1961, Sinclair 1991, Hoey 2005. Lexicographical practice has applied several of these approaches and developed new methods by means of using corpus-based techniques in order to represent the real use of lexical units in dictionaries. A notable example of this can be found in the work by Sinclair in the Cobuild Project (Sinclair 1987), which resulted in a real revolution in dictionary making, leading to a model for compiling corpus-based dictionaries in different languages.

The contextual approach has also transformed terminology. New terminological theoretical perspectives — e.g. Cabré 1999, Temmerman 2000, Bourigault and Slodzian 2000, Faber *et al.* 2005 — have stated that terms must also be considered as lexical units and, as such, they must be observed in their context of use in order to determine the specialised or non-specialised usage of these units. By observing the context of terminological units, it has been established that not only nouns carry over a specialised value, as stated in the classical orientation to terminology (Wüster 1979), but other categories such as verbs or adjectives may also take on a terminological sense (Lorente 2001, 2007, 2009, Estopà *et al.* 2002, Alonso and Torner 2010).

In this paper, we take into account these approaches so as to explore the possibilities of determining and analysing the specialised value of verbs by using one of the most recent methodologies of lexical analysis based on corpus and specifically applied to verbs: Corpus Pattern Analysis or CPA. CPA is a procedure developed by Patrick Hanks based on determining syntactic and semantic patterns of usage of verbs. The CPA method is associated to the Theory of Norms and Exploitations or TNE (Hanks 2004a, 2004b, 2013) and it is the foundation of the *Pattern Dictionary of English Verbs*, PDEV (<http://nlp.fi.muni.cz/pdev>). CPA is also being applied to other languages such as Italian, Chinese or Spanish — for a more specific explanation about how CPA can be used for Spanish, see Alonso (2009), Renau and Battaner (2012) and Renau (2012).

In a previous study (Alonso 2009), a first approach to the use of CPA for analysing lexical units used in thematic contexts in Spanish was developed. It was hypothesised that such a methodology facilitates the observation of specialised uses in contrast to general uses of a word. Hence, in this study, we shall concentrate on confirming our hypothesis based on the assumption that CPA is a valid methodology for the study of both general and specialised lexical units.

2 Verbal Lexical Units in Scientific Texts: Between Terminology and Phraseology

The observation of context has led to a discussion as to whether the dichotomy between general and specialised languages is a real one. Contextual approaches to terminology state that there is no such a division, but instead the two entities form a continuum (Cabré 1999, Meyer 2000, Hunston and Sinclair 2003, Myking 2007). Determining the termhood of a word is really a fuzzy task, as termhood is a matter of degree and not all terms are of the same type.

By looking at words in context, it can be observed that the noun is far from being the only category which plays a relevant role in scientific discourse, other categories such as verbs, adjectives or even adverbs may be set as domain-specific lexical units (Lorente 2001). From a classical approach to terminology, which pursued the standardisation of specialised languages, nouns were considered the prototypical lexical unit with referential character and, therefore, the only category to be considered a term. Most lexical analysis of scientific texts concentrated on the study of noun terminology. The contextual approach has also influenced terminology and broadened the scope of terminology to study other categories besides nouns. Considering a linguistic perspective to terminology, and as Cabré stated (1999), any word can be a term or not, depending on the context.

Thus, verbs are one of the most important categories in discourse, as ‘meanings are constructed around the verb, the pivot of the clause’ (Hanks 2010a: 3). Verbs are relational lexical units which link predicates with the actors who play a role in the predication. Verbs are the cornerstone of the sentence. They are cognitive nodes which form the text framework and help to organise discourse, to articulate and structure the text, to establish relations between lexical units, to express the author’s point of view, to interact with the reader, to create meaning.

Considering the idea shown in the previous lines, Lorente (2009) highlights the idea that verbs are not *per se* terminological units, but can acquire specialised value in context when their immediate environment also provides specialised knowledge. Nonetheless, a corpus-driven methodology that demonstrates how this specialised value is detected in context is needed.

Following a Sinclairian approach to the study of lexis and Sinclair’s distinction between the *open-choice principle* and the *idiom principle*, Hanks (2010b) considers that, in reality, some units have just a terminological tendency (*open-choice principle*) — e.g. *strobilation* — whilst others have a phraseological tendency (*idiom principle*) — e.g. *to blow*. Between these two extremes, there are words which can have both a terminological and a phraseological tendency depending on the

context — e.g. *to organise* or *to admit*. Verbs are mainly affected by phraseology and it is not possible to understand the meaning or meanings of the verb without taking into account the phraseological context in which the verb is used.

In this sense, if we consider the terminological tendency and phraseological tendency of words, a procedure to understand both tendencies and how the two influence each other is needed. We postulate that CPA provides us with such a mechanism.

3 Corpus Pattern Analysis (CPA) and the Theory of Norms and Exploitations (TNE)

Corpus Pattern Analysis is ‘a new technique for mapping meaning onto use’ (Hanks 2002). As noted in the introduction, CPA allows discovering how meanings arise from patterns of usage (words in context), rather than treating words as isolated elements in a compositional structure. It has its theoretical basis in TNE (Hanks 2013). TNE is a theoretical proposal based on the postulates of trends focused on words in contexts that we quoted in the introduction, and more precisely, on Sinclair’s work.

The set of patterns accounting for all normal uses of a given word is what the author refers to as *norms*. Uses that do not fit a norm exactly are called *exploitations*. Exploitations are mainly cases of metaphor, metonymy, and other creative uses of words, and they can also be related to syntax or combinatory.

By analysing concordances extracted from a corpus, CPA allows to identify lexico-syntagmatic patterns of a given word and measure their frequency. This facilitates the association of each pattern of usage to a meaning potential. Each pattern is associated to a prototypical use (normal and conventional) of a word or an exploitation of the normal use, which is related to an implicature or paraphrase of the pattern.

In the case of verb patterns, the analytical procedure consists in searching a verb in a corpus and selecting a random sample of concordances. At the moment, the Sketch Engine tool (Kilgarriff *et al.* 2004, <http://www.sketchengine.co.uk>) adapted to CPA is being used. By looking at the syntactic and argument structure of the verb, concordances are sorted into groups. Each concordance is numbered manually and associated to a pattern. Once the concordances have been related to a pattern, patterns are defined in an online database that is also linked to the Sketch Engine and the given corpus data.

Verb patterns are based on the structure of clause roles as described in systemic grammar (Halliday 1961): subject, predicator, object, complement and adverbial (SPOCA). Each clause role or argument is populated by a set of collocates which usually share a semantic aspect of meaning. Semantic values are given

for each set of collocates. It is what Hanks refers to as ‘semantic types’. The semantic type is an intrinsic property of the collocation and is represented in double brackets. All semantic types are stored in a hierarchically shallow semantic ontology. The semantic types are complemented by lexical sets and contextual roles. In some cases, one or more lexical items that cannot be grouped together into a semantic type populate the argument slot; these are considered as ‘lexical sets’. In other cases, the semantic type is complemented by a ‘semantic role’ which is an extrinsic property assigned by context. For instance, if we look up pattern 4 of the verb *sacrifice* at the PDEV, that is, `[[Human 1 = Leader]] sacrifice [[Human 2 = Politician]]`, `[[Human]]` is a semantic type, and “Leader” or “Politician” are semantic roles. The selection of semantic types, semantic roles and lexical sets is not easy and it is only by corpus evidence that this task can be achieved. Examples of English verbs patterns are freely available at the PDEV (<http://nlp.fi.muni.cz/pdev>).

4 Analysing Spanish Verbs: Between the General and the Scientific

In order to fulfil the main objectives mentioned earlier, some Spanish verbs frequent both in general language and in Environmental texts are analysed by means of CPA. Before going into the analysis, nevertheless, some methodological issues must be addressed.

4.1 Methodological Issues

For our purposes, we concentrate on two Spanish sub-corpus extracted from the *Corpus Tècnic de l’IULA – CTI-IULA* (Bach *et al.* 1997, <http://bwananet.iula.upf.edu/indexen.htm>): the IULA50, made of general texts (50 million words), and AquaCorp (Alonso 1999), formed by texts from the CTI-IULA sub corpus of Environment, and specifically related to water issues (less than 1 million words). We have focused on the Environmental domain as our research has its origins in previous work developed to characterise Environmental terminology (Alonso 2008 and 2009, Alonso and Torner 2010).

Despite the small size of the specialised corpus, we have obtained enough evidence for a first approach to the observation of specialised uses of verbs in contrast to general uses. In some specific cases in which the number of occurrences of a verb was not very high, the *Spanish Web Corpus* (SWC) available at the Sketch Engine tool and the *Corpus de Referencia del Español Actual*, CREA (<http://corpus.rae.es/creanet.html>) were also consulted to corroborate the results.

As for the corpus exploitation tools, Hanks’ collaborators at Masaryk University in Brno have implemented the corpus management system and CPA database for

Spanish. At the moment, the same software as for the compiling of the English pattern dictionary is being used. Finally, for the selection of verbal lexical units to be analysed, to narrow down our analysis, a conceptual class was arbitrarily chosen, specifically, that of ‘cleaning’, and in this class, only the four most frequent verbs were considered: *clorar* (‘to add chlorine’), *depurar* (‘to depurate’), *filtrar* (‘to filter’) and *tratar* (‘to treat’).

4.2 Analysis and Results

4.2.1 *Clorar* (‘to add chlorine’)

Clorar is a denominal verb derived from the noun *cloro* (‘chlorine’). It is clearly a terminological unit that denotes a concept related to Chemistry. It is surmised that the terminological value is carried over from the noun to the verb.

Only 2 occurrences of this verb pertain to the IULA50, and 20 are specific of the Environmental corpus. Due to the small number of occurrences in IULA50, the SWC and the CREA were consulted. In the general corpora, by applying CPA system, one pattern to explain the general use of the verb is established:

- 1 Pattern:** [[Human]] *clorar* [[Liquid = Water]]
Implicature: [[Human]] adds chlorine into [[Liquid = Water]]
(example: ‘Turistas empeñados en **clorar** el agua del depósito del pueblo en el que veranean’, IULA50)

Stereotypically, liquid into which chlorine is added is *water*. On the other hand, in AquaCorp the participle form of the verb is used modifying a noun that is a substance. The sense of ‘adding chlorine’ is maintained, but the semantic type is broadened to any [[Stuff]] which might not be necessarily *agua* (‘water’):

- 1 Pattern:** [[Human]] *clorar* [[Stuff]]
Implicature: [[Human]] adds chlorine into [[Stuff]]
(example: ‘...sustancias [‘substances’] **cloradas** de la industria de papel y pasta papelera’, AquaCorp)

We can see, then, that the terminological use of *clorar* has a wider meaning than the general one. Furthermore, if we take up Hanks’ view of the terminological and phraseological tendency of words, the verb *clorar* would be the case of a verb with a terminological tendency. The meaning in origin of the noun *cloro* is transferred to the verb in the derivation process. The specialised use of the verb is maintained in general language, though with a slight limitation on the object. The scope of the object lexical set is wider in the specialised corpus than in the general one, but the pattern usage is not really different. Clearly, this is a case of movement from the specialised to the general without change in use.

4.2.2 *Depurar* (‘to depurate’)

Depurar has the status of term in origin, also related to Chemistry, as the previous verb analysed. In the general corpus, *depurar* occurs 249 times. Once all the occurrences have been annotated according to the CPA procedure, the following patterns are determined:

- 1 Pattern:** [[Human]] *depurar* [[Liquid]]
Implicature: [[Human]] purifies a [[Liquid]], treats it in order to clean it
- 2 Pattern:** [[Human = Authority]] *depurar* [[Action]]
Implicature: [[Human = Authority]] investigates who is the person responsible for an [[Action]] and makes the decisions related to the facts
- 3 Pattern:** [[Human = Artist | Sportsman/woman]] *depurar* [[Activity = Technique]]
Implicature: [[Human = Artist | Sportsman/woman]] polishes or refines his/her [[Activity = Technique]]
- 4 Pattern:** [[Eventuality 1]] *depurar* [[Eventuality 2]]
Implicature: [[Eventuality 1]] leaves an [[Eventuality 2]] clean from all non-intrinsic elements in order to be shown or observed more clearly
- 5 Pattern:** [[Human 1 = Authority]] *depurar* [[Human 2 = Worker]]
Implicature: [[Human 1 = Authority]] purges a [[Human 2 = Worker]], gets out of his/her job or position by political or ideological reasons

Pattern 1 corresponds to the specialised use. It denotes the process of cleaning or purifying a liquid. The rest of patterns are figurative uses with different meanings in which the core notion of ‘cleaning’ is maintained.

In contrast to these results, in AquaCorp 24 occurrences are found, such as ‘... características de las aguas que hay que **depurar**’. The use activated in the Environmental corpus corresponds to the first pattern in the general corpus. Stereotypically, the subject is also [[Human]] and the object is also [[Liquid]], precisely, *agua* (‘water’) or other similar collocates such as *efluente* (‘effluent’) or *vertido* (‘discharge’), whilst in the general corpus other lexical items such as *sangre* (‘blood’) can populate the argument slot. The pattern in AquaCorp is as follows:

- 1 Pattern:** [[Human]] *depurar* [[Liquid= Water]]
Implicature: [[Human]] purifies a [[Liquid]], treats it by means of a process or chemical product in order to be reused

If we compare this pattern with pattern 1 of the previous analysis of IULA50, it can be observed that the implicature in the Environmental corpus is more precise and semantically restricted.

Depurar is another ‘near-term verb’, but it is different to *clorar*, in the sense that patterns show changes in use. The verb shows a terminological tendency and a phraseological tendency. *Clorar* maintains a termino-

logical tendency both in general and specialised contexts. On the contrary, *depurar* shows a terminological tendency in Environment while in general language the terminological tendency is complemented by the phraseological tendency. It may be argued that *clorar* is more terminological than *depurar* as there is only one prototypical pattern which is used both in general and specialised texts in the same way, while *depurar* shows a prototypical pattern used both in general and specialised texts, with a restriction in the implicature in the specialised text, as well as activation of different senses in general language, corresponding to exploitations of the first pattern.

4.2.3 Filtrar ('to filter')

The verb *filtrar* is not a term itself, contrary to *clorar* or *depurar*. In the general corpus *filtrar* has 641 occurrences. By analysing the random sample of occurrences, the following six patterns are determined:

- 1 **Pattern:** [[Human]] filtrar [[Information Source]]
 Implicature: [[Human]] reveals a confidence or secret
 [[Information Source]] to make it public
- 2 **Pattern:** [[Human]] filtrar [[Fluid | Wavelength]]
 Implicature: [[Human]] makes a [[Fluid | Wavelength]]
 go through a filter or similar object to clean
 or separate it from other stuffs
- 3 **Pattern:** [[Solid]] filtrar [[Fluid | Wavelength]]
 Implicature: [[Solid]] allows a [[Fluid | Wavelength]]
 go through it
- 4 **Pattern:** [[Fluid | Wavelength]] filtrarse [NO OBJ]
 ([Adv[Location]])
 Implicature: [[Fluid | Wavelength]] goes through
 ([Adv[Location]])
- 5 **Pattern:** [[Human | Machine]] filtrar [[Data]]
 Implicature: [[Human | Machine]] selects some [[Data]]
 to be shown or used, and prevents others from
 showing or using them
- 6 **Pattern:** [[Human = Footballer]] filtrar
 [[Action = Pass]]
 Implicature: [[Human = Footballer]] makes an [[Action = Pass]],
 taking the ball and passing it to another footballer

It can be observed that pattern 2 is the one related to the terminological use: [[Human]] filters a [[Fluid]] or [[Wavelength]], which is prototypically a liquid: 'La técnica realizada, patentada en Suecia, consiste en **filtrar** la sangre ['blood'] del paciente' (IULA50). Pattern 3 and 4 are syntactic and semantic variations of pattern 2, and the other patterns are more semantically separated.

In the specialised corpus, there are 28 occurrences of *filtrar*. Two of them do not correspond to the prototypical meaning, but are related to pattern 1 in the general corpus, referring to filter information: 'No es que se tema el despido por **filtrar** una información...'. Surprisingly, there is only one case of pattern 3 ('Diferentes especies

de *Daphnia* se han empleado, con desigual éxito, para el control biológico de algas, dado el elevado volumen de agua ['water'] que **filtran**') and no cases of pattern 4. The rest of concordances coincide with pattern 2: 'Lo más práctico es colocarlos inmediatamente en el disolvente, después de **filtrar** el agua' ['water'].

As well as in the general corpus, the subject role is populated by the semantic type [[Human]] and the object role by [[Fluid]], though is restricted to *water* or related nouns such as *efluente* ('effluent'), as observed in the verb *depurar*. One pattern is established:

- 1 **Pattern:** [[Human]] filtrar [[Liquid = Water]]
 Implicature: [[Human]] makes a [[Liquid = Water]] go through
 a filter or similar object to clean or separate it
 from other stuffs

The pattern in AquaCorp coincides with pattern 2 in the general corpus. There is no difference between the general and the specialised pattern, neither in relation to the arguments nor to the implicature. The subject in both cases is [[Human]] and the object a [[Fluid]], though in the specialised corpus the verb shows a preference for a smaller set of lexical items in the object clause role. The action expressed by the verb is the same one in both corpora. There is no evidence of semantic specialisation. *Filtrar* presents a prototypical pattern used both in general and specialised texts of Environment, with a more limited number of collocates when it is used in scientific texts.

4.2.4 Tratar ('to treat')

Tratar appears in 23,200 occurrences in the IULA50. From the total of occurrences, a random sample of 300 corpus lines is selected in order to be analysed following the CPA procedure. A total of nine patterns are found:

- 1 **Pattern:** [NO SUBJ] tratarse {de [[Anything]]}
 Implicature: [[Anything]] has been previously discussed,
 and will now be discussed again
- 2 **Pattern:** [NO SUBJ] tratarse {de} {[to/INF [V] |
 that-CLAUSE]}
 Implicature: {to/INF [V] | that [CLAUSE]} is important
 or interesting
- 3 **Pattern:** [[Human]] tratar {de} {[to/INF [V] |
 that-CLAUSE]}
 Implicature: [[Human]] tries to {to/INF [V] | that [CLAUSE]}
- 4 **Pattern:** [[Human | Information Source]] tratar
 [[Anything = Topic]]
 Implicature: [[Human | Information Source]] deals with
 [[Anything = Topic]]
- 5 **Pattern:** [[Information Source]] tratar {de} [[Concept]]
 | sobre [[Concept]]
 Implicature: [[Information Source]] deals with [[Concept]]
- 6 **Pattern:** [[Human 1]] tratar [[a Human 2]]
 Implicature: [[Human 1]] has a relationship with [[Human 2]]

- 7 **Pattern:** [[Human 1]] *tratar* [[a Human 2]]
[[Adv[Manner]]]
Implicature: [[Human 1]] behaves towards [[Human 2]]
in the way indicated by [[Adv[Manner]]]
- 8 **Pattern:** [[Human 1 = Health Professional]] *tratar*
[[a {Human 2 = Patient} | Injury]]
Implicature: [[Human 1 = Health Professional]] gives medical
treatment to [[Human 2 = Patient | Injury]]
- 9 **Pattern:** [[Human | {Process = Physical | Chemical}]]
tratar [[Stuff]]
Implicature: [[Human]] transforms a [[Stuff]] by
[[Process = Physical | Chemical]]

Among all these very different patterns, pattern 9 might be the most frequent in Environmental texts. In IULA50, the subject is whether a human or a physical or chemical process and the object treated is any kind of stuff—a solid, a wavelength or a fluid.

In AquaCorp 303 occurrences are found, from which 195 are associated to general patterns 1 and 2. These are discursive uses of the verb with the meaning of ‘to deal with’. Many examples of the verb *to treat* used in the sense of ‘to try’, corresponding to the general pattern 3, have also been found: ‘Cuando los ecólogos **trataban** de definir biocenosis [(they) tried to define biocenosis], buscaban entidades...’. The sense of ‘dealing with a topic,’ corresponding to the general patterns 4 and 5, is also quite frequent in Environmental texts: ‘Más constructivo es no **tratar** de la contaminación [(dealing with pollution)] desde el exterior’.

However, the verb *to treat* is used specifically in Environmental texts to refer to *water treatment* or other similar substances, in the sense indicated by pattern 9: ‘El proceso biológico generalmente utilizado para **tratar** los fangos primarios [(‘primary sludge’)] es la digestión anaerobia’. The collocates that normally populate the direct object slot are *agua/s* (‘water/s’), *agua/s residual/es* (‘waste water’), *caudal* (‘flow’), *compuesto* (‘compound’), *efluente* (‘effluent’), *entorno* (‘environment’), *fango/s* (‘mud’), *líquido* (‘liquid’), *materia* (‘matter’), *material* (‘material’), *residuo* (‘waste’), *vertido* (‘discharge’). They can be grouped in the semantic type [[Stuff]], more precisely, [[Liquid]] and, to a lesser extend, [[Solid]].

Though there are no great changes between the general and specialised use in relation to the arguments, apart from a more delimited range of lexical items, there is indeed a change concerning the meaning. Any stuff can be treated in general texts and in many specific domains in the sense of ‘putting the stuff through a process to transform it and use it for other purposes,’ for instance, *madera* (‘wood’), *cuero* (‘leather’) or *lana* (‘wool’). However, in Environmental texts related to water issues the sense is more precise, referring to ‘purify a dirty or polluted substance to be devoted to other uses’. The stuff is prototypically a ‘polluted’ substance that needs to be purified, so that it can be reused. The meaning is more precise than in general

language. Thus, pattern 9 determined in the general corpus has the following equivalent in Environmental texts:

- 1 **Pattern:** [[Human | {Process = Physical | Chemical}]]
tratar [[Liquid | Solid]]
Implicature: [[Human]] transforms a [[Liquid | Solid]] by
[[Process = Physical | Chemical]], manipulates it
to clean it and take it for other uses

This verb can be classified as a ‘phraseological verb’ but, as stated, its behaviour is not exactly as *filtrar*. The verb *tratar* exhibits a stronger phraseological tendency both in general and specialised texts.

The results obtained corroborate the idea that there are not only fuzzy boundaries between specialised and general uses, but also many gradations that are closely related to the terminological and phraseological tendency of verbs.

5 Conclusions

This paper has set out to show the possibilities of CPA in determining specialised uses of verbal lexical units in a Romance language such as Spanish. The CPA system has been proved to be very effective in relation to the following features:

- The CPA ontology is sufficient. The possibility of incorporating new concepts for a specific pattern results in a very flexible system adaptable to different languages.
- The metalanguage is also valid for Spanish.
- The general corpus compiled is sufficient for this first phase. It would be desirable, though, to refine the corpus in order to obtain a more balanced corpus in respect of the variety of themes and registers.

Our study has also demonstrated that CPA is a good methodology in determining specialised uses of verbal lexical units. CPA allows to determine the normal patterns of usage of each verb in general texts and specialised texts showing the conventional uses and the differences and similarities. The analysis developed shows that not all verbs behave in the same way. Each verb has their own particular patterns, sometimes used both in general and specialised texts without suffering any change in usage, sometimes with just slight changes and other times with domain-specific-patterns only used in specific texts. The particular patterns allow to differentiate the specialised from non-specialised uses, to determine whether these uses are context-dependent or not. The study highlights that verbs play a central role and deserve more attention in order to be able to understand specialised communication. CPA throws new light on the analysis of verbal lexical units in specialised discourses and on the aspects that should be emphasised in order to explain real usage and specific meanings of words in context.

The outcomes of CPA assist in the compilation of real corpus-driven specialised dictionaries. CPA helps to establish guidelines and criteria to decide on whether a verb must be entered in a specialised dictionary and which uses must be represented. From a lexicographical perspective, CPA allows to establish semantic divisions of an entry and relate them to a pattern as well as associate the corpus concordances to a specific pattern. It provides a systematic and coherent method for the dictionary-making process.

Finally, from a theoretical perspective, CPA highlights that lexical specialisation is also context-dependent and must be seen in terms of a gradation or continuum. This brings about the question of what it is really a *term*, as it might not always be the case that they designate a precisely defined concept with exact boundaries. ✿

References

- ALONSO, Araceli (2008). «Environmental Terminology in General Dictionaries». In: BERNAL, Elisenda; DECESARIS, Janet (ed.). *Proceedings of the XIII EURALEX*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada, p. 929-935.
- ALONSO, Araceli (2009). *Características del léxico del medio ambiente en español y pautas de representación en el diccionario general* [PhD dissertation]. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- ALONSO, Araceli; TORNER, Sergi (2010). «Adjectives and collocations in specialized texts: lexicographical implications». In: DYKSTRA, Anne; SCHOONHEIM, Tanneke (ed.). *Proceedings of the XIV EURALEX*. Leeuwarden: Fryske Akademy, p. 872-881.
- AUSTIN, John Langshaw (1955). *How to Do Things with Words*. 2nd edition, ed. J. O. Urmson and Marina Sbisa (1962). Harvard: Harvard University Press.
- BACH, Carme; SAURÍ, Roser; VIVALDI, Jordi; CABRÉ, M. Teresa (1997). *El Corpus de l'IULA: descripció*. Papers de l'IULA. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- BOURIGAU, Didier; SLODZIAN, Monique (2000). «Pour une terminologie textuelle». *Terminologies Nouvelles*, n° 19, p. 29-32.
- CABRÉ, M. Teresa (1999). *La terminología. Representación y comunicación. Elementos para una teoría de base comunicativa y otros artículos*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada. (Sèrie Monografies; 3.)
- ESTOPÀ, Rosa; LORENTE, Mercè; FOLGUERÀ, Rosa-Anna (2002). «El rol de los adjetivos en los textos especializados». In: *Actas del VIII Simposio Iberoamericano de Terminología*. Cartagena de Indias: Red Iberoamericana de Terminología.
- FABER, Pamela; MÁRQUEZ-LINARES, Carlos; VEJA-EXPÓSITO, Miguel (2005). «Framing Terminology: A process-oriented approach». *META* 50 (4) [on line]. <http://www.erudit.org/livre/meta/2005/000255co.pdf> [Consulta: 31 01 2010].
- FIRTH, John R. (1957). *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.
- HALLIDAY, Michael A. K. (1961). «Categories of the theory of grammar». In: KRESS, Gunther. R. (ed.) (1961). *Halliday: System and function in language*. Oxford: Oxford University Press, p. 52-72.
- HANKS, Patrick (2002). «Mapping Meaning onto Use». In: CORRÉARD, Marie Hélène (ed.). *Lexicography and Natural Language Processing: a Festschrift in honour of B. T. S. Atkins*. United-Kingdom: Euralex, Göteborg University, p. 156-198.
- HANKS, Patrick (2004a). «The syntagmatics of metaphor and idiom». *International Journal of Lexicography* 17 (3), p. 245-274.
- HANKS, Patrick (2004b). «Corpus Pattern Analysis». In: WILLIAMS, Geoffrey; VESSIER, Sandra (ed.). *Proceedings of the XI EURALEX*. Lorient: Université de Bretagne-Sud, p. 87-97.
- HANKS, Patrick (2010a). «How People Use Words to Make Meanings». In: SHARP, Bernadette; ZOCK, Michael (ed.) (2010). *Proceedings of the 7th International Workshop on Natural Language Processing and Cognitive Science, NLPCS 2010*. Madeira, p. 3-13.
- HANKS, Patrick (2010b). «Terminology, Phraseology, and Lexicography». In: DYKSTRA, Anne; SCHOONHEIM, Tanneke (ed.). *Proceedings of the XIV EURALEX*. Leeuwarden: Fryske Akademy, p. 1299-1308.

- HANKS, Patrick (2013). *Lexical Analysis: Norms and Exploitations*. Massachusetts: The MIT Press.
- HARRIS, Zellig (1954). «Distributional structure». *Word*, 10(23), p. 146-162.
- HOEY, Michael (2005). *Lexical Priming. A New Theory of Words and Language*. London: Routledge.
- HUNSTON, Susan; SINCLAIR, John (2003). «A local grammar of evaluation». In: HUNSTON, Susan; THOMPSON, Geoff (ed.) (2003). *Evaluation in Text: Authorial stance and the construction of discourse*. Oxford: Oxford University Press, p. 74-101.
- KILGARRIFF, Adam; RYCHLÝ, Pavel; TUGWELL, David (2004). «The Sketch Engine». In: WILLIAMS, Geoffrey; VESSIER, Sandra (ed.). *Proceedings of the XI EURALEX*. Lorient: Université de Bretagne-Sud. p. 105-116.
- LORENTE, Mercè (2001). «Tipología verbal y textos especializados». In: GONZÁLEZ, Miguel; SOUTO, Montserrat (ed.) (2001). *Cuestiones conceptuales y metodológicas de la lingüística*. Santiago de Compostela: Universidade de Santiago de Compostela, p.143-153.
- LORENTE, Mercè (2007). «Les unitats lèxiques verbals dels textos especialitzats. Redefinició d'una proposta de classificació». In: LORENTE, Mercè; ESTOPÀ, Rosa; FREIXA, Judit; MARTÍ, Jaume; TEBÉ, Carles (ed.) (2007). *Estudis de lingüística i de lingüística aplicada en honor de M. Teresa Cabré Castellví*. Vol. 2: *De deixebles 2*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada, p. 365-380.
- LORENTE, Mercè (2009). «Verbos y fraseología en los discursos de especialidad». In: CASAS, Miguel (dir.); MÁRQUEZ, Raúl (ed.) (2009). *XI Jornadas de Lingüística: homenaje al profesor José Luis Guijarro Morales (Cádiz, 22 y 23 de abril de 2008)*. Cádiz: Universidad de Cádiz, p. 55-84.
- MEYER, Ingrid (2000). «Computer Words in Our Everyday Lives: How are they interesting for terminography and lexicography?». In: HEID, Ulrich; EVERT, Stefan; LEHMANN, Egbert; ROHRER, Christian (ed.) (2000). *Proceedings of IX EURALEX*. Stuttgart: Universität Stuttgart, p. 39-57.
- MYKING, Johan (2007). «No fixed boundaries». In: ANTIA, Bassey Edem (ed.) (2007). *Indeterminacy in Terminology and LSP: Studies in Honour of Heribert Picht*. Amsterdam: John Benjamins, p. 73-91.
- RENAU, Irene (2012). *Gramática y diccionario: Las construcciones con se en las entradas verbales del diccionario de español como lengua extranjera [PhD dissertation]*. Barcelona: Universitat Pompeu Fabra. Institut Universitari de Lingüística Aplicada.
- RENAU, Irene; BATTANER, Paz (2012). «Using CPA to represent Spanish pronominal verbs in a learners' dictionary». In: VATVEDT FJELD, Ruth; TORJUSEN, Julie Matilde. *Proceedings of the XV EURALEX*. Oslo: University of Oslo, p. 350-361.
- SINCLAIR, John (ed.) (1987). *Looking Up: An account of the COBUILD Project in Lexical Computing*. London: Collins.
- SINCLAIR, John (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- TEMMERMAN, Rita (2000). *Towards New Ways of Terminology Description. The Sociocognitive-Approach*. Amsterdam: John Benjamins.
- WÜSTER, Eugen (1979). *Einführung in die Allgemeine Terminologielehre und Terminologische Lexikographie*. Viena: Springer.

Notes

1. Research for this article has been supported by the Spanish National Project HUM2009-07588/FILO: 'Agrupación semántica y relaciones lexicográficas en el diccionario' (J. DeCesaris, dir.), the LiCoRN Research group of the Laboratoire HCTI (Université Bretagne-Sud) and the Spanish Ministry of Science and Technology as part of the National Mobility Programme of Human Resources of the R&D National Programme 2008-2011 which has made possible the post-doctoral work of one of the authors. We would specially like to thank Patrick Hanks for his time and dedication to show us the CPA procedure and for all his teaching and clarification of many points, Geoffrey Williams for his constant support and careful proofreading and Mercè Lorente and the anonymous evaluators for their valuable comments.