

## The polysemy of the Spanish verb *sentir*: a Behavioral Profile analysis

**Abstract:** This study investigates the intricate polysemy of the Spanish perception verb *sentir* ('feel') which, analogous to the more-studied visual perception verbs *ver* ('see') and *mirar* ('look'), also displays an ample gamut of semantic uses in various syntactic environments. The investigation is based on a corpus-based behavioral profile (BP) analysis. Besides its methodological merits as a quantitative, systematic and verifiable approach to the study of meaning and to polysemy in particular, the BP analysis offers qualitative usage-based evidence for cognitive linguistic theorizing. With regard to the polysemy of *sentir*, the following questions were addressed: (1) What is the prototype of each cluster of senses? (2) How are the different senses structured: how many senses should be distinguished – i.e. which senses cluster together and which senses should be kept separately? (3) Which senses are more related to each other and which are highly distinguishable? (4) What morphosyntactic variables make them more or less distinguishable? The results show that two significant meaning clusters can be distinguished, which coincide with the division between the middle voice uses (*sentirse*) and the other uses (*sentir*). Within these clusters, a number of meaningful subclusters emerge, which seem to coincide largely with the more general semantic categories of physical, cognitive and emotional perception.

**Keywords:** Polysemy, lexical semantics, usage-based, Behavioral Profile, cluster analysis, Spanish, perception verbs, *sentir*

### 1. Introduction

The act of perception, a mechanism whereby a physical stimulus of the external world is captured, decoded and interpreted by a conscious entity, can rightly be considered one of the most sophisticated processes of nature. Indeed, as a fundamental cognitive process, the phenomenon of perception has received considerable attention among scientists in various research fields: from mathematicians and physicists, philosophers, psychologists, anthropologists to linguists, each of them from their own perspective and with different objectives, but all of them united by the same purpose to answer the question: What is perception? Moreover, since language is used primarily to communicate about the world we perceive, language and perception are inextricably interwoven (Miller & Johnson-Laird 1976). This relation is reflected in the rich linguistic bibliography and the numerous studies dedicated to verbs of perception (cf. for instance Enghels 2007; Hanegreefs 2008; Ibarretxe-Antuñano 1999; Vesterinen 2010 among many others). Indeed, the amount of interest in the

---

\* **Corresponding author: Marlies Jansegers:** Department of Linguistics, Ghent University.

E-mail: Marlies.Jansegers@UGent.be

**Clara Vanderschueren:** Department of Linguistics, Ghent University.

E-mail: Clara.Vanderschueren@UGent.be

**Renata Enghels:** Department of Linguistics, Ghent University.

E-mail: Renata.Enghels@UGent.be

study of this verbal category should not be surprising, since verbs of perception are the most tangible evidence of the direct interaction between physical perception and its linguistic codification (Fernández Jaén 2012: 155).

However, a closer look at that bibliography indicates that the degree of attention dedicated to the range of perception verbs is strikingly disproportional, since most studies tend to focus on visual and – to a lesser degree – auditory perception verbs (*see, look, hear, listen*). Consequently, these ‘prototypical’ perception verbs serve as a model for the description of other ‘inferior’ perception modalities such as olfaction, taste or touch (Enghels 2007: 5). This predominance of studies on visual and auditory perception is not unexpected since it reflects the fact that these ‘dominant’ perception modalities act in our Western culture as primary sources of objective information (Viberg 1984: 136; Sweetser 1990: 38). Undoubtedly, we tend to rely more on the information that we receive through our eyes, which is clearly illustrated by expressions such as ‘seeing is believing’. The existence of this kind of expressions also shows that verbs of perception frequently foster metaphorical mappings from concrete or physical meanings onto more abstract, mental domains, revealing in this case the privileged connection between visual perception and the field of cognition as an essential part of its polysemous character. Moreover, this complex polysemy also correlates with an ample gamut of syntactic contexts, as opposed to other verbs whose cognitive poverty results in a simpler – hence somehow less interesting – profile. Although the number of studies on the polysemy of this verbal category has increased considerably over the past years, one interesting perception verb has thus far hardly been discussed in Romance languages, including Spanish: *sentir* (‘feel’).

Indeed, comparable with the more-studied ‘prototypical’ perception verbs, Spanish *sentir* also displays a rich profile both semantically – ranging from meanings of direct physical perception (1), over cognitive perception (2) and more subjective or emotive meanings (3), to a rather discursive use (4) – and syntactically, combining with different kinds of complements.

- (1) Y ella **siente** la mano de Tomasa. [Chacón, 2002]  
‘And she *feels* Tomasa’s hand.’
- (2) Para que **sintieran** que su poder es en sí mismo frágil y pedante. [Prensa, 1995]  
‘So that they could *sense* that their power is in itself fragile and pedantic.’
- (3) Darse cuenta de esto le sirvió a Indalecio para **sentir** cierta ternura [...] por aquella chica rica. [Pombo, 2004]  
‘Realizing this helped Indalecio *feel* a certain tenderness for that rich girl.’
- (4) **Lo siento**, señor, pero [...] me encontraba un poco despistada. [Giménez Bartlett, 2002]  
‘*I am sorry*, Sir, but I was a bit distracted’

As a consequence, the fundamental question arises as to how these different senses relate to each other and how this intricate polysemy can be described in a uniform way.

More particularly, the following questions are posed: (1) Can we identify a prototypical meaning of the verb? (2) How many different meanings can be distinguished and which ones are more closely related than others? (3) What is the specific semantic structure of the network of *sentir*? (4) How do the semantic differences correlate with the morphosyntactic

behavior of the verb? Additionally, how can all these questions be approached in an empirical, systematic and verifiable way?

In practice, as for other linguistic fields, within the area of semantics several authors have been arguing for the need to pursue corpus-linguistic methods in order to avoid subjective interpretation (cf. among others Gibbs 2007; González-Márquez et al. 2007; Divjak 2010a-b; Geeraerts 2006, 2010; Glynn and Fischer 2010; Stefanowitsch 2010; Glynn and Robinson 2014). However, as Glynn (2010b: 90) points out, the application of such empirical, quantitative methods to the study of semantics in general, and polysemy in particular, is not straightforward. Indeed, how can meaning – as an intrinsically subjective and non-observable phenomenon – be investigated by means of quantitative methods?

Hence the main objective of the present paper is to contribute to addressing this general methodological concern identified in studies of polysemy from the empirical and quantitative points of view. More specifically, we propose to use the ‘behavioral profile’ (BP) approach (e.g. Divjak 2003, 2006; Divjak and Gries 2006; Gries 2006; Gries 2010a; Gries and Divjak 2009) in order to disentangle the intricate polysemy of the Spanish perception verb *sentir*. The outline of this paper is as follows: Section 2 provides an overview of the previous studies both of polysemy in general and of a Romance comparative study of *sentir* in particular. Through the application and combination of various methods, it will be shown how this study provides a gradual refinement of the semantic profile of the verb but at the same time raises some fundamental questions related to the polysemy of the Spanish verb. Section 3 discusses the method and data used in this study. Section 4 presents the results of a ‘behavioral profile’ of the verb, leading to a further refinement of the description of the (intralinguistic) polysemy patterns of *sentir*.

## **2. Antecedents**

### **2.1 The study of polysemy in linguistics: from burgeoning to nothingness and back again**

The interest in relations between words and meanings originates as early as Greek Antiquity and has kept many philosophers enthralled over the centuries. However, systematic linguistic research on the multiplicity of meaning only started tentatively in the mid-19<sup>th</sup> century by “linguists interested in meaning from the point of view of etymology, historical lexicography or historical semantics” (Nerlich and Clarke 1997: 351), and it is only until quite recently that the question of polysemy has risen to the fore in linguistic semantics (Cuyckens and Zawada 1997: xi). In the following sections we will first briefly describe the early treatment of polysemy in cognitive semantics and the methodological difficulties involved (Section 2.1.1.) before investigating how these problems can be addressed in a unified way (Section 2.1.2).

#### **2.1.1 Polysemy in early Cognitive Semantics**

Based on the insights of both the philosophy of language on family resemblance (Wittgenstein 1953) and the results from psychological research on categorization (Rosch 1978), cognitive semantics developed a prototype approach to word meaning. This prototype perspective in semantics involves the view of meaning as categorization and, thus, of lexical items as categories, which gave rise to notions such as *radial categories* (cf. the example of *mother* in

Lakoff 1987: 74-76) and *family resemblance categories* (cf. Wittgenstein's 1953 discussion of the word *Spiel*, 'game'). Once entering the field of semantics, the prototype approach was extrapolated to the study of polysemy and consequently, polysemous items were essentially considered to be categories of senses, interrelated by means of family resemblance and centered around a prototype (cf. Cuyckens and Zawada 1997: xii-xiii; Gries 2015).

A well-known representational formalism for such prototype-based polysemous structure is the radial network model (see Brugman 1988 or Brugman and Lakoff 1988 for a more detailed account). Its format has been adopted to various degrees by different authors, but all semantic network studies share the idea of (i) employing encyclopedic semantic features, (ii) without the notion of necessary and sufficient features for category membership, (iii) in order to distinguish senses and relate forms (cf. Glynn 2014: 18-19). However, this so-called 'full-specification approach' (Lakoff 1987) has been criticized, and it was demonstrated that the study of polysemy in cognitive semantics had inherited both the theoretical – discrete senses (cf. Geeraerts 1993) – and the methodological technique – introspection (cf. Sandra and Rice 1995) – of Structuralism. In sum, although Cognitive Linguistics theorizing has propounded a usage-based model of language as well as the commitment to empiricism and inductive research since its inception, in practice, the radial network analysis did not meet that purpose. These critiques cleared the way for the more experimental and corpus-driven methods in semantic analysis (Berez & Gries 2009).

However, the application of such empirical methods to semantics is not straightforward, given that the oft-cited main advantage of corpus linguistics, that is the possibility of quantifying results, raises two fundamental questions (Glynn 2010b: 90). First, how can meaning – as an intrinsically subjective and non-observable phenomenon – be investigated by means of quantitative methods? Second, what motivates this type of investigation? With respect to the latter question, Glynn (2010b) argues that quantification permits verification and the testing of hypotheses, thus facilitating the 'empirical cycle' (cf. Geeraerts 2010).

As to the former question, as Glynn (2010c: 240) points out, it is a severe misunderstanding to think that corpus-driven research and the quantitative assessment of its results is more objective than other methods, such as introspection. Indeed, the annotation of corpus data requires classificatory choices which are not always entirely objective. And this applies particularly to semantics: meaning is a non-observable relation in our mind and is therefore beyond the reach of absolute objectivity. However, it should be emphasized that quantitative analysis is not primarily aimed at objectivity but rather at a better and more principled way of verification of the results (Glynn 2010c: 242). Therefore, the main question that remains to be answered is how meaning can be defined in a measurable way, and thus be operationalized.

### 2.1.2 Towards a usage-based approach to polysemy

The answer to the slightly thorny question of operationalization of meaning lies precisely in two theoretical tenets at the root of Cognitive Linguistics, namely 'entrenchment' and 'conceptual categorization' (Glynn 2010a, 2014).

First, within the theory of entrenchment, Langacker (1987, 1988) correlates the frequency of occurrence of linguistic phenomena with their entrenchment in the cognitive system: the more frequent a form-meaning pair is used, the more entrenched it is in the speaker's

knowledge. Thus, through the notion of entrenchment, frequency operationalizes grammaticality.<sup>1</sup> Similarly, the question arises if meaning can also be operationalized.

Following Fillmore (1985), Lakoff (1982, 1987) advocated a theory of semantics based on world knowledge – encyclopedic semantics – as opposed to linguistic semantics in its narrow sense. More precisely, this implies a holistic approach to meaning which abandons the strict division between linguistic semantics and context pragmatics on the one hand and, between lexis and syntax on the other. There are no clear-cut distinctions between the distinct language levels, which are in constant interaction, and it is precisely this intricate merger that constitutes what is meaning. In an attempt to operationalize this complex encyclopedic meaning, Lakoff (1987: 5-6) recurs to the notion of ‘conceptual categorization’: every time we perceive and/or reason about *kinds* of things, we make use of categories based on shared properties. And therefore, categorization can be defined as a symbolic distinction between difference and similarity.

Yet, how can this abstract definition of conceptual categorization induce quantification and thus allow for an operationalization of meaning? Glynn (2010a, 2014) observes that this is already the underlying assumption that drives much corpus-driven research, given that corpus investigation essentially functions by grouping similar things (based on co-occurrence and correlation), distinct from those that are not similar.<sup>2</sup> From this, Glynn (2010a: 8) concludes that just as frequency can operationalize grammaticality, co-occurrence can operationalize categorization. Consequently, frequency of co-occurrence, which underlies all corpus investigation, is a quantitative operationalization of the fundamental theories of Cognitive Linguistics – entrenchment and categorization. Thus, in accordance with the broad scope of encyclopedic semantics explained above, this principle of co-occurrence must also be understood in a broad sense and goes beyond the dimension of formal co-occurrence in order to account for the complex interaction of all dimensions of meaning.

This aim to elucidate the complex, multidimensional nature of meaning is precisely the starting point of the Behavioral Profile approach (BP). Being corpus-based, the BP approach builds on the idea that corpus data provide distributional frequencies and that distributional similarity reflects functional or semantic similarity. Behavioral profiles and the proposed methods for their evaluation provide an ideal starting point for research concerning interfaces between different levels of linguistic analysis, and thus, for the study of the multidimensional nature of language (Gries & Divjak 2009).

In the present study, we will apply this BP approach to the polysemy of *sentir*. In order to fully appreciate the complexity and the particularity of the Spanish verb, its semantics will be first situated and discussed within the broader scope of its French and Italian cognates by applying various methods. In Section 2.2 it will be shown in a more concrete way how the study of the semantics of *sentir* also poses serious problems, and how the BP approach seems to be a promising answer to the methodological question.

---

<sup>1</sup> It must be mentioned that this assumption of a direct correlation between the frequency of occurrence of linguistic phenomena and their salience or entrenchment in the cognitive system, has been challenged (cf. Gilquin 2008; Arppe et al. 2010; Schmid 2010). We agree with Glynn who stresses that the frequency-based approach to entrenchment is indeed only one possible operational definition and that other operationalizations of the relationship between form and meaning may also be considered.

<sup>2</sup> For example, in order to evaluate their degree of attraction or repulsion, Collostructional Analysis (Stefanowitsch and Gries 2003) considers the formal co-occurrence of lexical items and their constructions.

## 2.2 *Sentir*: a Romance comparative study

A previous study (Enghels & Jansegers 2013) identified *sentir* as an interesting case within the spectrum of perception verbs in the Romance languages. By comparing the complex semantics of the etymologically related cognate verbs *sentir(e)* in Spanish, French and Italian, it has been found that the rich polysemy of these verbs only partly coincides, and that in each language the verbs have undergone semantic specializations. This cross-linguistic analysis has been particularly informative in that it implied a continuous methodological concern, taking into account the particular nature of the phenomenon of polysemy as argued in Section 2.1. Indeed, a lexicographic analysis was complemented by the results of a combined corpus approach involving a translation and comparable corpus. In what follows, the main outcomes of each case study are summed up, paying special attention to the methodological constraints which each method implies (Section 2.2.1) and which need to be met by the BP approach (Section 2.2.2).

### 2.2.1 Previous analyses

A first step to identify the *tertium comparationis* and the degree of equivalence of the verbs *sentir(e)* is studying the definition of their semantics with various dictionaries in the three languages concerned.<sup>3</sup> In analysis at a coarse-grained level the verbs were found to similarly cover a wide range of meanings linked to different semantic verb classes (amongst which perception, cognition and emotion), but they concurrently appeared to exhibit several language-specific features and uses. However, the lexicographic analysis left us with two main problems.

First, the information provided by these sources does not specify the precise extent of cross-linguistic differences and similarities in the polysemy of the verbs. Do they relate more closely to each other in some specific domains than they do in others? How can the distance between them be defined on a more accurate basis? Second, the consulted dictionaries list a quite unorganized inventory of definitions, without distinguishing between the major and minor meanings or more or less related meanings of the verbs. On the basis of these data it is thus quite impossible to reconstruct their polysemy, let alone to identify the prototypical meaning(s) of these verbs. It is precisely these limitations of the introspective lexicographic analysis that urged the need to apply other empirical methodologies.

In an attempt to measure the degree of equivalence between the three verbs and at the same time to gain better insight into the polysemy of each individual verb, a parallel multilingual corpus was compiled and subjected to a *Mutual Translation Correspondence Analysis* (Enghels/Jansegers 2013: 964-965). This method can be summarized as follows:

When Target Language<sub>1</sub> translates the semantic context of the Source Text by means of the verb *sentir(e)*, what are the correspondents in Target Language<sub>2</sub>? And conversely: When Target Language<sub>2</sub>

---

<sup>3</sup> The consulted dictionaries are *Diccionario del Español Actual* (Spanish), *Le Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française* (French) and the *Grande dizionario Italiano dell'uso* (Italian). See Bibliography for more details.

translates the semantic context of the source text by means of the verb *sentir(e)*, what are the different ways the same translation unit is expressed in Target Language<sub>1</sub>?

This analysis provided some general quantitative data. For instance, it showed that the *tertium comparationis* at its most basic level can be defined as ‘general physical perception without any modality of perception being specified’ (5):

- (5) a. Harry **felt** the warmth wash over him as though he'd sunk into a hot bath. (*Harry Potter and the Philosopher's stone*)
- b. Harry **sintió** que el calor lo cubría como si estuviera metido en un baño caliente. (Spanish)
- c. Harry **sentit** la chaleur se répandre autour de lui comme s'il venait de plonger dans un bain tiède. (French)
- d. Harry **sentì** il calore inondarlo come se si fosse immerso in un bagno caldo. (Italian)

Moreover, the idea of intra-linguistic semantic specializations, generated by the lexicographical approach, was confirmed but further refined:

- 1. In French, the verb is used most frequently in its cognitive meaning, often implying a sense of intuition (6);
  - 2. In Italian *sentire* most prototypically refers to auditory perception (6);
  - 3. The Spanish equivalent refers to the emotional meaning ‘regret, deplore’ in a unique way (7).
- 
- (6) a. He had **sensed** rather than **heard** it: someone or something was standing in the narrowgap between the garage and the fence behind him. (*Harry Potter and the prisoner of Azkaban*)
  - b. Il l'avait **sentì** plus qu'**entendu**: quelque chose ou quelqu'un se trouvait dans l'espace étroit entre le muret et le garage de la maison devant laquelle il s'était arrêté. (French)
  - b. Más que **oírlo**, lo **intuyó**: había alguien detrás de él, en el estrecho hueco que se abría entre el garaje y la valla. (Spanish)
  - c. Lo **avvertiva**, più che **sentirlo** con le orecchie: c'era qualcuno o qualcosa lì nello stretto passaggio tra il garage e la staccionata alle sue spalle. (Italian)
  - (7) a. **I'm sorry**, Potter, but that's my final word. (*Harry Potter and the prisoner of Azkaban*)
  - b. **Lo siento**, Potter; pero es mi última palabra. (Spanish)
  - c. **Je suis désolée**, Potter, reprit-elle, mais c'est mon dernier mot. (French)
  - d. **Mi dispiace**, Potter, ma è la mia ultima parola. (Italian)

Thus, Mutual Translation Correspondence Analysis of parallel corpus data proves to be suitable for the study of the degree of equivalence of cognate verbs in closely related languages. Given that the translated texts share the same meaning, one can be sure to compare similar semantic contexts. By observing that the equivalents of *sentir(e)* in all three languages fall into the same categories, the possible meanings that define the polysemy have been defined more precisely. The general semantic nuclei that emerged from the translation corpus were perception, cognition and emotion. However, a comparable corpus study was necessary to gain a clearer view of the spread and relative frequency of these nuclei for each individual verb.

Therefore, in a third phase, three monolingual samples were collected and manually annotated with the same semantic categories classified by the parallel corpus research. This corpus study

quantitatively confirmed the above-mentioned semantic specializations, but, to a much greater extent than the translation corpus did, hinted towards many ‘bridging contexts’ between the nuclei, thus suggesting continuity and gradualness in the polysemy of the verbs. To give an example, in (8) it is not possible to unambiguously classify the verb, given the presence of the NP *sabor a mala leche* (lit. ‘taste of bad milk’) evoking gustative perception but in fact referring to the idea of a bad feeling (emotion) of the speaker. Still, the PP *en mi propia voz* (‘in my own voice’) suggests an auditory perception meaning:

- (8) Debes de estar impaciente -dije, **sintiendo** el sabor a mala leche en mi propia voz, una voz insolente que no sabía de dónde venía. [Zafón C., 2003]  
‘You must be impatient, I said, *feeling* the taste of bad blood (lit. milk) in my own voice, an insolent voice, I didn’t know where it came from.’

These fuzzy boundaries between major semantic categories will be shown to identify the highly fertile areas for other metaphorical and pragmatic uses of the verb (cf. Section 4.3).

### 2.2.2 Methodological problems

However, the panorama of the semantics of *sentir(e)* in Romance and more particularly in Spanish, as it appears from these three case studies, still raises a number of research questions related to the general theme of polysemy. These problems concern:

1. The identification of the prototypical meaning: on the basis of which criteria can the prototypical meaning of the verb be defined?
2. The degree of distinction between different meanings: how many different meanings can be distinguished and how can one determine whether two meanings are clearly different or merely shades of meaning (cf. the well-known lumping vs. splitting issue)?
3. The structure of the semantic network: which meanings are more closely related to each other than others, and how does this influence the semantic network describing the polysemy of the verb?
4. Morphosyntactic correlates: how do semantic differences correlate with the morphosyntactic behavior of the verb?

In sum, the main challenge is to approach and study this complex polysemy in a more verifiable way. Therefore, and in order to minimize subjective and introspective knowledge, a corpus-based behavioral profile approach to the polysemy of the verb will be employed (e.g. Gries 2006, 2010). In what follows, we will present the general characteristics of the approach before applying the method to the study of *sentir* in Spanish.

## 3. Towards a behavioral profile analysis: method and data



As we have seen in the previous section, there are significant drawbacks to the intuitive approach. For several reasons, the BP method is a more objective and verifiable alternative to those intuitive approaches to semantics.<sup>4</sup>

First, the analysis implies manual annotation of objectively measurable characteristics. As the BP approach constrains one to code all the corpus examples for the same set of variables, it is extremely precise and explicit, yielding descriptions at a very high level of precision (cf. Divjak 2010a: 120). As a consequence, it can be argued that the method helps to minimize the share of subjective, implicit knowledge.<sup>5</sup>

Second, as mentioned in Section 2, linguistic phenomena have traditionally been described in terms of (im)possibility, by using, for example, the minimal pair test. This implies that a particular entity is either or not a full member of the category defined by a word. However, it has already been revealed that a graded phenomenon such as polysemy does not easily lend itself to this kind of tests. A fine-grained corpus-based analysis, by contrast, facilitates observing several parameters simultaneously and evaluating them in terms of probability instead of (im)possibility. In order to deal with an elusive concept such as meaning from a probabilistic point of view, it needs to be quantified, and thus operationalized, which highlights the challenge of the operationalization of meaning (cf. Section 2). Having risen to this challenge, the BP approach starts from the distributional hypothesis: distributional similarity reflects functional similarity. Therefore, it has been argued that behavioral profiles and the proposed methods for their evaluation provide an ideal starting point for research concerning interfaces between different levels of linguistic analysis, e.g. the syntax-lexis interface (cf. Divjak 2010a: 11; Divjak & Gries 2006: 52). Further, it has been argued that behavioral profiles are valuable for the analysis of polysemous items because they offer usage-based evidence for cognitive linguistic theorizing concerning network representations, prototypicality of senses, sense-distinctions and the polysemy-homonymy discussion (Gries & Divjak 2009: 72).

More precisely, the Behavioral Profile approach involves four steps (cf. Gries 2010a; Gries & Divjak 2009; Gries & Divjak 2010; Gries & Otani 2010):

---

<sup>4</sup> The Behavioral Profile approach, labeled as such by Gries, can be placed within the larger and ever-expanding family of corpus-driven quantified and multifactorial methods or the ‘Quantitative Multifactorial method’ in terms of Glynn (2009). Within the broad realm of Cognitive Semantics, the use of such quantitative, multifactorial and usage-based methods may be found, for example, in (the references are exemplary, not exhaustive): Geeraerts et al. (1994), Gilquin (2003, 2006), Gries 2003, Grondelaers and Geeraerts (2003), Newman and Rice (2004, 2006), Wulff (2006), Glynn (2009) see also many references therein). We consider this an umbrella term to cover a wide range of different approaches sharing the fine-grained annotation of corpus examples, but differing considerably in the ways in which this annotation is statistically explored. In the present paper, we explicitly apply one particular method, namely the Behavioral Profile approach as developed by Gries and Divjak.

<sup>5</sup> However, one could reasonably reject a corpus-based approach to meaning, objecting that there will always be some degree of introspection in the analysis of corpus data (an example of that corpus-critical point of view can be found in Raukko 1999, 2003). Of course, as Berez & Gries (2009: 158) point out, the analysis of corpus data is inextricably bound up with classificatory decisions which are not always entirely objective, and therefore a completely objective classification of corpus data is very unlikely. Indeed, the choice of parameters to be included in the BP analysis and the subsequent interpretation of the results contain elements of subjectivity, but an important part of the analysis is still entirely objective precisely because all the occurrences are coded for the same set of variables. This procedure ensures that all the analyzed information is made explicit, avoiding suggesting intuition-based findings and facilitating a better and more principled way of verification of the results.

- (i) Retrieving all instances of the verb in context in the form of a concordance.

For this study of the polysemy of the verb *sentir*, all the instances of the verb were retrieved from the Spanish databank CREA, taking as selection criteria only time (2000-2004) and diatopic variant (peninsular Spanish). Following this method, we retrieved a total of 6742 instances of the verb. Then, in order to obtain a more practical and workable corpus, a representative random sample was selected corresponding to the 25% of this total, yielding 1686 occurrences. Because of the lack of oral data, the CREA corpus was then complemented with the oral data available for the 21th century extracted from the PRESEEA and the COLAM corpus.<sup>6</sup> The outcome of this data-gathering method is a corpus of 1810 instances in total.

- (ii) Analysing and annotating manually a large set of properties of each match of the verb in the concordance. These properties are termed *ID tags* (Atkins 1987) and include morphological, syntactic, semantic, and other characteristics.

For the present study, a wide range of parameters were distinguished according to four general levels of analysis, that is (1) the properties of the verb itself, (2) the argument structure of the verb, (3) the characteristics of other adjuncts, and (4) discourse phenomena. Table 1 presents an example of such ID tags and their levels:

GENERAL LEVEL	TYPE OF ID TAG	ID TAG	ID TAG LEVEL
<b>VERB</b>	morphosyntactic properties	tense	present, past, future, infinite form
		person	1, 2, 3
		number	singular, plural
<b>ARGUMENT STRUCTURE</b>	properties of the Subject: form	lexical S	with S, without S
	properties of the Direct Object: form	lexical DO	with DO, without DO
		type of DO: form	nominal phrase, pronoun, gerund, infinitive, proper noun, etc.
<b>ADJUNCTS</b>	properties of the adverbial adjuncts	referent DO	person, concrete entity, abstract entity, situation, ambiguous
		presence adverbial adjunct	with adverbial adjunct, without adverbial adjunct
		type of adverbial adjunct: form	adverb, prepositional phrase, nominal phrase, etc.
<b>DISCOURSE</b>	scope	predicational autonomy	yes, no

Table 1. Examples of ID tags and their levels

This methodology yields a total of 32 different ID tags and 153 ID tag levels.

<sup>6</sup> Real Academia Española: *Corpus de Referencia del Español Actual*. [www.rae.es](http://www.rae.es) [CREA]; *Proyecto para el Estudio Sociolingüístico del Español de España y América*. <http://preseea.linguas.net> [PRESEEA]; *Corpus Oral de Lenguaje Adolescente*. <http://www.colam.org> [COLAM].

- (iii) Converting these data into a co-occurrence table that provides the relative frequency of co-occurrence of each sense of the verb *sentir* (columns) with each ID tag level (rows). As such, the percentages of ID tag levels sum up to 1 within each ID tag, as illustrated in table 2. Each column represents a set of co-occurrence percentages for one sense of the verb, and it is precisely this vector of co-occurrence percentages that is called a ‘Behavioral Profile’.

ID tag	ID tag level	<i>experience</i> - <i>physical</i> <i>perception</i>	<i>experience</i> - <i>emotional</i> <i>perception</i>	<i>auditory</i> <i>perception</i>	<i>consider,</i> <i>judge</i>	...
tense	present	0.30	0.36	0.29	0.55	...
	past	0.35	0.40	0.53	0.30	...
	future	0.01	0.01	0.00	0.02	...
	infinite form	0.34	0.23	0.18	0.13	...
lexical S	with S	0.18	0.41	0.24	0.41	...
	without S	0.82	0.59	0.76	0.59	...
...	...	...	...	...	...	...

Table 2. Example of BP Vectors

- (iv) Evaluating this table through exploratory and other statistical techniques. In this case we performed a hierarchical agglomerative cluster analysis. This is an exploratory technique that is aimed at identifying and representing (dis)similarity relations between items in the form of a hierarchical tree diagram (dendrogram) representing several clusters that are characterized by high within-cluster similarity and low between-cluster similarity (for a general introduction to cluster analysis for linguists, see for example Baayen (2008: Chapter 5) and Gries (2009: Chapter 5, Section 5). The statistical procedure was performed by employing Gries’s (2010b) BehavioralProfiles 1.01 using the R statistical software package.<sup>7</sup>

The hierarchical cluster analysis of the data leads to the dendrogram shown in Figure 1:

<sup>7</sup> The choice for exploratory HAC statistics over confirmatory statistics (like regression modeling), in combination with a series of subsequent significance tests (see *infra*), is justified, given the complex nature of the data. Indeed, the different senses of *sentir* cannot be straightforwardly interpreted as different levels of a categorical response variable: (a) the high number of different senses and the infrequency of certain senses are not an ideal basis for statistic modeling; (b) the exact number of senses is not fixed (this is exactly what the HAC tries to unravel), and there are a number of ambiguous senses in between two or more meanings; (c) consequently, the different senses of one and the same verb are not a matter of choice by the speaker, i.e. the speaker does not deliberately opt for one or the other sense of *sentir*. This is thus a conceptual objection to regression analysis, which ideally involves the clear choice for one or the other lexeme or construction.

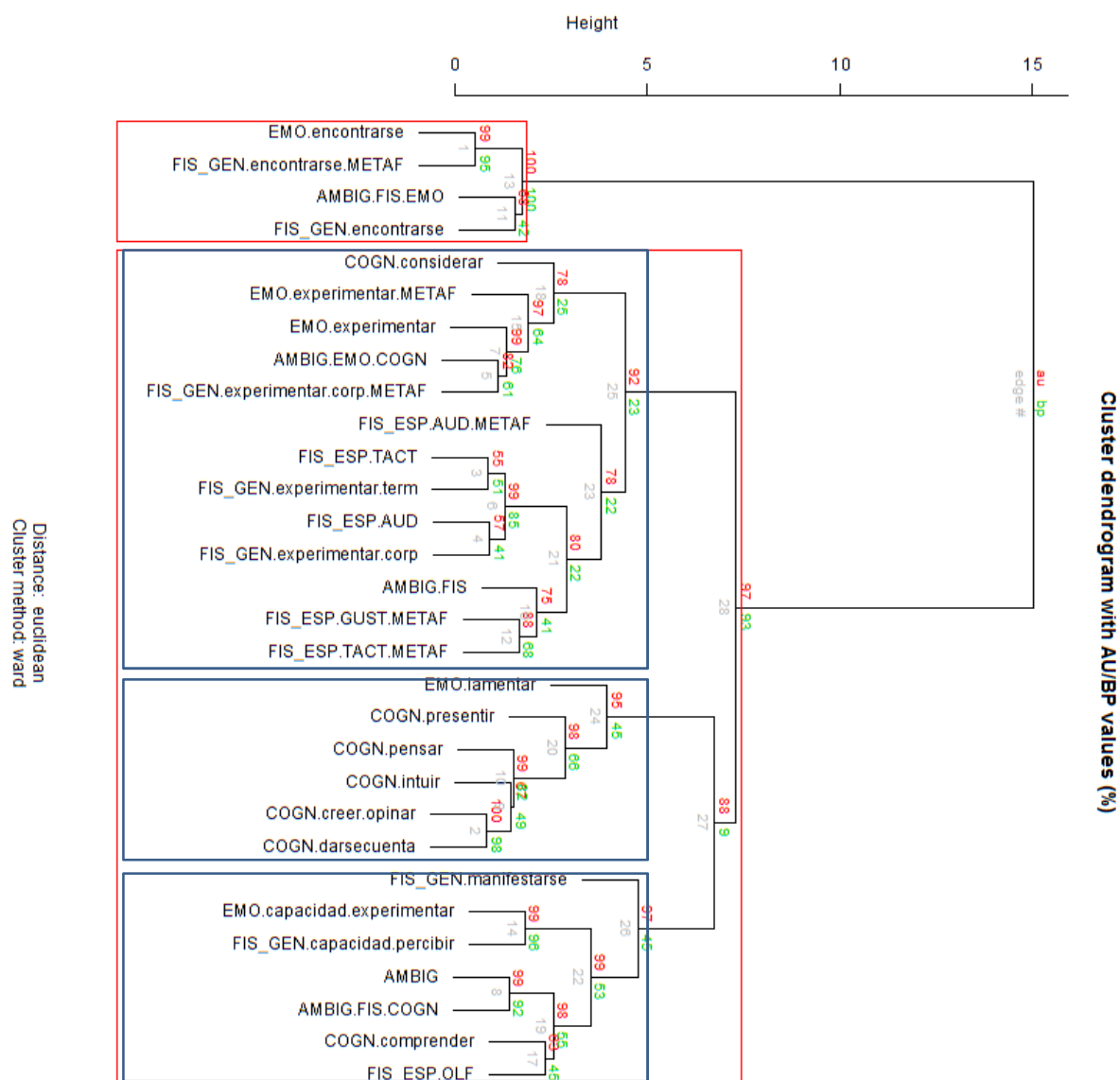


Figure 1. Dendrogram for the senses of *sentir* including the ambiguous cases.<sup>8</sup>

In the next section, we will discuss in detail the results of this cluster analysis of the verb *sentir* and its different senses and what it reveals about the intricate polysemy of the verb.

## 4. Analysis of *sentir*: results and discussion

### 4.1. General observations

<sup>8</sup> As a distance metric, we used the widely-used Euclidean distance for numerical data. For the creation of the tree, we followed Ward's minimum variance method (Ward 1963). A list of the different senses distinguished and their English paraphrase is provided in Appendix 1.

To begin with, the tree diagram in Figure 1 presents a clear image of how the different senses of *sentir* cluster together. As indicated through the two red boxes, two significant meaning clusters can be distinguished: the first one represents the cases of ‘encontrarse en un estado físico o psíquico’ (lit. ‘feel/perceive oneself to be in a physical or emotional state’, cf. resp. examples 9, 10), whereas the second large cluster groups together all the other senses of the verb.<sup>9</sup>

- (9) E Isabel de la Hoz, ahora, está de ocho meses, y **se siente** muy pesada y se le hinchon los tobillos, [...] [Pombo, 2004]  
 ‘And Isabel de la Hoz, now, is eight months [pregnant], and she *feels* very heavy and her ankles swell’
- (10) Personalmente tengo que decir que **me siento** muy satisfecho de haber trabajado con este hombre, discutido pero triunfador. [Del Rey del Val, 2002]  
 ‘Personally I have to say that I *feel* very pleased to have worked with this controversial but triumphant man.’

This bipartite structure thus clearly differentiates between a *sentirse*-cluster and a *sentir*-cluster, the first one corresponding to the middle voice uses and the second one to the active uses. Consequently, both uses call for a separate analysis and treatment. In the remaining sections of this paper, we will refer to these two main clusters as *sentirse* and *sentir*. The need to distinguish both clusters at this level is strongly supported by the very high Approximately Unbiased (AU) values (in red) reported in the dendrogram for these clusters (respectively 100% and 97%). These are p-values between 0 and 1, calculated through 1000 multiscale bootstrap resamplings with the R package *pv-clust*. High values indicate that there is strong evidence for the cluster in question.

Within the “active meaning” *sentir* cluster, indicated by the second red box, a number of subclusters arise, which, on the basis of the AU values above 90% (respectively 92%, 95% and 97%), are organized in three meaningful groups (cf. the three blue boxes within red cluster 2), namely (1) ‘experience a physical or mental feeling’ (‘experimentar una sensación física o psíquica’, cf. resp. examples 11, 12), (2) a cluster referring to more cognitive perceptions (cf. example 13) and (3) a cluster indicating the ability to experience or perceive something (example 14):

- (11)[...] pidió a un colega suyo que le extrajese una muela que tenía estropeada utilizando el gas de la risa, y no **sintió** dolor. [Sabadell, 2003]  
 ‘he asked his colleague to extract a tooth he broke using laughing gas, and he did not *feel* pain.’
- (12) Los marroquíes **sienten** una fuerte atracción hacia el lujo material [Silva, 2001]  
 ‘Morrocans *feel* a strong attraction towards material luxury.’
- (13) ¿Qué significado das a la Eucaristía? Como joven, ¿**sientes** que esta es el centro de tu vida? [Prensa, 2000]  
 ‘What meaning do you put on the Eucharist? As a young person, do you *feel* that this is the center of your life?’
- (14) Relájate, déjate acariciar, prueba a excitarte poco a poco. No me acaricies, no me toques, sólo disfruta y siente mis caricias. Cierra los ojos para **sentir** más en tu interior... [Llongueras, 2001]  
 ‘Relax, let yourself be caressed, try to excite yourself little by little. Do not caress me, do not touch me, just enjoy and feel my caresses. Close your eyes in order to *feel* more inside’.

<sup>9</sup> The same global picture with two large clusters can also be deduced from the lengths of the vertical lines in the dendrogram, as long vertical lines indicate more autonomous subclusters (cf. Gries 2009: 308).

Each cluster also contains ambiguous cases (indicated in the dendrogram as AMBIG.X) which present two (or more) concurring interpretations:

- (15) Todas aquellas casas, todo el campo que había atravesado para llegar hasta allí, toda aquella gente, lo que habían visto, lo que habían **sentido** y pensado, lo que habían hecho, aquel cielo, aquel aire... [Gavilanes, 2000]  
 ‘All those houses, the whole field he had crossed to get there, all those people, what they had seen, what they had *felt* and thought, that sky, that air....’

In this example, *sentir* can refer to physical (for example, physical pain or auditory perception), as well as to emotional (feelings such as fear, love) or cognitive perception, probably because of the coordination with the cognitive verb *pensar* ‘think’. As we will see in the following sections, these ambiguous cases turn out to be particularly revealing and interesting for the semantic processes underlying the polysemy of *sentir*.

In brief, the cluster analysis shows that two large significant meaning clusters can be distinguished (referred to as *sentirse* and *sentir*), which, predictably, coincide with the division between the middle voice uses (*sentirse*) and the other active uses (*sentir*). Within this second cluster, a number of meaningful subclusters emerge. In what follows, we will use this BP output in order to account for the four problems related to the study of the polysemy of *sentir* mentioned in section 2.2.2: the identification of a prototype (section 4.2), the degree of sense distinctiveness (section 4.3), the structure of the network (section 4.4) and the morphosyntactic correlates (section 4.5).

## 4.2. Prototype identification

In the first place, the question arises relating to if a prototypical sense of *sentir* can be defined. One way of approaching prototypicality is through frequency: the underlying rationale is that more prototypical senses simply occur more frequently (cf. Schmid’s 2000: 39 “From-Corpus-to-Cognition Principle”). Table 3 shows the four most frequent meanings. There turns out to be a striking predominance of emotional meanings: the three most frequent meanings all refer to the emotional pole of *sentir*. This result is not totally unexpected: former studies already emphasized the strong emotional meaning of contemporary Spanish *sentir* from both a cross-linguistic and a diachronic perspective (cf. Section 2.2 and Enghels & Jansegers 2013, Jansegers & Enghels 2013).

	Frequency		Variance
	#	%	
EMO.encontrarse	797	44.1	0.141
EMO.experimentar	391	21.6	0.128
EMO.lamentar	118	6.5	0.165
FIS_GEN.experimentar.corp	71	3.9	0.119

Table 3. Prototype. Frequency and Variance

This predominance of emotional perception is thus clearly reflected in the three most frequent senses. First, the middle voice uses (*sentirse*) turn out to be very productive. As illustrated in example 16, the marker *se* “indicates that the two semantic roles of Initiator and Endpoint refer to a single holistic entity” (Kemmer 1993:66). More precisely, *sentirse* describes an action that the subject performs upon himself, this subject being the experiencer of emotion and consequently the person par excellence to judge about his own feelings. Given this capability of the middle voice to bring into prominence the experiencer himself, it acts as an excellent host for the expression of personal feelings, and thus for the use of a lexical item like *sentir*.

Within the *sentir* cluster, the most frequent sense is the experience of a certain emotion (‘EMO.experimentar’, cf. example 17). Within the same cluster we see the very characteristic emotional meaning of the Spanish verb, namely the sense of regret (‘EMO.lamentar’) as illustrated in example 18. The general physical perception “experimentar una sensación física, corporal” only comes in the third place (cf. example 19).

- (16) Y me **sentí** feliz como un niño cuando su voz de reportero curtido en mil batallas comenzó a contarme lo que vio aquella jornada inolvidable en la que dio su gran noticia [...] [Jiménez, 2000]  
 ‘And I *felt* happy as a child when his experienced reporter voice that bore thousands of battles began to tell me what he saw that unforgettable day in which he told the big news.’
- (17) Al Douri, [...] subrayó que “nunca he **sentido** vergüenza [de representar a Iraq], siempre me he considerado como un servidor de mi país y mi pueblo, no del Gobierno”. [Prensa, 2003]  
 ‘Al Douri, [...] emphasized that “I have never *felt* embarrassment [to represent Iraq], I have always considered myself as a servant of my country and of my people, not of the government.’
- (18) Yo... No sé lo que me pasa. Me siento mal, muy mal, peor que nunca... Pero te quiero, Tam, y **siento** mucho haberme puesto así. [Grandes, 2002]  
 ‘I... don’t know what happens to me. I feel bad, really bad, worse than ever... But I love you, Tam, and I *regret* so much having behaved like that.’
- (19) Había apretado tanto los dientes que las mandíbulas me dolían, pero nada comparable al dolor agudo y penetrante que **sentía** en las costillas. [Giménez Bartlett, 2002]  
 ‘I had pressed my teeth so much that my jaws were aching, but nothing compared to the sharp and penetrating pain I *felt* in my ribs.’

In sum, according to frequency, the emotional sense of the verb turns out to be the most frequent one, and from that point of view, it can thus be considered as the most prototypical one.

However, as Gilquin (2008) rightly argues, the relationship between frequency and prototypicality is not always straightforward. In fact, prototypicality involves more than frequency because it also implies a high family resemblance with other members of the category (Rosch and Mervis 1975). Therefore, another way of handling prototypicality is by taking into account the multiplicity of contexts in which an element can occur: more prototypical elements are taken to be less formally constrained and thus to appear in a wider variety of contexts (cf. also Gries 2006, Divjak 2010a). On the basis of the BP vectors, we can test this by measuring the variance of the vectors. Since the BP vectors are constituted by proportions, a large variance presents a lower variability of contexts in this case.<sup>10</sup> So, in this

<sup>10</sup> This might seem counterintuitive, but the explanation is straightforward. All ID tag levels add up to 1 (100%) – a particular meaning is less formally constrained when the percentages are equally distributed over these

respect, within the *sentir*-cluster, general physical perception is less formally constrained (var = 0.119) compared to emotional experience (var = 0.128) and emotional regret (var = 0.165). From this perspective, the second column of table 3 illustrates that general physical perception is less marked and thus the most prototypical sense (cf. Lakoff 1987: 60-61 about the relationship between markedness and prototypicality). Emotional regret is the most formally constrained of these three senses.<sup>11</sup> Indeed, a quick glance at the possible forms of the Direct Object (DO) in the general physical perception sense reveals that this use admits not only nominal DOs (cf. example 11 above), but also *that* complementation (*que*, example 20), clitics (*las*, example 21) and other types of (adverbial) subordinate clauses, such as the one introduced by *como* in example (22):

- (20) Aquel día, el céltico saltó al campo para disputar un encuentro amistoso contra el Betis, con motivo del Memorial Quincho, y al poco **sintió** que la rodilla izquierda se le quebraba. [Prensa, 2001]  
 ‘That day, the Celtic player went into the field to play a friendly match against Betis, on the occasion of the Quincho Memorial, and soon he *felt* that his knee breaking’.
- (21) Curioso, que las dos costillas rotas no le duelan, el cirujano de campaña hizo un buen trabajo, es probable que no las **sienta** porque están en la mitad de su cuerpo [...]. [Torres, 2004]  
 ‘It is strange that the two broken ribs do not hurt you, the field surgeon did a good job, it is probable that you do not *feel* them because they are in the middle of your body [...]
- (22) [...] ha aumentado el ritmo de sus pasos, seguramente, tiene la respiración fatigosa y **siente** cómo le palpita el corazón [Salvador Caja, 2002]  
 ‘[...] the pace of his steps has increased, surely, he has a labored breathing and he *feels* how his heart beats.’

The emotional regret meaning, on the other hand, turns out to be the most formally constrained one. Indeed, besides some uses with the infinitive (cf. example 18 above), in this sense the verb predominantly occurs with the clitic *lo* (23):

- (23) Lo **siento**, no podemos confirmar ni negar. [Giménez Bartlett, 2002]  
 ‘I am sorry, we cannot confirm nor deny.’

In sum, from the perspective of a high family resemblance and the subsequent formally least constrained sense, the image obtained is rather different because of the three most frequent senses in the *sentir*-cluster, the one that displays the most formal variation is the less frequent one, namely general physical perception.

### 4.3. Degree of sense distinctiveness

---

100%, so that the variance (i.e. the deviation from the mean) is smaller. An ID tag with 3 levels, for instance, could yield 33%-33%-33% for a particular meaning, but 100%-0%-0% for another meaning. Clearly, the first one is more equilibrated in terms of the contexts in which it appears with that ID tag category, while the second has a larger variance and is strongly constrained and only occurs with one of the three levels.

<sup>11</sup> In fact, only this sense appears to have a statistically significant different variance from the other two senses, according to a Fligner-Killeen test for homogeneity of variances for non-normally distributed data: EMO.lamentar vs. EMO.experimentar:  $X^2 = 83.3$ ,  $df = 58$ ,  $p = 0.016$ ; EMO.lamentar vs. FIS.GEN.experimentar.corp  $X^2 = 33.8$ ,  $df = 24$ ,  $p = 0.089$  (marginally significant).



A second problematic issue of polysemy research is associated with determining whether different occurrences exemplify divergent senses or merely alterations of a more general sense (the lumping vs. splitting issue). Alternatively, which senses are more similar to each other and how can they be delineated from the other senses?

As we saw in Section 2.2.1, a quick glance at several dictionary entries reveals that they diverge significantly regarding the internal organization of the entry and the number of definitions proposed for the verb. Similarly, following the high degree of granularity of some (early) cognitive linguistics analyses (cf. for example the *full-specification approach* referred to in Section 2.1.1), we could simply interpret the obtained dendrogram as instantiating 25 different senses of the verb (or 30 if we take into account the ambiguous cases). However, as already mentioned in Section 2.1.1, this type of full-specification approach is not without problems and therefore, other methods of analyses have been proposed to complement those intuition-based ones.

A corpus-based perspective, however, distinguishes senses in terms of formal patterns, and, as such, facilitates a construction-based approach to the delineation problem. The underlying assumption when comparing distributional patterns is that the greater the overlap is in syntactic structures, the closer certain senses must be in their semantic structure. Indeed, as Divjak (2010a: 17-18) notes, more recent cognitive linguistics generally acknowledges that the constructions of a particular lexeme can correlate with its semantic characteristics, and consequently, both lexemes and constructions convey meaning. Moreover, it has to be stressed that, in accordance with their reject of discrete senses (cf. Section 2.1.1), cognitive linguists have also posited a continuum of semantic distinctness where the cases of polysemy are located between the extremes of vagueness on the one hand, and homonymy on the other hand (cf. Tuggy [1993] or Croft [1998]). In other words, the distinctness of different senses is considered a matter of degree. Below we examine how these two cognitive semantic principles – a construction-based approach to sense distinctness and a continuum of semantic distinctness – can be concretely applied to the motivation of sense distinction of *sentir*.

At first glance, the overall picture resulting from the cluster analysis looks fairly cluttered. However, just as human beings are constantly categorizing and organizing the apparent complexity of nature on the basis of similarity and dissimilarity between entities, the same reasoning can be applied to *sentir*. In fact, within its continuum of sense distinctness, some clear focal points or clusters can be distinguished (Section 4.1): a first distinction is between de middle voice uses (*sentirse*) en the active uses (*sentir*). Within this second cluster, three meaningful groupings arise, namely (1) emotional or general physical experiencing (2) cognitive perception – clustered curiously enough with the regret cases, and (3) ability of perception. Moreover, it is striking that both between these clusters and within each cluster separately, ambiguous cases arise, reflecting and visualizing those fuzzy boundaries and transitions between the focal categories. These turn out to be very fertile areas for the generation of metaphorical uses of the verb. For example, within the first cluster between the focal points of EMO.encontrarse and FIS\_GEN.encontrarse (resp. ‘being in an emotional or general physical state’), we find both metaphorical cases of the general physical state (24) and ambiguous instances between physical and emotional perception (25):

- (24) El horror de aquella escena había desfilado ante mis ojos en apenas unos segundos. Me **sentía** paralizado, incapaz de actuar o de articular un solo pensamiento. [Ruiz Zafón, 2001]  
 ‘The horror of that scene had passed before my eyes in just a few seconds. I *felt* paralyzed, unable to act or to articulate one single thought.’
- (25) Por la ventana comenzaba a filtrarse suavemente la luz del amanecer. Algunas velas se habían apagado... Yo me **sentía** bien en mi cuerpo, como se siente un hombre sano después de hacer el amor. [Llongueras, 2001]  
 ‘The light of dawn began to filter gently through the window. Some candles were extinguished... I *felt* good in my body, how a healthy man feels after making love.’

The presence of this kind of examples is intuitively totally reasonable, because it is generally known that physical perception metaphorically relates to emotions (cf. among other Kurath 1921, Sweetser 1990, Ibarretxe-Antuñano 1999, Kövecses 2008). As example (24) fittingly illustrates, the physical state of being paralyzed, that is the physical inability to move, metaphorically refers to the emotional meaning of being afraid (cf. for example, expressions such as ‘paralyzed with fright’). Sentence (25), on the other hand, is ambiguous between physical (for example referring to the good physical health of the person) and emotional state (referring to his mental health, feeling good). The same type of metaphorical use between cases of general physical and emotional experiencing also manifests in the second cluster, and facilitates the transition between both senses and thus concretizes the semantic continuum of *sentir*.

This semantic continuum is also reflected in the syntactic distribution of the verb, which in turn helps to interpret and disentangle some, at first sight, ‘strange’ clusterings on the basis of shared formal patterns. Two particular clusterings need to be inspected in more detail, namely (1) the presence of the cognitive cases equivalent to ‘considerar’ (‘consider’, ‘judge’) in the cluster with emotional and physical perception and, (2) the clustering of the emotional regret meanings with cognitive perception.

First, if there is a separate grouping of cognitive perception meanings, why does it not include the sense ‘consider’? Indeed, the cognitive cluster contains very close senses of ‘presentir’ (‘have a presentiment’), ‘intuir’ (‘intuit’; cf. example 26), ‘pensar’ (‘think’, ‘believe’: cf. example 13, *supra*), ‘opinar’ (‘opine’), ‘darse cuenta de’ (‘realize’; cf. example 27):

- (26) El día que viniste a casa por primera vez **sentí** que ya te conocía. [Ruiz Zafón, 2001]  
 ‘The day you came home for the first time I *felt* I knew you’
- (27) [...] lo confortable de aquellas reuniones, lo burgués, la sensación de seguridad y de reposo, invadían el corazón hasta anegarlo. Isabel de la Hoz **sintió** muy pronto que era obligatorio rebelarse contra aquel confortable reducto interior, como contra una tentación perversa: la tentación de no salir. [Pombo, 2004]  
 [...] the comfort of those meetings, its bourgeois character, the feeling of safety and rest, invaded the heart and even inundated it. Isabel de la Hoz very soon *felt* that it was obligatory to rebel against that comfortable interior stronghold, as against a perverse temptation: the temptation to not leave.’

Notice that in these epistemic examples, the verb does not express a cast-iron certainty but rather a kind of knowledge that does not require much “epistemic commitment” from the part of the speaker (Fernández Jaén 2012: 431). So in these contexts the verb introduces suspicions (26) or epistemic uses expressing a kind of involuntary finding based on experience (27).

Therefore, the senses within this cluster can be lumped together under the common denominator of “attenuated or mitigated epistemicity”, characterized by a lower control over the mental activity when compared to more prototypical and agentive cognitive verbs such as ‘saber’ (‘know’).

Moreover, what favors this lumping argument is their shared formal patterns. In this respect, it is worth noting that the senses in this cluster share the presence of a particular verbal complement type, namely a complement referring to a situation (event or state), formalized in the large majority of the cases (62/79 examples, or 78.5%) by means of a *that* complementation pattern (cf. ex. 26, 27), and – to much smaller amount – by means of the infinitive, neuter clitics (*lo*) and deverbal NPs all referring to events.

On a similar basis, these formal patterns justify the separation of the ‘considerar’ type, which turns out to be restricted to a very specific construction, namely the predicative complement oriented towards the object:

- (28) Me complace que me **sienta** ya tan amiga como para confiarme esas cosas. [Salvador Caja, 2002]  
‘I am pleased that she already *considers* me that much as a friend to entrust me this kind of things’

From a semantic point of view, it comes as no surprise that these cases cluster with the emotional or physical experiencing, because they are primarily based on a general experience to which, in a second stage, a certain valorization or evaluation from the part of the speaker is associated. In other terms, these examples situated at the margins between the cognitive and the emotional/physical perception instantiate an epistemic shift towards subjectivity.

Second, a focus on the distributional similarity also explains why the emotional regret sense clusters with the cognitive senses. As has just been explained, this cluster is characterized by the presence of verbal complements that refer to a situation. This is precisely what also distinguishes the regret sense of *sentir* from the other emotional senses: whereas the latter tends to occur with an NP (cf. example 12 above), the regret sense is restricted to verbal complements, and principally, to the construction with the neuter clitic pronoun ‘lo’ (29):

- (29) Son instrucciones del comandante. Debe usted abandonar el avión. - Aquí debe de haber un error. - En absoluto, señor. **Lo siento** mucho pero tengo que insistirle. [Silva, 2000]  
‘These are instructions from the captain. You should leave the plane. - There must be some misunderstanding here. - Not at all, sir. I am really sorry but I have to insist’.

To summarize, a construction-based method helps to distinguish senses in terms of formal patterns, which refines introspective analyses and explains at first sight counter-intuitive clustering results by formulating the (dis)similarities in a more precise way.

#### 4.4. Structure of the network

A related question to the problem of sense distinctiveness is how exactly this (dis)similarity of senses can be measured and quantified in a more precise way. Specifically, certain senses have been found to be more closely related than others, but how close is ‘closely’? The present section deals with this question and aims to determine the precise structure of the

semantic network of senses and their relations: which of them are more (dis)similar and how can this structure be motivated on the basis of objective evidence?

The answer can be found through a correlation analysis. More precisely, 435 pairwise correlations of the 30 senses were calculated, yielding correlation coefficients that range from 0.27 to 0.99. If we focus on the highest values, the most similar senses (0.99) turn out to be ‘EMO.encontrarse’ (‘be in an emotional state’) and ‘FIS.GEN.encontrarse.metaf’ (metaphorical uses of the general physical state). Indeed, as mentioned in Section 4.3 (cf. examples 24, 25), metaphorical physical uses essentially refer to emotions. This outcome of the correlation analyses is thus intuitively reasonable. Similarly, between the most similar senses are also all those referring to cognitive perception: ‘presentir’ (‘have a presentiment’), ‘intuir’ (‘intuit’), ‘opinar’ (‘opine’), ‘pensar’ (‘think’, ‘believe’), ‘darse cuenta’ (‘realize’), with correlation coefficients ranging between 0.83 and 0.98. This strong correlation measure could serve as an extra argument in favor of the lumping together of these senses instead of splitting them (cf. Section 4.3).

Next, concerning the least similar senses, ‘EMO.encontrarse’ (‘be in an emotional state’) differs most from all other senses except (obviously) from the ‘FIS\_GEN.encontrarse.metaf’ (metaphorical uses of the general physical state), the ‘AMBIG.FIS.EMO’ (ambiguous uses of the emotional perception), and the ‘FIS\_GEN.encontrarse’ (‘be in a physical state’) sense, which coincide perfectly with the first cluster equivalent to the middle voice uses of *sentirse*. The correlation measures thus endorse the separation of these middle voice uses from the other uses of *sentir* and consequently the existence of two large clusters of the verb.

They also affirm the grouping of the regret senses (EMO.lamentar) with the cognitive cluster (COGN.X) (correlation coefficients ranging from 0.70 to 0.84, cf. also the AU value of 95%) rather than with the emotional/physical experience cluster (resp. 0.66 and 0.70), although the difference does not turn out to be that high. These values probably point at the situation of this specific sense at the borderline between cognitive and emotional perception.

However, the most distant sense from all the others turns out to be ‘FIS\_GEN\_Manifestarse’ (‘manifest, reveal itself’):

(30) La influencia de la arquitectura se deja **sentir** en todas manifestaciones artísticas como prueba este retablo-relicario de madera pintada que conjuga las formas ligeras y caladas góticas con mozárabes y tracerías mudéjares. [Beltrán Martínez, 2000]

‘The influence of the architecture is *felt* in all art forms as testifies this painted wooden altarpiece-reliquary combining the light and soaked gothic forms with Mozarabic and Mudejar tracery.’

This sense of the verb is restricted to the specific construction with causative verbs in the pronominal passive form (in this example ‘dejarse + infinitive’). As visualized in the dendrogram, it clusters with the ‘capacidad’ (‘ability’) senses of the verb, and the AU value of 97% suggests that there is good evidence for this. This can easily be interpreted semantically, given that the causative construction ‘hacerse / dejarse + infinitive’ expresses a kind of creation of a possibility for the experiencer to ‘feel’ something. The extreme negative values obtained in the correlation analysis could thus indicate a privileged construction linked to a particular sense of the verb. Put another way, accepting that both lexemes and constructions have meanings, we can assume that both meanings interact when they are put into contact (cf.

Divjak 2006: 20). The next section investigates the underlying motivation for the verb to enter in this specific construction.

#### 4.5. Morphosyntactic correlates

A final problem involved in the study of polysemy, and linked to all the previous ones, relates to the question of which are the specific morphosyntactic correlates of the semantic network obtained for the verb. In other words, what is the underlying morphosyntactic motivation for the clustering and the specific properties that discriminate between clusters?

To this end, Backhaus et al. (1996: 310-312) proposed computing *t*-values: these *t*-values help to detect which are the most strongly represented (in the case of positive *t*-values) and underrepresented (negative *t*-values) variables in a particular cluster. Therefore, in order to highlight the most discriminating variables responsible for the clustering, we focused on the variables displaying positive *t*-values for one cluster and negative values for the other clusters, and the other way around. Moreover, since the *t*-values mark characteristics that appear more frequently than average with a particular cluster rather than exclusive features related to a specific cluster, our focus of attention is not on the absolute, but rather on the relative values based on the ranking of these *t*-values (cf. Divjak 2010a, 141-142). In what follows, we will discuss the most revealing *t*-values for each cluster separately.

##### 1) Cluster 1: ‘encontrarse’ (*sentirse*) – be in a physical or emotional state

The most revealing *t*-value within the cluster of middle voice uses *sentirse* (the first red cluster) is – not surprisingly – the presence of a predicative complement of the subject ( $t = 2.437$ ). This subject complement can adopt a wide range of forms, the adjective resulting the most important one ( $t = 2.452$ ; cf. example 31), but also cases introduced by means of *como* ( $t = 2.195$ ; cf. example 32), an adverb ( $t = 1.749$ ; cf. example 33), NP ( $t = 2.002$ ; example 34) or PP ( $t = 1.521$ ; cf. example 35) are found:

- (31) El viajero se **siente** libre en Australia, sacó en conclusión D. H. Lawrence. [Leguineche, 2000]  
‘The traveler *feels* free in Australia, concluded D.H. Lawrence.’
- (32) Cuando eres nuevo, como no tienes ni sitio ni nada, te **sientes** como un mueble, pero de los que estorban. [AAVV, 2001]  
‘When you are new, since you don’t have a place or anything, you *feel* like a piece of furniture, but of those that disturb.’
- (33) Me hablaban con total naturalidad y yo me **sentía** muy bien con ellas. [Llongueras, 2001]  
‘They spoke with complete spontaneity to me and I *felt* very good with them.’
- (34) Afortunadamente, en ningún momento logró que me **sintiera** un apestado o un leproso, simplemente estaba muy sorprendido, casi divertido. [Llongueras, 2001]  
‘Fortunately, he never succeeded to make me *feel* like an outcast or a leprous, I was just very surprised, almost amused.’
- (35) Después de vivir una desgracia tras otra, desde que se había casado con Domingo se **sentía** a salvo. [Gavilanes, 2000]  
‘After having passed one misfortune after another, since she had married Domingo she *felt* safe.’

A more striking observation concerns the scores obtained for the presence of adjuncts, which more specifically, refer to a cause ( $t = 1.484$ ; example 36) or an agent ( $t = 1.187$ ; example 37):

- (36) Han pasado ya siete años desde el descubrimiento inicial y ahora es el momento de divulgar y ofrecer este trabajo al mundo. Por ello me **siento** inmensamente feliz y agradecido [...]. [Iborra Montells, 2001]  
 'It has already been seven years ago since the initial discovery and now is the time to disclose and offer this work to the world. For that reason, I *feel* extremely happy and grateful.'  
 (37) Me he **sentido** muy maltratada por usted. [Salvador Caja, 2002]  
 'I *felt* very abused by you'.

As illustrated in example (37), *sentirse* largely occurs in combination with the participle form used as an adjective, allowing accordingly a more passive reading in comparison with a more standard adjective (as in example 31). Examples where both structures occur in conjunction endorse this passive interpretation:

- (38) No se **sienten** ni aceptan ser representados por las organizaciones políticas actuales y tampoco tienen confianza en la lucha y en la fuerza de los trabajadores. [Prensa, Pueblos, 2000]  
 'They do not *feel* nor accept being represented by the actual political organizations and neither do they have confidence in the struggle and strength of the workers.'

## 2) Cluster 2: 'experimentar una sensación psíquica o física' – experience a mental or physical sensation

Within the active cluster (the second red cluster), a first grouping comprises the cases of mental or physical experiencing. In this cluster, the most revealing  $t$ -values are related to the combination of a DO in the form of an NP ( $t = 0.859$ ), both referring to concrete ( $t = 0.581$ ; example 39) and abstract entities ( $t = 0.553$ ; example 40):

- (39) Piensa en las lagartijas que ahora duermen bajo las piedras calientes y a salvo de navajazos, [...], y **siente** el frío hocico de Chispa, que prolonga su existencia pegado a sus tobillos lastimados, [...]. [Marsé, 2000]  
 'He thinks of lizards that are now sleeping under the hot stones and safe from stabs, [...] and he *feels* the cold muzzle of Chispa, which prolongs his existence stuck on his injured ankles [...]'  
 (40) Isabel observa de reojo a Zamacois y **siente** una ligera repugnancia que en parte es ternura: como quizás se siente al mirar niños muy pequeños [Pombo, 2004]  
 'Isabel looks at Zamacois out of the corner of her eye and feels a slight repugnance which is partly tenderness: perhaps like one *feels* when watching very little children.'

As these examples illustrate, a concrete NP (*el frío hocico de Chispa*) is related to the expression of a physical perception (in this case referring to touch), whereas the combination with an abstract NP (*una ligera repugnancia, ternura*) triggers an emotional reading.

Within the combinations with an NP, cases of presence ( $t = 0.791$ ) or absence ( $t = 0.424$ ) of a determiner should also be distinguished. The latter cases turn out to be intrinsically related to the emotional experience sense with an abstract NP:

- (41) Tenía suerte de que me hubieran parido así, sin **sentir** miedo ni desesperación. [Llongueras, 2001]

‘I was lucky that I had been given birth that way, without *feeling* fear or desperation.’

Other examples of the same type are *sentir admiración*, *sentir vergüenza*, *sentir alegría* etc. In these cases, *sentir* behaves like a light verb within the complex predicate V + DO.

Finally, the *t*-values also highlight the explicit presence of an adverbial adjunct specifying place within this cluster:

(42) No me imagino todas las consecuencias de opción tan arriesgada, pero estoy convencido de que al menos los espectadores hubieran **sentido** miedo en el cuerpo, tal como yo quería. [Boadella, 2001]

‘I cannot imagine all the consequences of such a risky option, but I am convinced that at least the spectators have *felt* fear in their body, just like I wanted.’

(43) Siguió caminando y se quitó los zapatos cuando **sintió** bajo sus pies la arena blanda. [Pérez-Reverte, 2002]

‘She continued walking and took off her shoes when she *felt* the soft sand under her feet.’

As these examples illustrate, both the emotional and physical perception meaning affirm the explicit specification where (mostly a body part) the experience takes place: on the inner side of the body in the case of emotional experience (cf. example 42 *en el cuerpo*) or in the outer side in the case of physical perception (cf. example 43 *bajo sus pies*). This reference to the human body thus seems to underlie the clustering of these emotional and physical feelings. However, once again, the corpus shows some borderline metaphorical in-between examples, which combine a clearly physical description (cf. example 44 words such as *pinchazo*, *aire*, *pulmones*) with references to emotion (cf. words as *en el centro del pecho*, *conmoción*):

(44) y él **sentía** un pinchazo agudo y delicioso en el centro del pecho, mientras el aire abandonaba a toda prisa sus pulmones para dejar que se ahogara en su propia conmoción. [Grandes, 2002]

‘and he *felt* a sharp and delightful prick in the center of his chest, while the air quickly abandoned his lungs in order to let him drown in his own shock.’

### 3) Cluster 3: Cognitive perception + emotional regret

This cluster consists of the (counter-intuitive) joining of cognitive perception with the particular sense of emotional regret. However, as suggested in Sections 4.3-4.4 this junction can be understood on the basis of their distributional similarity. Indeed, as the *t*-values confirm, the most important variable within this cluster turns out to be precisely the presence of a DO referring to an event or situation by means of *that* complementation (*t*-value = 1.76). It is exactly this complement type that also distinguishes the ‘regret’ sense of *sentir* from the other emotional senses. Whereas the latter tend to occur with a NP (cf. examples 12, 40-42 above), the former, i.e. the ‘regret’ sense is restricted to verbal complements, and principally, to the construction with the neuter clitic pronoun ‘lo’ (‘it’), which is the clitic par excellence to refer to situational complements (cf. examples 23, 29).

With respect to this specific construction, a closer examination of the *t*-values alerts to the important collocation with *pero* (‘but’; *t*-value = 0.73) and the vocative or explicit address to someone (*t*-value = 0.73):

- (45) Son instrucciones del comandante. Debe usted abandonar el avión. - Aquí debe de haber un error. - En absoluto, señor. Lo **siento** mucho pero tengo que insistirle. [Silva, 2000]  
 'These are instructions from the captain. You should leave the plane. - There must be some misunderstanding here. - Not at all, sir. I am really sorry but I have to insist'

In this example (see also example 29, repeated here for convenience), the presence of the vocative (*señor*) and the adversative conjunction *pero*, indicate that rather than implying 'regret' in the strict sense, *lo siento* is an autonomous predication functioning as an intersubjective pragmatic marker which announces that the speaker will say something that may be contrary to the opinion or expectation of his interlocutor. This (implicit) announcement of a differing opinion, attenuating an assertion, reveals how, even in its pragmatic use, an underlying epistemic sense is maintained. However, concurrently, another striking element is the frequent collocation of this construction with adverbs of quantity (*mucho*). In their capacity of 'gradable quantifiers', these adverbs occur preferably with verbs of emotion, which allow more easily being quantified than, for example, cognitive verbs as *pensar* ('think'; \**lo pienso mucho?*). In other words, the derivative use as an intersubjective pragmatic marker continues testifying the borderline position of the 'regret' cases between the epistemic uses on the one hand and the emotional sense on the other.

#### 4) Cluster 4: Ability to perceive + 'manifestarse'

Within the fourth cluster, the most important variable turns out to be the occurrence of *sentir* in the infinitive form ( $t = 1.148$ ), both in the absolute construction ( $t = 1.126$ ; example 46) and in pronominal passive causative constructions yielding the 'manifestarse' sense equivalent to 'appear, show up' ( $t = 0.781$ ; cf. example 47):

- (46) Y Sofía quiere meterse algo al cuerpo y no **sentir**. [Beccaria, 2001]  
 'And Sophia wants to put something into her body and not *feel*.'  
 (47) La profunda afinidad electiva que existió entre liberalismo y ciencia social se dejó **sentir**, desde el primer momento, en el desarrollo de la ciencia económica y, en parte, en el de la teoría política. [Giner, 2001]  
 'The high elective affinity that existed between liberalism and social science was *felt* from the first moment, in the development of economics, and partly, in that of political theory.'

As we already mentioned in Section 4.4, the occurrence of the verb in this construction needs some further comments. Accepting the idea that both lexemes and constructions have meanings, we can assume that the presence of the verb in this construction convey a particular meaning. Therefore, the question arises as to what attracts *sentir* to this construction, that is to say, which element of this construction enables *sentir* to do something that it cannot do on its own? Or in short, what is the added value of their alliance?

In accordance with its semantic profile, *sentir* is a verb of which the grammatical subject coincides with the experiencer. This is the case in all its senses and different clusters: whether it refers to an emotional or physical state, an emotional or physical experiencing, a cognitive perception or ability, the subject of *sentir* is always a human experiencer. In the causative construction, however, the semantic role of the grammatical subject is the stimulus of the perception and the experiencer very often is not even mentioned. Correspondingly, this



facilitates the occurrence of inanimate, abstract subjects such as *los efectos* (the effects), *la influencia* (the influence) etc. This kind of non-prototypical subjects contrasts sharply with the usual grammatical subjects of the verb where ‘NP<sub>1</sub> feels NP<sub>2</sub>’ – this NP<sub>1</sub> coinciding generally with the concrete, human, dynamic, topical subject of the verb. In other words, the specific construction with a causative verb in the reflexive passive form allows focusing on the stimulus of perception in lieu of the experiencer, which fades into the background. As we already mentioned, on the semantic level, the clustering with the ‘ability’ senses can be explained by the fact that the causative construction ‘hacerse / dejarse + infinitive’ precisely expresses a kind of creating or enabling the possibility for the experiencer to ‘feel’ something, which fits perfectly well with its quality as a topicalizing construction.

In sum, computing the *t*-values turns out to be an entirely valuable method to gain more insights into the underlying motivation for the clustering in order to discover the discriminating variables responsible for the clustering. Going beyond the more specific level of the individual senses, the calculation of the *t*-values allows an important generalization on the cluster level. In this respect, it turns out that the most discriminating variable relates to the argument structure. It is thus clear that a construction-based approach is a thoroughly rewarding method to distinguish senses in terms of formal patterns, refining considerably the introspective analyses or offering clarification of the seemingly counter-intuitive clustering in a more effective way.

## 5. Conclusion

Through a corpus-based behavioral profile analysis, this study has examined the polysemy of the Spanish verb *sentir*, which has led to a number of significant insights gained at different levels. First, from a methodological point of view, it has been demonstrated that the BP analysis is a highly fruitful method for the analysis of polysemy, thus validating some important previous research findings based on the same approach, albeit associated with other phenomena in lexical semantics such as near-synonymy (Divjak 2010a; Divjak & Gries 2006) and antonymy (Gries & Otani 2010) or with the polysemy of verbs in other languages (Berez & Gries 2009; Gries 2006). Despite the challenge of the operationalization of the elusive concept of meaning, the quantitative nature of the analysis offers a systematic and verifiable alternative to more intuitive approaches to lexical semantics. The method not only offers usage-based evidence for cognitive linguistic theorizing concerning (1) prototypicality of senses, (2) sense-distinction, (3) network representations, and (4) the semantics-morphosyntax interface, but also leads to a gradual refinement of the results of the previous cross-linguistic approach to the polysemy of *sentir*, facilitating fundamentally the empirical cycle of proposing hypotheses and testing them.

Besides its methodological merits as a quantitative usage-based approach to polysemy, this method also offers the qualitative evidence of and insights into the intricate semantics of *sentir*. Specifically, according to frequency, it turns out that the most prototypical sense of the verb refers to the emotional perception – and related to that, the specific meaning of ‘regret’, a particularity of the Spanish verb only. However, from the perspective of a high family

resemblance and the subsequent formally least constrained sense, general physical perception is the most prototypical sense.

Second, starting from the distributional hypothesis, the BP facilitates a construction-based approach to the delineation problem distinguishing senses in terms of formal patterns. Within its continuum of sense distinctness, some clear focal points or clusters can be distinguished: a first distinction is between de middle voice uses (*sentirse*) en the active uses (*sentir*). Within this second cluster, three meaningful groupings arise, namely (1) emotional or general physical experiencing (2) cognitive perception – curiously enough clustered with the regret cases, and (3) ability of perception. According to the traditional definition of a polysemous verb, the existence of some focal clusters in the dendrogram thus indicates that *sentir* indeed has more than one distinct sense. However, both the inclusion of ambiguous and metaphorical cases between and within those clusters show that the senses are not discrete but intrinsically related and connected at various levels of analysis, emphasizing profoundly the continuous nature of the semantic universe of *sentir*. This continuum is also reflected in the syntactic distribution of the verb, which helps to interpret and disentangle some seemingly counter-intuitive clusterings on the basis of shared formal patterns. For example, sharing a *that* complementation pattern, the senses within the cognitive perception cluster can be lumped together under the common denominator of “attenuated or mitigated epistemicity”. This presence of verbal complements linked to a situation characteristic of the cognitive perception cluster also explains why the emotional ‘regret’ sense clusters with these cognitive senses, since this is precisely what distinguishes the ‘regret’ sense of *sentir* from its other emotional senses. While the latter tends to occur with an NP, the ‘regret’ sense is restricted to verbal complements, and principally, to the construction with the neuter clitic pronoun ‘lo’.

Third, a related question to this problem of sense distinctiveness is how exactly this (dis)similarity of senses can be measured and quantified in a more precise way. Explanation was sought through a correlation analysis. The calculation of these measures affirm the grouping of the ‘regret’ senses with the cognitive cluster rather than with the emotional one, although it does not yield a great difference. These values probably point at the situation of this specific sense at the borderline between cognitive and emotional perception. However, the most distant sense from all the others, based on the correlation coefficients, turns out to be ‘appear, show up’, restricted to the specific construction with causative verbs in the pronominal passive form. The extreme negative values obtained in the correlation analysis could thus suggest a privileged construction linked to a particular sense of the verb.

Finally, in order to investigate the underlying motivation for the verb to enter in this kind of specific constructions, the morphosyntactic correlates have been studied. Going beyond the more specific level of the individual senses, the calculation of the *t*-values allows an important generalization on the cluster level. In this respect, it turns out that the most discriminating variable relates to the argument structure: the first cluster (*sentirse*) correlates with the presence of a predicative complement oriented at the subject, the emotional or physical perception is related to the combination with a NP, with or without a determiner. The cognitive and ‘regret’ sense, by contrast, prefers a situational complement and finally, the ability cluster relates to the absolute use. These values concurrently highlight the attraction of the verb to the specific pronominal passive causative constructions yielding the meaning of ‘appear, show up’.

Through its application to a new family of languages (Romance languages) and to a verb that is part of the semantic field of perception, which has not yet been analyzed using BP, the present study has shown how the method can be exploited to its full potential, suggesting some theoretical and methodological implications for future cognitive semantic research in the field of perception and beyond. Clearly, a vast array of rich bibliography has been dedicated to the phenomenon of perception in (Cognitive) Linguistics in general, and to verbs of perception in particular, as the most tangible evidence of the direct interaction between physical perception and linguistic codification. Yet, up to the present, no common empirical ground seems to have been established for the semantic study of perception verbs and the concepts they give access to. It is evident that a comprehensive and principled study of perception verbs is crucial to access the complex conceptual structures behind the linguistic forms. In light of this, the BP approach provides essential clues to the representation of conceptual structures associated with the semantic field of perception. Moreover, extending this methodology as a common empirical method could yield interesting insights on both the cross-linguistic and the intra-linguistic level, and on both the general level of perception modalities and the specific verb level (cf. also Divjak 2015 for a first important step towards that goal, applying the BP approach to the sensory concept formation on the basis of perception types in Russian).

Finally, in order to corroborate the findings of this BP analysis about the usage patterns of the polysemous verb and the concepts of perception, this corpus research can be complemented with other types of data, such as experimental (elicited) data, which has been shown to be fruitful, especially when used in combination with corpus data (cf. recent studies of Arrpe and Järvikivi 2007a-b; Divjak and Gries 2008, Featherston 2005; Gries 2002; Gries et al. 2005a-b; Liu 2013; Vanderschueren & Diependaele 2013).

## References

- Aijmer, Karin & Anne-Marie Simon-Vandenberg. 2011. Pragmatic markers. In Jan Zienkowski, Jan-Ola Östman & Jef Verschueren (eds.), *Discursive Pragmatics*, 223-247.
- Arppe, Antti & Juhani Järvikivi. 2007a. Take empiricism seriously! In support of methodological diversity in linguistics. *Corpus Linguistics and Linguistic Theory* 3(1). 99-109.
- Arppe, Antti & Juhani Järvikivi. 2007b. Every method counts: Combining corpus-based and experimental evidence in the study of synonymy. *Corpus Linguistics and Linguistic Theory* 3(2). 131-159.
- Arppe, Antti, Gaëtanelle Gilquin, Dylan Glynn, Martin Hilpert, & Arne Zeschel. 2010. Cognitive Corpus Linguistics: five points of debate on current theory and methodology. *Corpora* 5(1). 1-27.
- Atkins, Beryl T. 1987. Semantic ID tags: Corpus Evidence for Dictionary Senses. *Proceedings of the Third Annual Conference of the UW Centre for the New Oxford English Dictionary*, 17-36.
- Baayen, R. Harald. 2008. *Analyzing Linguistic Data. A practical introduction to statistics*. Cambridge: Cambridge University Press.

- Backhaus, Klaus, Bernd Erichson, Wulff Plinke & Rolf Weiber. 1996. *Multivariate Analysemethoden: eine anwendungsorientierte Einführung*. Berlin, Heidelberg, New York: Springer.
- Berez, Andrea L. & Stefan Th. Gries. 2009. In defense of corpus-based methods: a behavioral profile analysis of polysemous *get* in English. In Steven Moran, Darren S. Tanner, & Michael Scanlon (eds.), *Proceedings of the 24th Northwest Linguistics Conference*. University of Washington Working Papers in Linguistics Vol. 27, 157-166. Seattle, WA: Department of Linguistics.
- Brugman, Claudia. 1988. *The Story of Over: Polysemy, Semantics and the Structure of the Lexicon*. New York: Garland.
- Brugman, Claudia & George Lakoff. 1988. Cognitive topology and lexical networks. In: Steven L. Small, Garrison W. Cottrell & Michael K. Tanenhaus (eds.), *Lexical Ambiguity Resolution*, 477-508. San Mateo, CA: Morgan Kaufman.
- Croft, William. 1998. Linguistic evidence and mental representations. *Cognitive Linguistics* 9(2). 151-173.
- Cuyckens, Hubert & Britta Zawada. 1997. *Polysemy in Cognitive Linguistics. Selected papers from the Fifth International Cognitive Linguistic Conference*. Amsterdam & Philadelphia: John Benjamins.
- De Mauro, Tullio. 1999. *Grande dizionario Italiano dell'uso*. Torino: Utet.
- Divjak, Dagmar. 2003. On trying in Russian: a tentative network model for near(er) synonyms. In *Slavica Gandensia 30. Belgian Contributions to the 13th International Congress of Slavists*, Ljubljana, 15-21 August 2003, 25-58.
- Divjak, Dagmar. 2006. Ways of Intending: Delineating and Structuring Near-Synonyms. In: Stefan Th. Gries, St. & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics. Corpus-based Approaches to Syntax and Lexis*, 19-56. Berlin & New York: Mouton de Gruyter.
- Divjak, Dagmar. 2010a. *Structuring the lexicon. A clustered model for near-synonymy*. Berlin & New York: Mouton de Gruyter.
- Divjak, Dagmar. 2010b. Corpus-based evidence for an idiosyncratic aspect-modality relation in Russian. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 305-330. Berlin & New York: Mouton de Gruyter.
- Divjak, Dagmar. 2015. Exploring the grammar of perception. A case study using data from Russian. *Functions of Language* 22.
- Divjak, Dagmar & Stefan Th. Gries. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1). 23-60.
- Divjak, Dagmar & Stefan Th. Gries. 2008. Clusters in the mind? Converging evidence from near synonymy in Russian. *The Mental Lexicon* 3(2). 188-213.
- Enghels, Renata. 2007. *Les modalités de perception visuelle et auditive. Différences conceptuelles et répercussions sémantico-syntaxiques en espagnol et en français*. Tübingen: Niemeyer.
- Enghels, Renata & Marlies Jansegers. 2013. On the cross-linguistic equivalence of *sentir(e)* in Romance languages: a contrastive study in semantics. *Linguistics* 51(5): 957-991.

- Featherston, Sam. 2005. The Decathlon Model. In Stephan Kepser & Marga Reis (eds.), *Linguistic Evidence. Empirical, Theoretical and Computational Perspectives*, 187-208. Berlin & New York: Mouton de Gruyter.
- Fernández Jaén, Jorge. 2012. *Semántica cognitiva diacrónica de los verbos de percepción física del español*. Alicante: Universidad de Alicante dissertation.
- Fillmore, Charles. 1985. Frames and the Semantics of Understanding. *Quaderni di Semantica* 6. 222-254.
- Geeraerts, Dirk. 1993. Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4(3). 223-272.
- Geeraerts, Dirk. 2006. Methodology in Cognitive Linguistics. In Gitte Kristiansen, Michel Achard, René Dirven & Francisco J. Ruiz de Mendoza Ibáñez (eds.), *Cognitive Linguistics: Current applications and future perspectives*, 21-50. Berlin & New York: Mouton de Gruyter.
- Geeraerts, Dirk. 2010. The doctor and the semantician. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 63-78. Berlin & New York: Mouton de Gruyter.
- Geeraerts, Dirk, Stefan Grondelaers & Peter Bakema. 1994. *The structure of lexical variation. Meaning, naming, and context*. Berlin & New York: Mouton de Gruyter.
- Gibbs, Raymond. 2007. Why cognitive linguists should care more about empirical methods. In Monica González-Márquez, Irene Mittelberg, Seana Coulson & Michael J. Spivey (eds.), *Methods in Cognitive Linguistics*, 2-18. Amsterdam & Philadelphia: John Benjamins.
- Gilquin, Gaëtanelle. 2003. Causative *get* and *have*. So close, So different. *Journal of English Linguistics* 31. 125-148.
- Gilquin, Gaëtanelle. 2006. The place of prototypicality in corpus linguistics: Causation in the hot seat. In Stephan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*, 159-192. Berlin & New York: Mouton de Gruyter.
- Gilquin, Gaëtanelle. 2008. What you think ain't what you get: highly polysemous verbs in mind and language. In Jean-Rémi Lapaire, Guillaume Desagulier & Jean-Baptiste Guignard (eds.), *Du fait grammatical au fait cognitif. From Gram to Mind: Grammar as Cognition*, 235-255. Pessac: Presses Universitaires de Bordeaux.
- Glynn, Dylan. 2009. Polysemy, syntax, and variation. A usage-based method for Cognitive Semantics. In Vyvyan Evans and Stéphanie Pourcel (eds.), *New Directions in Cognitive Linguistics*, 77-106. Amsterdam & Philadelphia: John Benjamins.
- Glynn, Dylan. 2010a. Corpus-driven Cognitive Semantics. Introduction to the field. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 1-41. Berlin & New York: Mouton de Gruyter.
- Glynn, Dylan. 2010b. Synonymy, lexical fields, and grammatical constructions. A study in usage-based cognitive semantics. In Hans-Jörg Schmid & Susanne Handle (eds.), *Cognitive Foundations of Linguistic Usage Patterns*, 89-118. Berlin & New York: Mouton de Gruyter.
- Glynn, Dylan. 2010c. Testing the hypothesis: Objectivity and verification in usage-based Cognitive Semantics. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in*

- cognitive semantics: corpus-driven approaches*, 239-269. Berlin & New York: Mouton de Gruyter.
- Glynn, Dylan. 2014. Polysemy and Synonymy. Cognitive theory and corpus method. In Dylan Glynn & Justyna A. Robinson (eds.), *Corpus Methods for Semantics. Quantitative Studies in polysemy and synonymy*. Amsterdam & Philadelphia: John Benjamins.
- Glynn, Dylan & Kerstin Fischer (eds.). 2010. *Quantitative Methods in Cognitive Semantics: Corpus-Driven Approaches*. Berlin & New York: Mouton de Gruyter.
- Glynn Dylan & Robinson Justyna (eds.). (2014). *Polysemy and synonymy: corpus methods and applications in Cognitive Linguistics*. Amsterdam & Philadelphia: John Benjamins.
- Gries, Stefan Th. 2002. Evidence in linguistics: Three approaches to genitives in English. In Ruth M. Brend, William J. Sullivan, & Arle R. Lommel (eds.), *LACUS Forum XXVIII: what constitutes evidence in linguistics?*, 17-31. Fullerton, CA: LACUS.
- Gries, Stefan Th. 2003. *Multifactorial Analysis in Corpus Linguistics: A study of Particle Placement*. London: Continuum.
- Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: The many senses of *to run*. In Stefan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis*, 57-99. Berlin & New York: Mouton de Gruyter.
- Gries, Stefan Th. 2009. *Quantitative corpus linguistics with R: a practical introduction*. London & New York: Routledge, Taylor & Francis Group.
- Gries, Stefan Th. 2010a. Behavioral Profiles: a fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, 5(3). 323-346.
- Gries, Stefan Th. 2010b. BehavioralProfiles 1.01. A program for R 2.7.1 and higher.
- Gries, Stefan Th. 2015. Polysemy. In Ewa Dąbrowska & Dagmar Divjak (eds.), *Handbook of Cognitive Linguistics*. Berlin & New York: Mouton de Gruyter.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005a. Converging evidence: Bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4), 635-676.
- Gries, Stefan Th., Beate Hampe & Doris Schönefeld. 2005b. Converging evidence II: More on the association of verbs and constructions. In John Newman & Sally Rice (eds.), *Empirical and Experimental Methods in Cognitive/Functional Research*, 39-72. Stanford: CSLI Publications.
- Gries, Stefan Th. & Dagmar Divjak. 2009. Behavioral profiles: a corpus-based approach to cognitive semantic analysis. In Vyvyan Evans & St  phanie Pourcel (eds.), *New Directions in Cognitive Linguistics*, 57-75. Amsterdam & Philadelphia: John Benjamins.
- Gries, Stefan Th. & Dagmar Divjak. 2010. Quantitative approaches in usage-based Cognitive Semantics: Myths, erroneous assumptions, and a proposal. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 333- 353. Berlin & New York: Mouton de Gruyter.
- Gries, Stefan Th. & Naoki Otani. 2010. Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34. 121-150.

- González-Márquez, Monica, Irene Mittelberg, Seana Coulson & Michael J. Spivey (eds.). 2007. *Methods in Cognitive Linguistics*. Amsterdam & Philadelphia: John Benjamins.
- Grondelaers, Stefan & Dirk Geeraerts. 2003. Towards a pragmatic model of cognitive onomasiology. In Hubert Cuyckens, René Dirven & John R. Taylor (eds.), *Cognitive Approaches to Lexical Semantics*, 67-92. Berlin & New York: Mouton de Gruyter.
- Hanegreefs, Hilde. 2008. *Los verbos de percepción visual. Un análisis de corpus en un marco cognitivo*. Leuven: KULeuven dissertation.
- Ibarretxe-Antuñano, Iraide B. 1999. Metaphorical mappings in the sense of smell. *Metaphor in Cognitive Linguistics, Selected Papers from the 5<sup>th</sup> International Cognitive Linguistics Conference*, Amsterdam, July 1997, 29-45. Amsterdam & Philadelphia: John Benjamins.
- Jansegers, Marlies & Renata Enghels. 2013. De verbo de percepción a marcador de disculpa: la evolución diacrónica del verbo *sentir* en español, *Revue de Linguistique Romane* 305-306. 139-166.
- Kemmer, Suzanne. 1993 *The middle voice*. Amsterdam & Philadelphia: John Benjamins.
- Kövecses, Zoltán. 2008. Metaphor and Emotion. In: Raymond W. Gibbs (ed.), *The Cambridge handbook of Metaphor and Thought*, 380-396. Cambridge: Cambridge University Press.
- Kurath, Hans. 1921. *The semantic sources of the words for the emotions in Sanskrit, Greek, Latin, and the Germanic languages, Department of comparative philology, general linguistics, and Indo-Iranian philology*. Chicago: University of Chicago Press.
- Lakoff, George. 1982. Categories. An essay in Cognitive Linguistics. In Linguistic Society of Korea (ed.), *Linguistics in the Morning Calm*, 139-194. Seoul: Hanshin.
- Lakoff, George. 1987. *Women, Fire, and Dangerous Things. What categories reveal about the mind*. Chicago & London: University of Chicago Press.
- Langacker, Ronald. 1987. *Foundations of Cognitive Grammar*. Vol. 1. *Theoretical prerequisites*. Stanford: Stanford University Press.
- Langacker, Ronald. 1988. A usage-based model. In Brygida Rudzka-Ostyn (ed.), *Topics in Cognitive Linguistics*, 127-161. Amsterdam & Philadelphia: John Benjamins.
- Miller, George A. & Philip N. Johnson-Laird. 1976. *Language and perception*. Cambridge: Cambridge University Press.
- Nerlich, Brigitte & David D. Clarke. 1997. Polysemy: Patterns of Meaning and Patterns in History. *Historiographica Linguistica* 24. 349-385.
- Newman, John & Rice, Sally. 2004. Patterns of usage for English SIT, STAND and LIE: A cognitive exploration in corpus linguistics. *Cognitive Linguistics* 15(3): 351-396.
- Newman, John & Rice, Sally. 2006. Transitivity schemas of English EAT and DRINK in the BNC. *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*, St. Th. Gries & Anatol Stefanowitsch (eds.), 225-260. Berlin & New York: Mouton de Gruyter.
- Raukko, Jarno. 1999. An “intersubjective” method for cognitive-semantic research on polysemy: The case of *get*. In Hiraga Masako K., Chris Sinha & Sherman Wilcox (eds.), *Cultural, Psychological and Typological Issues in Cognitive Linguistics. Selected papers of the bi-annual ICLA meeting in Albuquerque, July 1995*, 87-105. Amsterdam & Philadelphia: John Benjamins.

- Raukko, Jarno. 2003. Polysemy as flexible meaning: experiments with English *get* and Finnish *pitää*. In Brigitte Nerlich, Zazie Todd, Vimala Herman, and David D. Clarke (eds.), *Polysemy. Flexible patterns of meaning in mind and language*, 161-193. Berlin & New York: Mouton de Gruyter.
- Robert, Paul, Alain Rey & Josette Rey-Debove. 2008. *Le Nouveau Petit Robert. Dictionnaire alphabétique et analogique de la langue française. Nouvelle édition du Petit Robert de Paul Robert*. Paris: Dictionnaires Le Robert.
- Rosch, Eleanor. 1978. Principles of Categorization. In Eleanor Rosch & Barbara B. Lloyd (eds.), *Cognition and Categorization*, 27-48. Hillsdale, N.J.: Lawrence Erlbaum.
- Rosch, Eleanor & Carolyn B. Mervis. 1975. Family resemblances: studies in the internal structure of categories. *Cognitive Psychology* 7. 573-605.
- Sandra, Dominiek & Sally Rice. 1995. Network analyses of prepositional meaning: Mirroring whose mind – the linguist's or the language user's? *Cognitive Linguistics* 6(1). 89-130.
- Seco, Manuel, Andrés Olimpia & Ramos Gabino. 1999. *Diccionario del español actual*. Madrid: Aguilar.
- Schmid, Hans-Jörg. 2000. *English abstract nouns as conceptual shells: from corpus to cognition*. Berlin & New York: Mouton de Gruyter.
- Schmid, Hans-Jörg. 2010. Does frequency in text instantiate entrenchment in the cognitive system? In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 101-133. Berlin & New York: Mouton de Gruyter.
- Stefanowitsch, Anatol. 2010. Empirical Cognitive Semantics: Some thoughts. In Dylan Glynn & Kerstin Fischer (eds.), *Quantitative methods in cognitive semantics: corpus-driven approaches*, 355-380. Berlin & New York: Mouton de Gruyter.
- Stefanowitsch, Anatol & Stefan Th. Gries. 2003. Collocations: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2). 209-243.
- Sweetser, Eve. 1990. *From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- Tuggy, David. 1993. Ambiguity, polysemy, and vagueness. *Cognitive Linguistics* 4(3): 273-290.
- Vanderschueren, Clara & Kevin Diependaele. 2013. The Portuguese inflected infinitive: an empirical approach. *Corpus Linguistics and linguistic theory* 9(1). 161-186.
- Vesterinen, Rainer. 2010. The relation between iconicity and subjectification in Portuguese complementation: Complements of perception and causation verbs. *Cognitive Linguistics* 21(3). 573-600.
- Viberg, Åke. 1984. The verbs of perception: a typological study. *Linguistics* 21(1). 123-162.
- Ward, Joe H. 1963. Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58 (301). 236-244.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Oxford: Blackwell.
- Wulff, Stefanie. 2006. Go-V vs. go-and-V in English: A case of constructional synonymy? In Stephan Th. Gries & Anatol Stefanowitsch (eds.), *Corpora in Cognitive Linguistics. Corpus-Based Approaches to Syntax and Lexis*, 101-126. Berlin & New York: Mouton de Gruyter.