

**TOWARDS A DYNAMIC BEHAVIORAL PROFILE:  
A DIACHRONIC STUDY OF POLYSEMIOUS *SENTIR* IN SPANISH**

**Abstract:** This study examines the diachronic evolution of the polysemy of the Spanish verb *sentir* ('to feel') by means of a corpus-based dynamic Behavioral Profile (BP) analysis. Methodologically, it presents the first application of the BP approach to historical data and proposes some methodological innovations not only within the current body of research in historical semantics, but also with regard to previous applications of the BP approach. First, whereas the majority of existing studies in quantitative historical semantics are largely based on observed frequencies or percentages of collocational co-occurrence, our study leverages more complex historical data that are based on the similarities of vectors. Second, this study also provides an extension of the methodological apparatus of the BP approach by complementing the traditional Hierarchical Agglomerative Cluster analysis (HAC) with a dynamic BP approach derived from Multidimensional Scaling maps (MDS). Theoretically, this methodology contributes to a comprehensive perspective on the process of Constructionalization and the nature of networks, which is illustrated on the basis of the development of the Discourse Marker (DM) *lo siento* ('I'm sorry').

**Keywords:** polysemy, semantic change, constructionalization, Behavioral Profile, Multidimensional Scaling, perception verbs, Spanish

## **1. Introduction**

One of the areas of linguistic research that is trickiest to study is meaning. Questions of semantics such as what an expression means, how many senses an expression has, how exactly they differ and which one(s) is/are prototypical/central/primitive, how semantically similar different expressions are, etc. have long proved extremely hard to tackle for lexicographers, theoretical linguists, applied linguists, and others. For lexicographers, many of these questions have been problematic for centuries and were addressed with example collections and a generally keen eye for linguistic meaning differences. In the last, say, 30 years of theoretical linguistics, it was especially Cognitive Linguistics and Construction Grammar, which, given the central role they assign to meaning and function in the definition of units and constructions, had to face many of these essential questions. Over the last few decades, however, both disciplines, in spite of their very different emphases and goals, have been influenced by a strong empirical and quantitative turn in linguistics, a development which saw a huge increase in analyses using corpus data and statistical methods.

However, these developments of course require that one has a valid way of operationalizing, or at least beginning to approximate, meaning in corpus data for some later statistical analysis. The one conceptual step that probably virtually all corpus-linguistic analyses make is based on the distributional hypothesis, the notion that differences in function/meaning are reflected in differences in distribution (see Harris 1954, Firth 1957, Bolinger 1968). This notion has been instrumental for lexicographers to tease apart subtle meaning differences in example collections (as in the Cobuild or the Hector project, see Sinclair 1987 or Atkins 1992 respectively), and it has been just as instrumental in general semantic studies and cognitive-linguistic studies, where both lexical co-occurrence (aka *collocation*) and lexicogrammatical co-occurrence (aka *colligation* or *collostruction*) have given new impulses to semantic analyses of both synchronic and diachronic corpus data.

The main assumption underlying such studies, then, is that, for instance, different sense of the same form will co-occur with different lexical items or constructions. However, one big difference between different kinds of analyses is concerned with the amount of co-occurrence information used and how statistical analysis enters into the picture. It is probably fair to say that much lexicographic work used a lot of co-occurrence information from concordance lines, but was, until recently, relatively non-quantitative – co-occurrence information was included on a qualitative intuitive level based on lexicographer’s intuitions. Much work based on collocation and collocations was based on co-occurrence of senses or forms with elements in a particular window or slot around the expression being analyzed, but involved more statistical sophistication. And finally, some vector-space kind of analyses used large windows of text around the expression of interest and use sophisticated statistical tools on large number of items to explore differences between forms and/or senses.

The approach we use and extend in this study seeks a compromise between these different approaches: Our method, the Behavioral Profile (BP) approach, is based on co-occurrence information from concordance lines, but (i) it does not just use co-occurrence information from one slot or lexical co-occurrence information from a huge but noisy window, but it is based on comprehensive manual annotation of dozens of features characterizing each usage event, and (ii) it does not rely on a largely intuitive analysis of this huge co-occurrence information, but uses multivariate exploratory statistics to differentiate senses and forms as well as corresponding visualization.

This method has proven useful for the analysis of different phenomena in lexical semantics such as near-synonymy (Divjak & Gries 2006; Divjak 2010), antonymy (Gries & Otani 2010) and polysemy (Gries 2006; Berez & Gries 2009), but we will also make and exemplify two suggestions for how it can be extended. First, while most if not all BP studies were synchronic in nature, we apply now this BP approach to the diachronic evolution of the polysemy of the Spanish verb *sentir* (‘to feel’). Second, most BP studies used hierarchical agglomerative cluster analysis (HAC) as their main exploratory tool but in order to deal with diachronic data, we will show how the recent addition of multidimensional scaling (MDS) to diachronic semantics can also improve BP analyses.

This methodology will be illustrated by means of a case study on the diachronic evolution of the Spanish verb *sentir* (‘to feel’). In a previous synchronic BP study on the polysemy of *sentir* (Jansegers et al. 2015), it has been shown that this verb displays a rich profile, both semantically – extending from meanings of direct physical perception (both general physical perception (example 1) and specific modalities of perception, cf. example 2), through cognitive perception (3) to emotional values (4) – and syntactically – taking different kinds of complements:

- (1) Había apretado tanto los dientes que las mandíbulas me dolían, pero nada comparable al dolor agudo y penetrante que **sentía** en las costillas.  
‘I had clenched my teeth so much that my jaws ached, but nothing compared to the sharp, penetrating pain I *felt* in my ribs’.
- (2) Le hablé de [...] cómo me habían temblado las manos al **sentir** el roce de los labios de Nuria Monfort en la piel apenas unas horas atrás. [CREA: Ruíz Zafón, 2001]  
‘I told her how [...] my hands had trembled as I *felt* the touch of Nuria Monfort’s lips on my skin just a few hours ago.’
- (3) En el año y medio que estuvo en España, fue herido dos veces y estuvo a punto de morir, pero, entonces, **sentía** que ésa era su única razón de ser. [CREA: Press, 2003]  
‘In the year and a half he was in Spain, he was wounded twice and nearly died, but then he *felt* that this was his only *raison d’être*.’

- (4) Hilarón, [...], es el gallardo guardabosques fiel amigo de infancia de Giselle que **siente** un profundo amor por ella aunque no es correspondido. [CREA: Prensa, 2002]  
'Hilarón, [...] is the gallant ranger, faithful old friend of Giselle who *feels* a deep love for her although it is not reciprocated'

However, although the contemporary polysemy of the Spanish verb covers three general semantic domains related to the verb – physical (1,2), cognitive (3), and emotional perception (4) – only the first two were inherited directly from Latin: in its etymon, *sentīō*, *sentīre* only encodes physical and cognitive perception. Moreover, it turns out that contemporary Spanish even has developed a more discursive use of the verb in the apologetic construction 'lo siento' ('I'm sorry'):

- (5) Por eso no te escuchaba, **lo siento**. [CREA: Grandes, 2002]  
'That's why I didn't listen to you, I'm sorry.'

As we will see throughout this paper, from a quantitative point of view, it turns out that the most frequent sense of the verb in present-day Spanish refers to the emotional perception. As a consequence, the question arises as to how this semantic domain of emotion – and the related discursive use of 'lo siento' – has developed over the course of history.

In sum, we hope to not only contribute to the quantitative study of polysemy from a diachronic point of view, but also offer a comprehensive perspective on processes such as constructionalization. More precisely, in the present study we focus on the rise of *lo siento* ('I'm sorry') as a pragmatic marker and its constructionalization within the bigger picture of the changing polysemic profile of the verb through the course of history. Whereas most studies in diachronic semantics tend to draw straight lines between a specific construction and one single historical predecessor, our dynamic BP analysis shows that the path of change might not always be as linear. It allows thus to visualize "multiple inheritance" of constructions (Trousdale 2013) and the existence of multiple source constructions (De Smet et al. 2013). In doing so, the proposed analysis offers thus simultaneously a dynamic and multidimensional picture of its developing polysemy and a very fine-grained profile of the verb.

The outline of the paper is as follows: Section 2 describes the BP methodology in more detail according to the traditional four steps of analysis, namely (1) the retrieval of the data (Section 2.1); (2) the annotation of a large set of properties (Section 0); (3) the conversion of a co-occurrence table into vectors and (4) its evaluation by means of statistical techniques (Section 2.3). Section 3 provides the results of the dynamic BP analysis. First, the new approach will be validated on the basis of the synchronic data set by contrasting the HAC and the MDS technique (Section 3.1). It will be shown that the extension of the BP approach using MDS on synchronic data validates and even goes beyond the traditional HAC analysis. Then, we will discuss the results of the diachronic dynamic BP on the basis of the sequentially ordered MDS plots (Section 3.2), which will lead to a detailed case study on the constructionalization of the discourse marker *lo siento* (Section 3.3).

## 2. Data and method

Similar to previous studies that apply the BP method, the present paper adopts the following four-step procedure (cf. Gries & Divjak 2009; Gries 2010a; Gries & Divjak 2010; Gries & Otani 2010):

- (i) Step 1: the retrieval of all instances of the verb in context in the form of a concordance (see Section 2.1).
- (ii) Step 2: the manual analysis and annotation of 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 (see Section 2.2).
- (iii) Step 3: the generation of a co-occurrence table specifying which ID tag level is attested how often with each sense, i.e. the conversion of this table into vectors (see Section 2.3).
- (iv) Step 4: the evaluation of this table by means of exploratory and other statistical techniques (see Section 2.3).

From this, it becomes clear that, while our analysis follows the established BP practice in the first three steps, the fourth step is different: Whereas traditionally, the statistical exploration of BP analyses is based on HAC, our study presents a dynamic BP approach derived from MDS-based semantic maps in order to deal with the changing dynamic relationships within and between senses over time.

### 2.1. Corpus data: retrieval

Our data come from tens of thousands of manually annotated data points based on data from the Spanish *Corpus Diacrónico del Español*, (CORDE), which extends from the 13th until the 20th century, and the *Corpus de Referencia del Español Actual* (CREA) for the contemporary data. For a representative sample, we first manually retrieved all occurrences of the verb lemma *sentir*, starting from the last third of the 13th century ( $\pm 1270-1290$ ), which is generally considered the onset of Castilian prose (Alfonso X). Next, the same time period was retrieved with systematic intervals of 200 years each, leading to five chronological cutoff points (1270-90, 1470-90, 1670-90, 1870-90, beginning 21st century). This approach yielded a total of 14,782 instances of the verb lemma.

Second, we randomly sampled from each period containing >1000 occurrences of the lemma: 50% of the 15th century data and 25% of the 19th and 21st century data. Third, because of the lack of contemporary oral data, the CREA corpus was complemented with the oral data available for the 21st century extracted from the PRESEEA and the COLAM corpus. In order to ensure a maximum degree of representativeness, no previous selection regarding register and genre was made, taking as selection criteria only time (1270-90, 1470-90, 1670-90, 1870-90, beginning 21st century) and diatopic variant (only peninsular Spanish). The outcome of this data-gathering method is a corpus of 4488 instances in total. Table 1 summarizes the selected chronological cutoff points and the total number of analyzed occurrences per period:

Table 1: Our data as retrieved and sampled from the corpus

| Period           | Total | Total <i>sentir</i><br>(after random sample) | Total <i>sentirse</i> <sup>1</sup><br>(after random sample) |
|------------------|-------|--|---|
| XIII (1270-1290) | 187   | 151 (80.7%)                                  | 36 (19.3%)  |
| XV (1470-1490)   | 1254  | 549 (87.6%)                                  | 78 (12.4%)  |
| XVII (1670-1690) | 326   | 311 (95.4%)                                  | 15 (4.6%)   |
| XIX (1870-1890)  | 6149  | 1337 (86.9%)                                 | 201 (13.1%)   |
| XXI (2000-2004)  | 6742  | 826 (49.0%)                                  | 860 (51.0%)   |
| oral corpus XXI  | 124   | 86 (69.4%)                                   | 38 (30.6%)  |
| Total            | 14782 | 3260   | 1228  |

## 2.2. Corpus data: annotation

Next, we prepared the data for a BP analysis and our extensions: All 4488 occurrences were annotated for a large set of properties, called ID tags. Our ID tags are based on those used in a previous synchronic BP study of the polysemy of *sentir* (Jansegers et al. 2015), namely 34 different ID tags describing the presence or absence of 197 morphological, syntactic, semantic, and pragmatic characteristics. Table 2 is an excerpt of the ID tags and their levels:<sup>2</sup>

Table 2: Examples of ID tags and their levels (Jansegers et al. 2015: 393)

| General level      | Type of ID tag            | ID tag                    | ID tag levels  |
|--------------------|---------------------------|---------------------------|--|
| Verb               | Morphosyntax              | tense                     | present, past, future, infinite form   |
|                    |                           | person number             | 1, 2, 3<br>singular, plural  |
|                    | Semantics                 | general semantic category | general physical P, specific physical P, emotional P, cognitive P, ambiguous                         |
|                    |                           | specific sense            | experience physical perception, experience emotional perception, auditory perception, consider, etc. |
| Argument structure | Subject                   | lexical S                 | with S, without S  |
|                    | Direct Object             | form                      | nominal phrase, pronoun, gerund, infinitive <sup>3</sup> , proper noun, etc.                         |
| Adjuncts           | Adverbial adj. properties | referent semantics        | person, concrete, abstract, situation, ambiguous   |
|                    |                           | presence                  | with adverbial ad., without adverbial adj.   |
|                    |                           | form                      | adverb, prepositional phrase, nominal phrase, etc.   |
| Discourse          | Scope                     | predicational autonomy    | no, yes  |

Since the sense annotation is an essential part of the analysis, this aspect merits some comments. It should be noted that for the semantic analysis, we resorted to a very fine-grained annotation of the different possible senses of the verb. This was done manually and mainly on the basis of a previous tripartite case study (Enghels & Jansegers 2013), where (1) a lexicographic analysis was complemented by the results of a Romance comparative study based on a combined corpus approach involving both (2) translation data and (3) a comparable corpus.<sup>4</sup> This analysis has been performed in two phases: First, we distinguished very fine-grained senses that were minimally different, which then allowed classification into four general semantic categories, namely (1) general physical perception, (2) specific modality of physical perception, (3) emotional perception and (4) cognitive perception.

## 2.3. Statistical analysis

In a third step, we convert the annotations into a co-occurrence table providing the relative frequency of co-occurrence of each sense of the verb *sentir* (in the columns) with each ID tag level (in the rows). This procedure was performed with Gries's (2010b) BehavioralProfiles 1.01 script written for the R programming language. As exemplified in Table 3, the percentages of ID tag levels add up to 1 within each ID tag such that each column represents a set of co-occurrence percentages for one sense of the verb, which is the Behavioral Profile after which the method is named.

Table 3: Examples of BP vectors (Jansegers et al. 2015: 394)

| ID tag    | ID tag level | experience<br>physical<br>perception | experience<br>emotional<br>perception | auditory<br>perception | ‘consider’,<br>‘judge’ | ... |
|-----------|--------------|--------------------------------------|---------------------------------------|------------------------|------------------------|-----|
| tense     | present      | 0.30                                 | 0.36                                  | 0.29                   | 0.55                   | ... |
|           | past         | 0.35                                 | 0.4                                   | 0.53                   | 0.3                    | ... |
|           | future       | 0.01                                 | 0.01                                  | 0                      | 0.02                   | ... |
|           | infinite     | 0.34                                 | 0.23                                  | 0.18                   | 0.13                   | ... |
| lexical S | with         | 0.18                                 | 0.41                                  | 0.24                   | 0.41                   | ... |
|           | without      | 0.82                                 | 0.59                                  | 0.76                   | 0.59                   | ... |
| ...       | ...          | ...                                  | ...                                   | ...                    | ...                    | ... |

The last step is the exploration of this table using statistical techniques. As mentioned above, due to the diachronic nature of our data, we did not employ the traditional cluster-analytic approach but opted for a method that would not only discover discrete groupings of elements, but also visualize their dynamic interrelationships over time, MDS. A bit like principal component analysis, MDS is an exploratory dimensionality-reduction technique that seeks to condense a large number of dimensions in a multivariate data set into a smaller number of dimensions – typically two or three (cf. Wheeler 2005; Hilpert 2011, 2013: 66-74; Levshina 2011, 2015; Croft & Timm 2013). The basic assumption underlying this technique is that (dis)similarities between the entities under study can be represented as spatial distances. Applied to our concrete example, this means that the senses in Table 3 can be transformed into points in a typically two-, but theoretically also higher-, dimensional plane (or sphere) where distances between points representing senses reflect the (dis)similarities between the senses as well as possible. This procedure involves three main steps:

- (i) Quantification of the data. Here, our data are that already: We have the BP co-occurrence vectors for the senses of the lemma.
  - (ii) Distance computation: Here, we computed Euclidean distances between the BP vectors of all pairs  $\frac{30 \times 29}{2} = 435$  pairs of the BP vectors of the 30 senses into a distance matrix. This shows, for instance, that the two most similar senses are ‘EMO.encontrarse’ (‘to be in an emotional state’, as in (6)) and ‘FIS.GEN.encontrarse.metaf’ (metaphorical uses of the general physical state, as in (7)):
- (6) Personalmente tengo que decir que **me siento** muy satisfecho de haber trabajado con este hombre, discutido pero triunfador. [CREA: Del Rey del Val, 2002]  
‘Personally I have to say that I *feel* very pleased to have worked with this controversial but triumphant man.’
  - (7) 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. [CREA: 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.’

This outcome is intuitively very reasonable, because it is generally known that physical perception metaphorically relates to emotions (cf. among others Kurath 1921, Sweetser 1990, Kövecses 2008).

On the other hand, the two most different senses turn out to be ‘FIS\_GEN.Manifestarse’ in the pronominal passive causative construction (‘appear, show up’, as in (8) versus ‘COGN\_presentir’ (‘have a presentiment’, as in (9)).

- (8) 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. [CREA: 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.’
- (9) Después de dar alguna vuelta escuché arriba un lloro o un gemido. Subí a la habitación de mi hermana, la puerta estaba entornada y puedo jurarle que en aquel momento, antes de abrirla, **sentí** que me iba a dar algo. [CREA: Díez Rodríguez, 2002]  
 ‘After going for a walk I heard a cry or a moan upstairs. I went to my sister’s room, the door was left ajar and I can swear to you that at that time, before opening it, I *felt* I was going to go nuts.’

- (iii) Third, an MDS algorithm transforms the distance matrix into a set of coordinates for each sense, which can be presented in a two-dimensional plot; we used the `cmdscale` function in the R statistical software package. Figure 1 visualizes the result of an MDS analysis for the contemporary corpus by showing the main two dimensions on the *x*- and *y*-axes; the sizes of bubbles represent the relative frequency of each particular sense per century, the colors represent the general semantic category.

As one might hope for, the graph displays the cognitive senses in close proximity (see the group of blue dots on the right) as well as the close relationship between the physical metaphorical uses and emotional perception (resp. at coordinates [-3.01, 0.85] and [-3.15, 0.61] in the *sentirse* cluster on the left, where a green bubble of metaphorical physical uses is plotted within the red bubble of emotional senses). On the other hand, we find large distances between, say, FIS\_GEN.manifestarse (‘appear, show up’, see example (8) above) and EMO.lamentar (‘regret’, see (10)).

- (10) 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. [CREA: 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.’

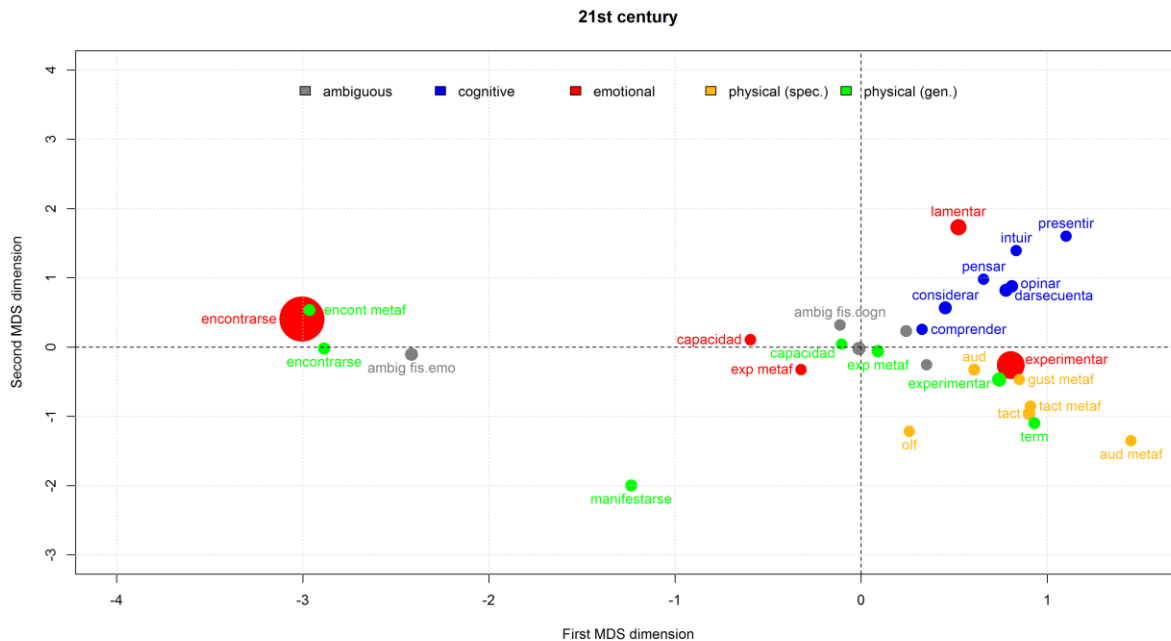


Figure 1: MDS plot of behavioral profile vectors (all data from the 21st century)

### 3. Results and discussion

#### 3.1. Validation of the approach using the synchronic data: HAC vs. MDS

Both HAC and MDS are bottom-up exploratory techniques to detect and represent structure in multivariate data, but they do so very differently. Since we are doing the first BP analysis with an MDS rather than an HAC, this section validates the MDS approach by comparing it to the results of the more traditional HAC approach. We use the contemporary data mentioned in Table 1 above with altogether 1810 concordance lines.

A previous synchronic BP analysis of the polysemy of the verb *sentir* using HAC (Jansegers et al. 2015) led to the dendrogram in Figure 2, in which different clustering solutions are indicated with red and blue rectangles.

The two red boxes indicate that the cluster analysis finds two large significant meaning clusters, which coincide with the division between middle voice uses (*sentirse*, on the top) and other active uses (*sentir*, on the bottom); henceforth, we refer to these two clusters as *SENTIRSE* and *SENTIR* respectively. The first one represents the cases of being in an emotional state as exemplified in (6) above and being in a physical state as exemplified in (11) below.

- (11) E Isabel de la Hoz, ahora, está de ocho meses, y **se siente** muy pesada y se le hinchan los tobillos, [...] [CREA: Pombo, 2004]  
 ‘And Isabel de la Hoz, now, is eight months [pregnant], and she *feels* very heavy and her ankles are swelling’.



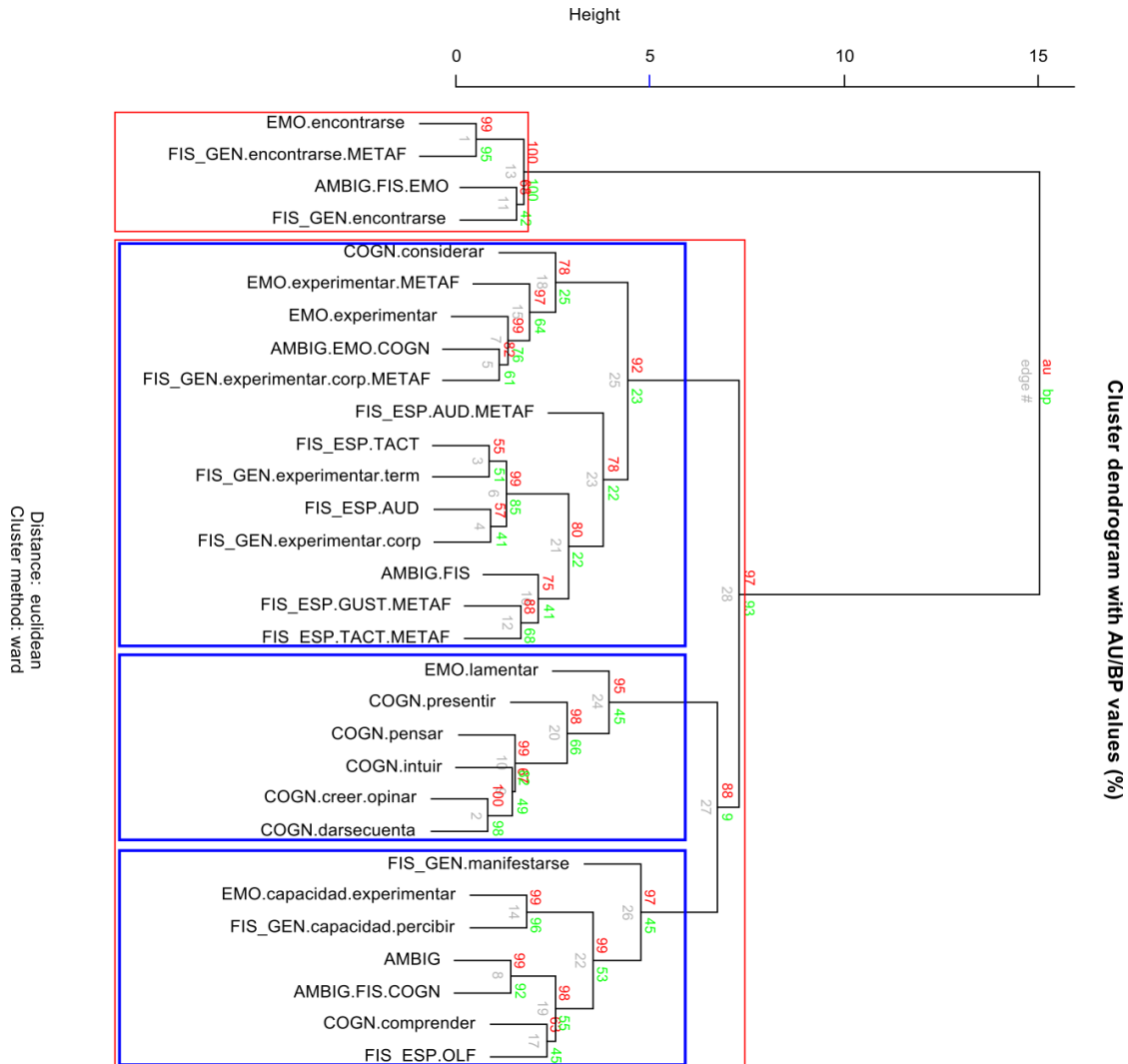


Figure 2: HAC dendrogram for the senses of *sentir*

Within the *SENTIR* cluster, three subclusters emerge (see the blue boxes), namely (a) ‘experience a physical or mental feeling’ (‘experimentar una sensación física o psíquica’, see (12) and (13) respectively), (b) some cognitive perception senses (e.g. example (9) above) and – curiously enough – the emotional regret cases (cf. (10) above) and (c) a cluster grouping the ability to experience or perceive something (see (14)) and the sense of ‘appear, show up’ (see (8) above). Each cluster also contains ambiguous cases (indicated in the dendrogram as AMBIG.X) which present multiple possible interpretations: (15) turns out to be ambiguous between a physical (here, physical health of a person) and an emotional state (here, mental health/well-being):

- (12) [...] pidió a un colega suyo que le extrajese una muela que tenía estropeada utilizando el gas de la risa, y no **sintió** dolor. [CREA: Sabadell, 2003]  
‘he asked his colleague to extract a tooth he broke using laughing gas, and he did not *feel* pain.’
- (13) Los marroquíes **sienten** una fuerte atracción hacia el lujo material [CREA: Silva, 2001]  
‘Moroccans *feel* a strong attraction towards material luxury.’

- (14) Y Sofía quiere meterse algo al cuerpo y no **sentir**. [Beccaria, 2001]  
 ‘And Sophia wants to put something into her body and not *feel*.’
- (15) Yo me **sentía** bien en mi cuerpo, como se siente un hombre sano después de hacer el amor. [CREA: Llongueras, 2001]  
 ‘I *felt* good in my body, as a healthy man feels after making love’.

These semantic clusters have syntactic correlates, which explain some seemingly counterintuitive clusterings: The first cluster (*SENTIRSE*) correlates with the presence of a predicative complement oriented at the subject. Within the second large *SENTIR* cluster, the first subcluster (physical/mental feelings) tends to occur with an NP; The second subcluster (cognition/regret) features situational complements (e.g., *that* complementation or neuter clitic pronouns); the third cluster (ability) relates to the absolute use and pronominal passive causative constructions meaning ‘appear, show up’ (Jansegers et al. 2015: 415).

Thankfully, the MDS results in a very compatible picture: As in the HAC, the MDS plot in Figure 1 shows a clear division of the middle voice uses of *sentirse* on the left vs. the other active uses of *sentir* on the right. Within this second large cluster, the bottom-right part of this cluster groups the physical perception uses (general/green and specific/orange) together with general emotional experience, which corresponds to the first subcluster within the *SENTIR* cluster in the HAC results. Also, the cognitive/blue senses group together in the top-right part of this cluster, which corresponds to the second subcluster within the *SENTIR* cluster in the HAC analysis. It is also worth noticing that the majority of the metaphorical uses of the verb together with the ambiguous cases are close to 0 along the y-axis.

Focusing on the periphery, the extreme positions along the y-axis in the MDS plot seem to be occupied by the cases identified as (more or less) grammaticalized instances (i.e. the ‘regret’ sense, ‘EMO.lamentar’), or non-prototypical uses of the verb (i.e. the specific pronominal passive causative constructions of FIS\_GEN.manifestarse, ‘appear, show up’) on the basis of the HAC analysis. The extreme value of the FIS\_GEN.manifestarse sense on the y-axis seems to confirm its status as a particular construction, suggested in the HAC analysis. In these cases, *sentir* combines with a non-prototypical subject in the sense that it does not act as the experiencer (as in the large majority of the cases), but as the stimulus of perception, as was illustrated in example (8). Similarly, the ‘EMO.lamentar’ sense in the extreme top includes the fixed expression *lo siento* (‘I’m sorry’) functioning as an apologetic discourse marker. This construction may be associated with that extreme position because of its features, which we discuss in more detail in Section 3.3.

In sum, the MDS results are compatible with the so far predominant HAC approach. However, some subtle advantages of MDS need to be mentioned as well (cf. also Levshina 2011: 90). First, the application of HAC implies categorical splits of the data – something is a member of one and only one cluster, or it isn’t – which is certainly not a cognitively realistic representation of phenomena like polysemy or (near) synonymy and can be misleading for analytical purposes and in how it can suggest comparatively clear distinctions when other methods and/or linguistic intuition would lead one to expect comparatively high degrees of similarity or overlap. Therefore, the potential to more finely visualize the continuity of the semantic space can be an important advantage of MDS over HAC. For a concrete example, note in the transition area (along the x-axis) between the *SENTIRSE* cluster (middle voice use) on the left and the *SENTIR* (active) cluster on the right, we find precisely the ‘FIS\_GEN.manifestarse’ sense, i.e. the non-prototypical cases where the subject of the verb is not the prototypical experiencer (EXP), but the stimulus (STIM). In the same transition area between the two main clusters, although more towards the periphery of the *SENTIR* cluster, we find the absolute cases of the verb, conveying the sense of ‘capacity to feel something’, both in the physical and the emotional sense. This corresponds to the third subcluster in the *SENTIR*

cluster from the HAC analysis. Their position in the MDS solution makes sense, because contrary to the middle voice uses (*SENTIRSE* cluster), the absolute uses – just as the active uses (*SENTIR* cluster) – always imply a certain stimulus (although implicitly). The MDS analysis allows thus to represent this underlying syntactic continuum: Represented on a cline, this is what we see reflected here in the plot: *SENTIR* = EXP + explicit STIM > *sentir*<sub>[abs]</sub> = EXP + implicit STIM > *sentir*<sub>[manifestarse]</sub> = implicit EXP + explicit STIM > *SENTIRSE* = only EXP.

It is thus clear that this extension of the BP approach using MDS on synchronic data validates but also insightfully extends the previous HAC analysis. In the next section, we turn our attention to how MDS handles the diachronic dimension.

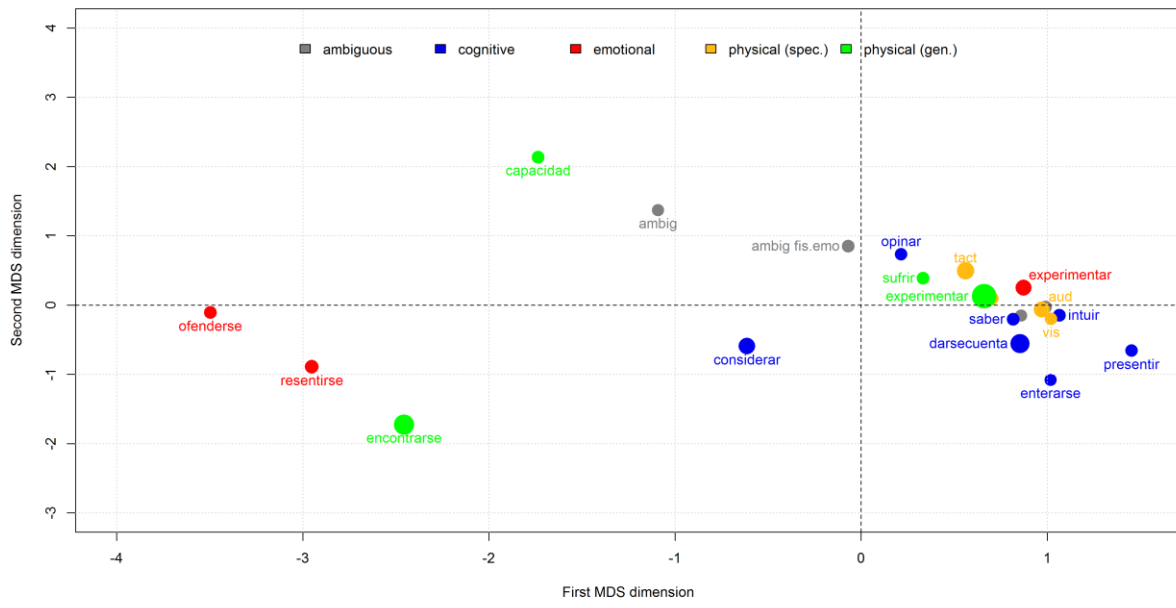
### 3.2. *Visualizing semantic change: a dynamic Behavioral Profile analysis*

MDS has already been applied in different areas of linguistic research, such as sociolinguistics (Szmrecsanyi & Kortmann 2009; Szmrecsanyi 2011; Szmrecsanyi & Wolk 2011), typology (Croft & Poole 2008), and lexical semantics (Majid *et al.* 2008; Levshina 2011) but it is only recently that it has edged its way towards the enigmatic matter of historical semantics (e.g. Hilpert 2011; Sagi *et al.* 2012).<sup>5</sup> This innovation is largely due to the work of Martin Hilpert (2011, 2013, 2016), who introduced MDS as a means for the dynamic visualization of language change. Starting from the distributional hypothesis, the essential idea is that different distance matrices for different time periods can be used to generate a set of MDS maps that are displayed in sequence to visualize diachronic semantic development. From a diachronic perspective, each element is thus not only compared to a set of other elements, but essentially also to itself at different points in time, which allows thus both a holistic and a more detailed interpretation of the change (Hilpert 2011: 437; 2013: 67-68).

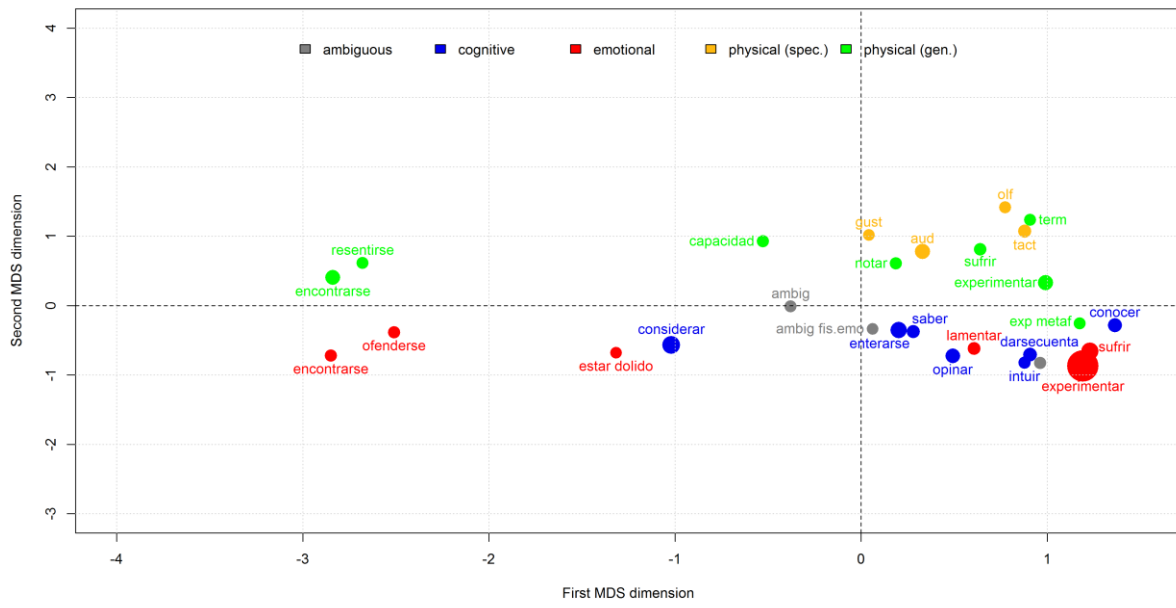
However, just like its synchronic counterpart, the majority of the existing MDS work on diachronic data focuses on changes in collocational profiles over time (Sagi *et al.* 2012; Hilpert 2016), or is essentially geared towards morphosyntactic change based on relative frequency profiles (Hilpert 2011, 2013). In other words, restricting the analysis to collocational and/or frequency profiles, the distributional hypothesis and its application to historical data has not yet been exploited to its full potential and we propose to extend the existing lines of research by including the fine-grained BP vectors as input to MDS analyses.

To that end, the procedure explained in Section 2 was performed separately for each time slice distinguished in the diachronic corpus, yielding five different distance matrices and, subsequently, five separate MDS analyses. Apart from the specific restrictions and challenges that all quantitative diachronic corpus methods face due to the particular nature of the data (bias towards specific registers, authors and genres, discontinuity of certain genres, predominance of certain diachronic periods, sparseness of data, etc.), the present study also has to address another concern: Since we are analyzing the dynamic development of a polysemic verb, some senses and distributional features will not be attested in all time periods, i.e., the general frame of reference and comparison may change over time. To maximize interpretability and comparability, for each time slice, we calculated the cumulative percentage of all the senses – starting from the most frequent one – that together represent 95% of the data. This is a compromise between a too fine resolution of the senses (taking into account all different senses per time slice, but thereby including infrequent senses and many hardly attested features) and a too coarse-grained one (only taking into account the senses shared by all time periods and throwing out much of the data): We cover the vast majority of the data for which the most data points are available. Figure 3 visualizes the development of the polysemy of the verb *sentir* by means of different MDS plots, one for each time slice.

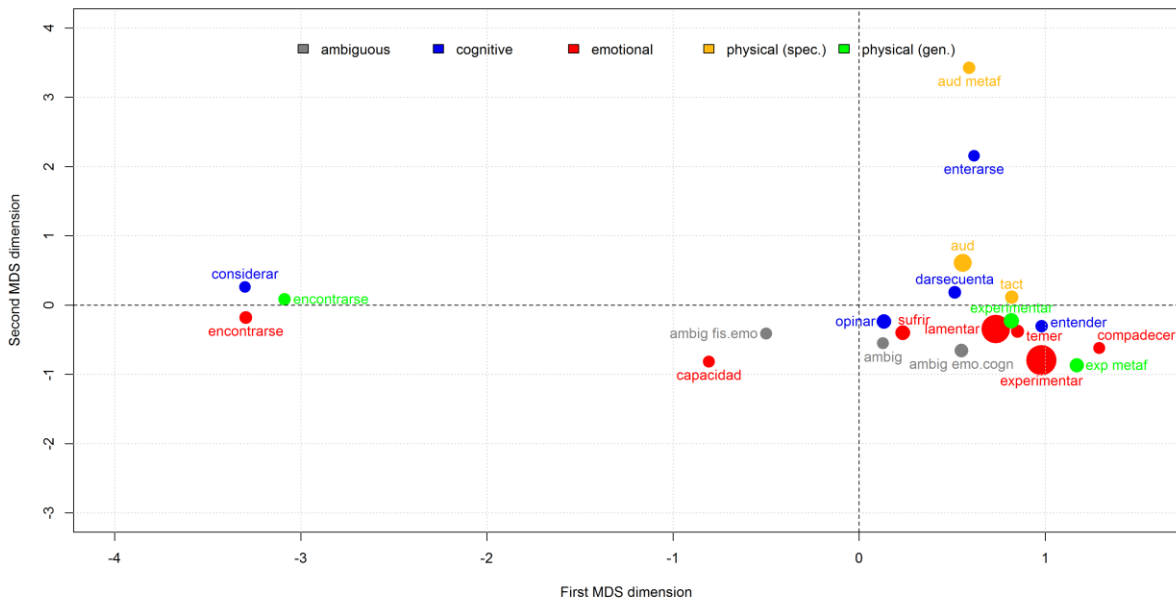
13th century



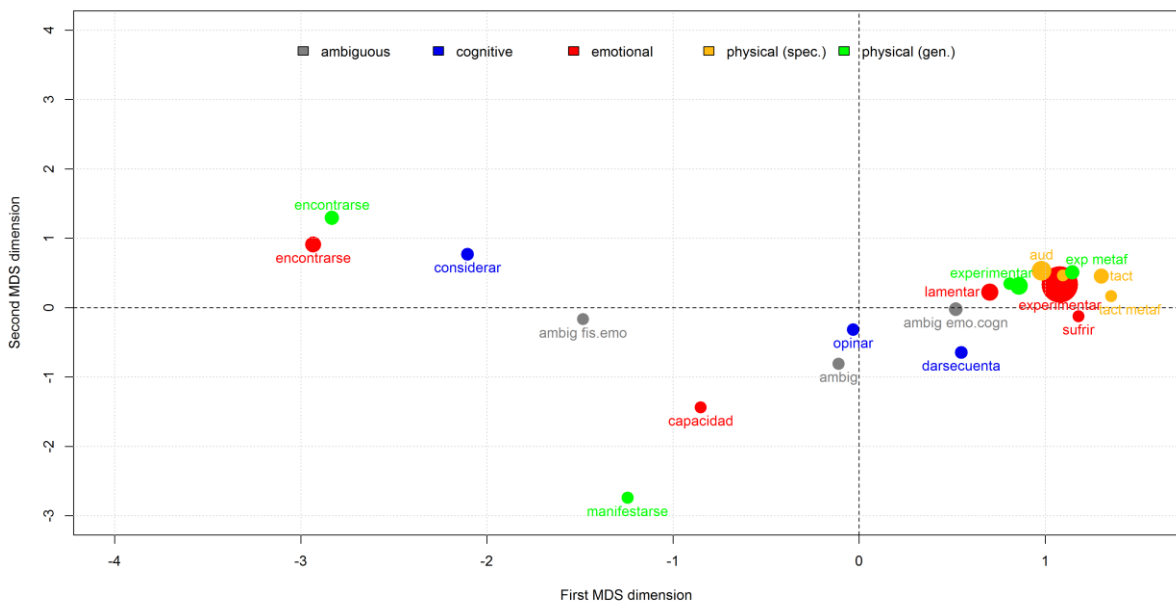
15th century



17th century



19th century



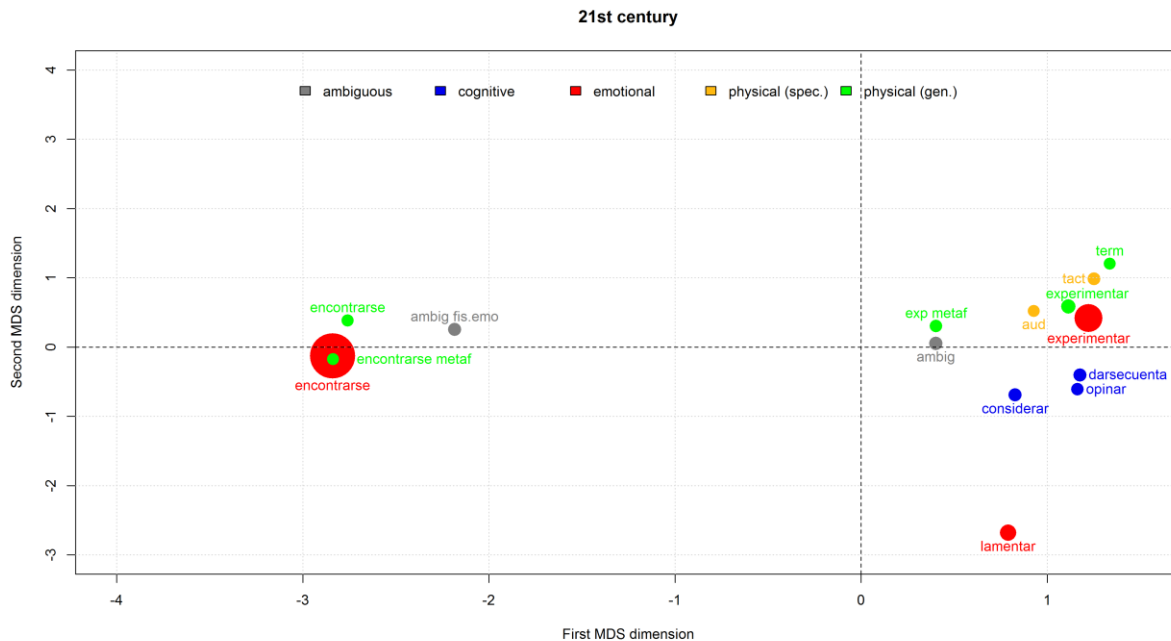


Figure 3: MDS plots of behavioral profile vectors (95% of tokens of each century)

To begin with, it is useful to make some general observations. At first glance, the comparison across centuries shows that the basic division along the  $x$ -axis into two big clusters (*SENTIRSE* vs. *SENTIR*) related to argument structure (cf. also Section 3.1) remains constant across all centuries, but it is most pronounced in the 21st century, where a large white space around  $x=-1$  sharply separates both clusters. Similarly, the 21st century plot shows a semantic continuum along the  $y$ -axis from concrete to abstract and pragmatic meanings: From physical perception at the top and emotional perception in the lower half of the top, via the ambiguous and metaphorical cases around point 0, towards cognitive senses in the lower half and ultimately the extreme negative position of the pragmatic ‘regret’ sense. Interestingly, the fairly clear division we see in the contemporary uses has not always been this clear: In earlier centuries, the center of the plot is often more populated and there is considerable variation along both axes over time. This raises the question as to how the relationships between the different senses have changed over time and what the driving historical forces behind the contemporary 21st century picture are. Therefore, a more detailed analysis per century is required.

Starting from the 13th century, the first two dimensions of the MDS account for 51.4% of the semantic variation between the senses that are displayed. First, within the *SENTIRSE* cluster on the left-hand side of the graph, apart from the general physical meaning (FIS\_GEN.encontrarse; cf. (16)), the middle voice use seems to be explicitly related to the negative pole of emotional perception, and more precisely in the senses of ‘be offended’ (EMO.ofenderse; cf. (17)), and ‘to resent’ (EMO.resentirse; cf. (18)). More than that, it turns out that if the middle voice use expresses emotion in the 13th century, it is exclusively negative – the more neutral meaning of ‘being in an emotional state’ (EMO.encontrarse) expressed by the middle voice use appears only from the 15th century onwards.

- (16) Diag ordonnez otrossi quando se **sintió** mal ferido fue contra Rodrig arias & diol una ferida por somo de la cabeça; [CORDE: Alfonso X, 1270-1284]  
 ‘Moreover, when Diego Ordóñez *felt* seriously injured he went after Rodrigo Arias and wounded him on top of his head’.

- (17) [...] por amor de dios uos Ruego que non uos **sintades** en uuestro coraçon del mal daquellos falsos de carrion que Rescebiestes [...].[CORDE: Alfonso X, 1270-1284]  
 ‘for the love of God, I beg you that you are not *offended* in your heart because of the harm that you received from those impostors from Carrion’.
- (18) de los dannos & males que rreçibe. **siente sse** & enssanna sse commo varon non llora commo muger al onbre es propia la ssanna con obras & a la muger bozes & lagrimas [CORDE: Alfonso X, 1270-1284]  
 ‘be resentful and cruel as a man because of the harms and damages, do not cry as a woman, cruelty with actions is typical of a man, moaning and tears of a woman’.

Note that in (17) and (18), syntactically, *sentirse* enters into a very specific construction with a prepositional object introduced by *de*. Semantically, this is different from the others in that the negative semantic weight is not in the complement, but in the verb itself (equivalently in this example to ‘feel remorse’). Moreover, a considerable number of our concordance examples expressing physical perception relate explicitly to physical pain (cf.(16)).<sup>6</sup> In other words, from its onset, *sentirse* was intrinsically related to a pronounced negative semantic prosody, and more precisely, the expression of pain (both physically and emotionally).<sup>7</sup>

Within the *SENTIR* cluster, we see a division along the *y*-axis between cognitive (lower part) and physical senses (upper part). The only emotional sense is the general sense of ‘experience an emotional perception’ (EMO.experimentar). This predominance of physical and cognitive senses in the 13th century is in line with the Latin etymon of the verb: The Spanish verb *sentir* derives directly from the latin verb *sentīō*, *sentīre* (‘perceive by means of the senses’, ‘to realize’; cf. DCELC 4, 190sq., *sentir*). In other words, in its etymon, *sentir* encodes both general physical perception and cognitive perception, which we still see reflected in its early Spanish profile from the 13th century. Interestingly, between the cognitive perception group, we also find one physical perception sense such as visual perception (FIS\_GEN.VIS, as in (19)). This makes perfect sense because it is generally known that from all the physical perception modalities, vision is most closely related to cognition, as reflected by expressions such as ‘seeing is believing’ (cf. the SEEING IS KNOWING metaphor and the tautology *I saw it with my own eyes*, Sweetser 1990: 33). This also shows that – just like other verbs of perception – *sentir* frequently fosters metaphorical mappings from concrete or physical meanings onto more abstract, mental domains as an essential part of its polysemous character.

- (19) Et assi commo entro a la puerta daquella casa del Minotauro & fue yendo por las calleias della a adelant. yua dexando el filo como las Infantes le enuiaran consseiar.[...] Et des que fue muy adentro en la casa yl **sintió** el Minotauro; lleuantos & uenie contra el. abierta la boca; por coger le en ella. [...] [CORDE: Alfonso X, 1275]  
 ‘And as he entered the door of the house of the Minotaur he walked through the passages while he unwound the ball of wool as the Infants had recommended to him. When he went deep within the house, there he *saw* the Minotaur that raised himself and went against him with open mouth in order to draw him into it’.

In the transition zone between both clusters along the *x*-axis, apart from some ambiguous senses, we see ‘FIS.GEN\_capacidad’, i.e. the general ability to perceive a physical experience, which illustrates that the syntactic continuum along this axis referred to above is already present in the 13th century: Between the explicit presence of the STIM (*SENTIR*) and its complete omission (*SENTIRSE*), we find the absolute cases of the verb. Note also the central

location of the COGN.considerar sense ('to consider, to judge'): From a syntactic point of view, this sense is situated between both clusters because of the external stimulus (as in the *SENTIR* cluster; cf. *la mayor* ('the bigger animal') in (20)) to which the experiencer attributes a certain quality/characteristic by means of a predicative complement (as in the *SENTIRSE* cluster; cf. *más fuerte* ('stronger') in (20)). The only difference with the *sentirse* cluster is that this predicative complement is geared towards the external stimulus (expressed by the direct object) and not interiorized towards the experiencer (expressed by the subject).

- (20) [...] ca assí contece esto e dalo la natura que en todas las animalias la que menor es e más flaca que teme a la mayor que **siente** más fuerte que sí e obedecel [...]. [CORDE: Alfonso X, 1275]  
 'because that is how it occurs in nature, that in all the animals the one that is smaller and thinner fears the bigger one, which he *considers* stronger than himself and he obeys him'.

This specific cognitive use of the verb behaves somewhat differently from the other cognitive uses of the verb, as was illustrated already on the basis of the HAC analysis (cf. Section 3.1). Indeed, 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 speaker. These examples instantiate thus an epistemic shift towards subjectivity. Therefore, this sense can be paraphrased as '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') (Jansegers et al. 2015: 404). This lower control is also what relates it to the middle voice uses (cf. Maldonado 1999) and what we see reflected in this plot; we might therefore have some diachronic evidence for the contemporary BP profile. However, following its historical path in the successive plots, it turns out that this sense is meandering back and forth between both clusters, which could point towards changing semantic and syntactic forces in the course of its evolution (cf. Section 3.3).

In the 15th century plot, the first two dimensions explain 52% of the total variance. First, the *SENTIRSE* cluster on the left has become more diversified: Not only can the middle voice express negative emotion (EMO.ofenderse, 'to offend'), it also expands its use towards the general and more neutral meaning of 'being in an emotional state' (EMO.encontrarse; cf.(21)). A similar kind of expansion characterizes the physical perception senses, which not only expresses the neutral meaning of 'being in a physical state' (FIS\_GEN.encontrarse), but can also express the negative pole of physical perception in the sense of 'to suffer, weaken' (FIS\_GEN.resentirse; cf.(22)):

- (21) E la infanta no salió a la fiesta porque se avía **sentido** enojada. [CORDE: Valera, 1487-1488]  
 'And the princess did not go to the party because she had been *feeling* angry'.  
 (22) Galpano que se **sintió** de una ferida que tenía en la cabeça, que la sangre le caía sobre los ojos, se tiró afuera por los limpiar [...] [CORDE: Rodríguez de Montalvo, 1482-1492]  
 'Galpano, who *suffered* from a wound he had on his head, the blood rolling over his eyes, jumped outside to wash them'.

This FIS\_GEN.resentirse sense, parallel to the EMO.resentirse sense already existent in the 13th century, enters in the specific construction with prepositional object introduced by *de*. Semantically, this construction differs from the others because the semantic weight resides in the verb itself, not in the complement. Parallel with this evolution, a new absolute use arose



conveying the resultative meaning of ‘being negatively affected, in pain’ (EMO.estar afectado.dolido):

- (23) [...] pero como el rrey estava **sentido** y enojado de las mentiras pasadas, no le dió el crédito que solía, [...]. [CORDE: Enríquez del Castillo, 1481-1502]  
‘but since the King was *offended* and upset because of the past lies, he did not give him the credit that he used to’.

This sense is related to the copular construction where *sentir* functions as a predicative complement of the subject. Although, here, the status of the verb as nucleus of the predicate is subject to discussion, note that, similarly to the examples discussed above, in this sense the verb does not require the explicit presence of a negative complement in order to express negative emotion, but absorbs the negative semantic load itself. In other words, the negative prosody related to *SENTIRSE* in the 13th century becomes even more pronounced in the 15th century.

Within the *SENTIR* cluster, in this plot the top-right quadrant exclusively groups both general physical perception senses and more specific perception modalities such as tactile, gustatory, and olfactory perception. The bottom-right quadrant, by contrast, groups cognitive and emotional senses. The distance between the basic two original senses of the verb – physical and cognitive perception in Latin – thus becomes more pronounced in the 15th century. Interestingly, also within this cluster a group of senses seems to emerge that is explicitly related to negative emotion; here, we find the first instances of the particular ‘regret’ meaning (EMO.lamentar), which is close to the negative emotional meaning ‘EMO.sufrir’ (‘to suffer’) and to the cognitive senses of ‘opinar’ (‘to believe, think’) and ‘darse cuenta de’ (‘to realize’, ‘become aware of’).

This negative emotional cluster becomes even more noticeable in the 17th century. Indeed, in the bottom right-hand side of the plot, we see that the ‘regret’ sense (EMO.lamentar; (24)) is situated in close proximity to other strong negative emotions such as ‘to suffer’ (EMO.sufrir; (25)) and ‘to fear’ (EMO.temer; (26)). The 17th century is characterized thus by an explicit negative prosody related to the *SENTIR* cluster. At the same time, within this emotional cluster, EMO.compadecer (‘to feel pity, sorry for someone’; (27)) expresses shared pain. This empathetic sense could be key in the evolution towards a discourse marker use *lo siento* (‘I’m sorry’) in later centuries (cf. Section 3.3).

- (24) [...] el rey Felipe Segundo, que **sintió** grandemente la muerte temprana de hijo que, en su tierna edad, daba muestras de gran prudencia. [CORDE: Abarca de Bolea, 1679]  
‘King Philip the Second, who extremely regretted the early death of his son, who in his tender age, showed signs of great caution’.
- (25) Procura mortificar  
tu querer propio, y **sentir**,  
pues el grano, sin morir,  
no puede fructificar.[CORDE: Panes, 1675]  
‘try to mortify your own desire, because the grain, without dying, cannot bear fruit’.
- (26) Y tú, que fuiste el motivo [...] de esta desdicha, conoce  
que el no quitarte la vida  
es porque **sientas** al doble  
mi rigor y tu castigo,  
llorando en penas mayores,  
que él recibiendo la herida,

eres tú quien siente el golpe. [Barrios, 1672]

‘And you, who were the cause [...] of this misery, know that the fact of not taking your life, is so that you *fear* both my rigor and your punishment, crying in the largest griefs, that he who receives the wound, it is you that feels the blow’.

- (27) Nuestro rey Felipe IV (que está en el cielo) entró en la ciudad de Lérida a siete de agosto, año 1644, y viendo su Majestad lágrimas en uno de los consellers, que pedían el perdón por toda Cataluña, le levantó con su real mano y le dijo **sentía** su pena y que siempre los había amado mucho. [CORDE: Abarca de Bolea, 1679]

‘Our King Philip IV (who is in heaven) entered into the city of Lérida on August 7th of the year 1644, and His Majesty, seeing tears in the eyes of one of his ministers who implored forgiveness for Catalonia, lifted him with his royal hand and told him that he *feels* his pain and that he had always loved them a lot’.

Observing the *SENTIR* cluster in general, it is interesting to notice that, in the 17th century, the emotional perception meanings seem to become predominant and more diversified compared to earlier, when the physical and cognitive senses inherited from its Latin etymon were still dominant. Also, this century sees the rise of emotional perception in the absolute construction (EMO.capacidad), which is situated in the center of the plot, according to the syntactic continuum along the *x*-axis. Regarding the *SENTIRSE* cluster, the evolution is almost the reverse, since the senses that are explicitly related to negative emotion that characterized previous centuries, now disappear in favor of the general physical emotional perception (EMO.encontrarse). The first two dimensions of the 17th century MDS plot account for 59% of the semantic variation between the senses that are displayed.

In the 19th century plot, the two plotted dimensions account for 65% of the total variance. In general, the senses seem to be more dispersed and the center of the graph is more populated. In the top right, as in other centuries there is an important cluster of auditory, tactile and physical perception senses (FIS\_ESP.AUD; FIS\_ESP.tact; FIS\_GEN.experimentar.corp; FIS\_GEN.experimentar.term respectively) as well as their metaphorical counterparts (FIS\_ESP.AUD METAF; FIS\_GEN.experimentar.corp.METAF). Interestingly, the cases of general emotional perception (EMO.experimentar) are closely to these physical senses, which points to the relatedness of the emotional and physical perception as in the HAC results for the 21st century, where physical and emotional perception together form the first subcluster within the *SENTIR* cluster.

Overall, compared to the 17th century, in the 19th century, the number of explicit negative emotional senses decreases and (general) physical perception rises again. Exploring some examples suggests this importance of the physical perception sense is related to the omnipresence of auditory-perception expressions in this century:

- (28) Rosalía, que estaba sola en la habitación interior, **sintió** los pasos de Horacio, **sintió** el rechinar de la puerta, miró y le vio entrar. [CORDE: Pérez Galdós, 1872]

‘Rosalie, who was alone in the interior room, *heard* the footsteps of Horace, *heard* the gnashing of the door, looked and saw him entering’.

Going further down along the *y*-axis, we then find ‘regret’ (EMO.lamentar) and the metaphorical uses of tactile perception (FIS\_ESP.TACT.METAF). This location between the physical and emotional perception of the ‘regret’ meaning is quite striking compared to its extreme position in the 21st century. We will discuss this particular evolution in more detail in Section 3.3.

Compared to the 19th century, the semantic divisions and boundaries in the 21st century turn out to be more clear-cut: the total amount of variance explained by both

dimensions is 73%. After nine centuries of peripeteias along and across both dimensions, the 21st century displays a rather straightforward image: Semantic division along the *y*-axis from concrete, physical perception towards an abstract cognitive cluster via emotional perception and an extreme pragmatic use; syntactic division along the *x*-axis between middle voice uses (*SENTIRSE*) and active uses (*SENTIR*).

Up until now, we have traced the semantic evolution of the verb exclusively based on its context. Besides context, another important aspect in language change is frequency: A focus on the general frequency evolution allows us to go beyond the particular level of the individual senses and to make generalizations on the level of overarching semantic categories as well as reveal more general underlying tendencies in semantic change. Indeed, if we focus on the frequency and color evolution based on the general semantic categories, the picture changes from mainly large green/orange (i.e. general/specific physical perception) bubbles in earlier centuries towards a very pronounced red (emotional perception) panorama in the present century. This evolution matches the well-known tendency typical of polysemic entities to extend their more concrete, physical meanings towards more abstract domains (Sweetser 1990). In other words, in accordance with its Latin heritage, *sentir* starts off as a predominantly physical – and cognitive – perception verb. However, as the data show, its contemporary Spanish homologue is dominated by a semantic domain that was non-existent in Latin. Indeed, although in present-day Spanish we distinguish three general semantic domains related to the verb – physical, cognitive, and emotional perception – only the first two were inherited from Latin. Moreover, it turns out that the present-day emotional pole of the verb fosters even more discursive uses as an apologetic discourse marker, visualized in the contemporary plot by its extreme isolated position in the bottom part of the *y*-axis:

- (29) Creo que darle demasiada importancia al aniversario acabaría siendo peor. Por eso no te escuchaba, **lo siento**. [CREA: Grandes, 2002]  
'I think that attaching too much importance to the anniversary would make it worse. That's why I didn't listen to you, *I'm sorry*'.

That way, as visualized clearly by means of the sequential MDS plots, the three most frequent emotional senses in the contemporary corpus are exactly the ones that one would assume to be maximally different: (a) being in an emotional state (correlated with the middle voice use *SENTIRSE*), (b) experience an emotional perception (correlated with the *SENTIR* cluster) and finally (c) the typical 'regret' sense (related to the specific apologetic construction).

As a consequence, the question arises as to how this particular sense arises and develops in the course of history. That is, can we identify any diachronic explanation and justification for the synchronic construction *lo siento*? In the next section, we will analyze this diachronic development within the theoretical framework of Construction Grammar.

### 3.3. Theoretical implications. Constructionalization of *lo siento* ('*I'm sorry*').

Originally, Construction Grammar (CxG) was primarily applied to synchronic data. More recently, however, there is an increasing body of research exploring diachronic applications of CxG (e.g., Traugott & Trousdale 2013; Hilpert 2013; Traugott 2014; Boogaart et al. 2014; Barðdal et al. 2015; and contributions in Giacalone Ramat et al. 2013). While these studies largely focus on (morpho)syntactic phenomena, Hilpert (2013: 210) argues:

constructional change can proceed at very different levels of linguistic structure from allomorphy to syntax. The next logical step would be to go

beyond syntax and into the realm of discourse in order to examine constructional change as it unfolds in actual instances of spoken conversation.

This is precisely the aim of the present section, which aspires to apply a CxG approach to the diachronic development of the discourse marker *lo siento* ('I'm sorry'), based on the previously discussed dynamic BP that takes into account different levels of linguistic analysis. As such, the basic claim of this section will be that an encompassing methodology such as the dynamic BP is instructive – maybe even necessary – to do justice to the multifaceted notion of constructional change/constructionalization.<sup>8</sup>

In general, constructionalization (Cxzn) and grammaticalization (Gzn) are two tightly intertwined theoretical approaches, and an extensive analysis of the differences between these two frameworks goes beyond the scope of the present study (see e.g. Hilpert 2013, Traugott & Trousdale 2013 for exhaustive comparisons). Most importantly for us, a Cxzn perspective has several advantages over the Gzn model. First, whereas Gzn research has tended to focus either on change in form or change in meaning, the Cxzn model combines both form and meaning with equal weight. Second, the treatment of discourse markers (DM) has always posed a major challenge for the theory of Gzn. A Cxzn perspective, by contrast, assumes that “a linguistic model should in principle be able to account for all facets of a speaker’s knowledge about their language” (Boas 2013: 234). This all-embracing view of language implies that no one linguistic domain is core: The basic unit of grammar is a form (phonology, morphology, syntax)-meaning (semantics, pragmatics, discourse-function) pair and DM are simply discourse management constructions. Thus, we are here adopting a Cxzn perspective, where both form and meaning are treated with equal weight and where *meaning* can be understood in the broad sense of *function* to include not only (lexical) meaning but also discourse function, information structure, and other pragmatic phenomena.

One particular group of DM are the so-called verbal DM, which cover different elements, among others comment clauses that express stance (Stenström 1995; Quirk et al. 1997; Biber et al. 1999; Brinton 2008) such as *I believe, I mean, you see* and different types of conversational routines such as *thank you, goodbye, (I'm) sorry* (Aijmer 1996). The latter have received relatively little scholarly attention and have been characterized mainly in terms of lexicalization (Norrick 1979, Aijmer 1996). In the present section, we will precisely focus on the development of one such a conversational routine in Spanish from a more encompassing perspective of Cxzn, namely the rise of the apologetic marker *lo siento* ('I'm sorry'):

- (30) [...] perdóname, papá, por lo de antes, la verdad es que soy un imbécil, no debería haberte dicho eso porque no lo pienso, **lo siento** mucho, en serio, no sé lo que me ha pasado... [CREA: Grandes, 2002]  
'Forgive me, father, for what happened in the past, the truth is I'm an idiot, I should not have said that to you, because I don't believe it, *I'm really sorry*, seriously, I don't know what happened to me'.

In accordance with the general definition of a construction presented above, the apologetic meaning of *lo siento* is not fully predictable from its component parts, that is, it is not the sum of its individual parts, since this would literally yield 'I feel it'. In what follows, we will concentrate on the DM *lo siento* as a construction, specifically on (i) the diachronic development both of its form and meaning and (ii) its development within the polysemous network and frequency changes of *sentir* in the course of time. More particularly, the contemporary construction *lo siento*, on the form side, has a first person singular in the present tense combined with the clitic *lo*, and, on the meaning side, a negative emotional

meaning. Consequently, several questions arise that are of interest from the diachronic perspective of Cxzn:

- 1) Are there significant frequency shifts in the semantic evolution of the verb in general and the emotional meaning in particular that could have facilitated the emergence of a DM?
- 2) From a (morpho)syntactic point of view, where does the clitic *lo* originate?
- 3) From a semantic point of view, where does the negative emotional meaning come from?
- 4) Finally, how did *lo siento* emerge and evolve diachronically with regard to the general semantic landscape and network of *sentir*? In other words, what is the source construction of this DM?

First, the calculation of the number of occurrences per century yields already a general image of its frequency evolution:

Table 4: The change of *sentir*'s frequency per 100K words over time

| Century                  | 13  | 15   | 17   | 19   | 21   |
|--------------------------|-----|------|------|------|------|
| Frequency per 100K words | 4.8 | 15.4 | 28.8 | 39.7 | 36.4 |

As shown in Table 4, the general frequency of the verb increases steadily from the 13th century to the 19th century, after which it seems to level off in the contemporary corpus. Zipf's (1949:16) *economic versatility* principle suggests that this frequency increase reflects an increase in its semantic versatility and indeed as shown above in Section 3.2, *sentir*'s relative frequencies of senses have changed from two main semantic domains in Latin – physical and cognitive perception – to three semantic domains in contemporary Spanish, where the emotional meaning even seems to predominate. This explosive growth of the emotional senses is visualized in Figure 4.

Just like the MDS plots presented in the previous section, bubble size reflects relative frequency of the particular emotional senses, but what this plot adds is the diachronic path indicated by means of the arrows. For example, in the *sentirse* cluster on the left, the diachronic evolution of the 'EMO.encontrarse' sense is summarized by four connected points. As in the general MDS plots, this meaning only starts off in the 15th century, at coordinates [-2.8, -0.7]. Then the arrow goes up to [-3.3, -0.2] in the 17th century, towards [-2.9, 0.9] in the 19th century and finally stops at [-2.8, -0.1] in the 21st century and, over this time, its frequency increases dramatically. A similar evolution is observed in the *sentir* cluster on the right for the EMO.experimentar sense, and for the specific EMO.lamentar ('regret') sense. However, in the latter, the evolution looks somewhat different: Starting in the 15th century, this sense first becomes very frequent in the 17th century, after which its frequency drops, but at the same time, it adopts a very extreme position in the plot, a position that reflects more distinct formal and semantic features from the rest, that is, a process of Cxzn.

Indeed, the 21st century dot in the plot basically collects the cases of *lo siento* in the first person singular of the present tense in combination with the clitic *lo*. In other words, between the 17th century (cf. example (24) presented here as (31) and its contemporary use (32), *sentir* in the 'regret' sense has suffered a process of syntactic reduction or, more generally, *deategorialization* by which a prototypical member of the Verb category becomes less prototypical in its distribution (Hopper & Traugott 2003: 106).

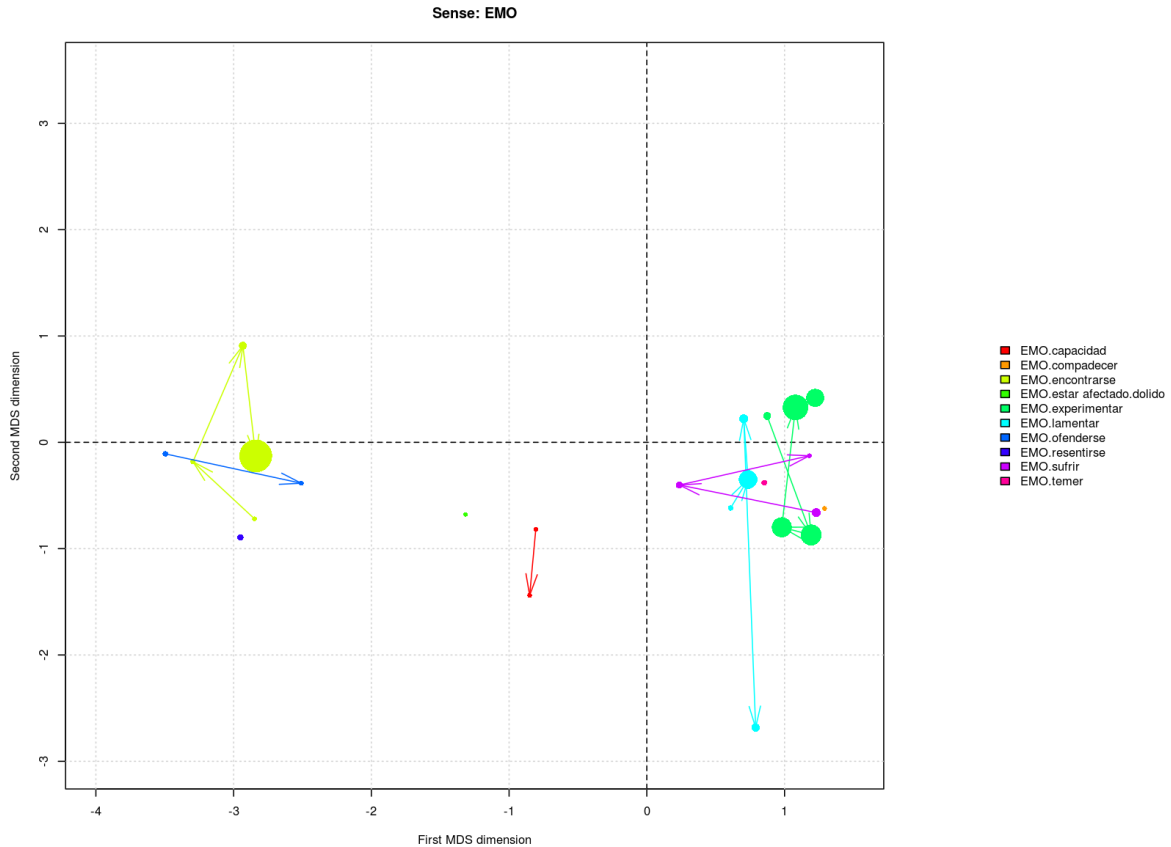


Figure 4: Stationary motion chart of senses in 2-dimensional MDS space

- (31) [...] el rey Felipe Segundo, que **sintió** grandemente la muerte temprana de hijo que, en su tierna edad, daba muestras de gran prudencia. [CORDE: Abarca de Bolea, 1679]  
 ‘King Philip the Second, who extremely *regretted* the early death of his son, who in his tender age, showed signs of great caution’.
- (32) Tengo una cita con su padre. - Oh, no, no... Mi padre ha fallecido. - Perdone. Lo **siento** mucho... Entonces con su marido. [CREA: Press, 2003]  
 ‘I have an appointment with your father. – Oh, no, no... my father passed away. – Excuse me. I’m very *sorry*... Then with your husband’.

More particularly, in its DM function, *sentir* loses some of its prototypical properties as a verb such as its capacity to display variation of tense, mode, aspect, number and person. This loss of some verbal morphosyntactic properties manifests itself in two concrete aspects, (1) the fixation of the form and (2) certain restrictions regarding its argument structure. The fixation of the form is clear: *Sentir* in the DM function appears in the first person singular of the present tense. Concerning its argument structure, a clear restriction appears with regards to the form of its DO: Whereas in the 17th century, apart from the absolute use (4.1%), the verb in this sense combines with NP (37%), *that*-clauses (20.5%), infinitives (12.3%), clitics (15.1%), pronouns and periphrastic relatives (11%), in the contemporary corpus, this sense of the verb is almost exclusively related to the combination with the clitic *lo* (89%) (cf. also Jansegers & Enghels 2013, Jansegers 2017 for a discussion about the morphosyntactic reduction in *lo siento*).

The clitic *lo* in Spanish has been characterized by Hanegreefs (2008: 133-137) and Fernández Jaén (2012: 317-318) as “semantically unstable”, that is, representing complements

whose referent is inaccessible or generalized, characterized by a completely underspecified content, which supplies the sentence with ambiguity. Over time, the semantic ambiguity of *sentir* rises very often in relationship with these kinds of unstable complements that pronominally substitute unspecific, propositional contents:

- (33) que sin dubda, si buen consejo toviésemos, ni oviera tantos males, ni sufriérades tantos malos. E lo más graue que yo **siento**, es que aquella libertad que natura nos dió, e nuestros progenitores ganaron con buen esfuerço, nosotros la avemos perdido e cada ora perdemos con cobardía & caimiento [...]. [CORDE: Pulgar, *CRC*, 1480-1484]  
 ‘Without a doubt, if we had had good advice, there would not have been so many misfortunes, and we would not have suffered so many disasters. And the thing that I *feel* most serious, is that the freedom that nature gave us, and that our ancestors won with a lot of effort, we have lost it, and every hour we lose with cowardice and decline’.

In this example, the speaker expresses a negative valorization with regard to a specific situation or event, which is reported explicitly in the broader context. As mentioned in Section 3.2, this type of evaluative perception – frequently linked to the presence of a predicative complement oriented towards the DO – implies an epistemic shift towards subjectivity. Moreover, particularly in the 15th century, this valorization frequently is related to explicitly negative contexts of pain and suffering, which causes a pronounced ambiguity between two alternative, competing readings: The cognitive reading COGN.considerar (‘to consider, to judge’) or the negative emotional meaning of EMO.lamentar (‘to regret’). That is why this specific construction V + clitic + Compl Pred Obj conveying a negative evaluative perception can be characterized as a source construction for the DM *lo siento*, and not only from the syntactic point of view (the construction with the neuter clitic *lo*, which is the clitic par excellence to refer to situational complements), but also from a semantic point of view: Both the increase of subjectivity incorporating the attitude of the speaker and the emergence in negative contexts yields a considerable ambiguity that results in the competing interpretations of ‘to regret’ or ‘to consider, judge’.<sup>9</sup>

In sum, *lo siento* inherits the collocation with a neuter clitic from the cognitive perception domain. This syntactic relationship is reflected in the MDS plots, where we see that at its first appearance in the 15th century, the ‘regret’ sense of the verb emerges in the midst of cognitive senses. This relationship persists into the contemporary era, where the cognitive senses are situated closest to the ‘regret’ sense despite the extreme position of the latter.

However, in its contemporary use, the explicit emotional meaning of *lo siento* does not fit its cognitive form (the construction with the neuter clitic *lo* referring to situational complements). Therefore, a second, clearly emotional path, must be at the origin of the apologetic DM. This path is clearly visualized by means of the MDS plots: Not only does the emotional meaning of the verb increase in frequency over time, from its first onset, the verb was intrinsically related to a pronounced negative semantic prosody related first to the middle voice *sentirse* cluster, and from the 15th century onwards also in the *sentir* cluster. In this case, the semantic weight lies in the complement of the verb (*dolor* in (34)):

- (34) Tu dolor es el que **siento**, este dobla mi tormento, [CORDE: Manrique, 1474]  
 ‘Your pain is the one that I *feel*, this one doubles my torment’.

In this particular example, it is interesting to notice that *sentir* frequently appears in contexts that point towards a sentiment of compassion, or literally, shared pain. It is clear that

this capacity of feeling someone else's pain on the concrete level, corresponding to a feeling of empathy on the abstract level, is a requirement for the development of an apologetic marker such as *lo siento*.

In sum, then, the contemporary construction *lo siento* can be seen as the result of not only one, but multiple source constructions: On the form side, its input has been basically the cognitive construction with the so-called unstable objects, that is, the construction with the neuter clitic pronoun *lo* ('it'), the clitic par excellence to refer to situational complements. On the meaning side, *lo siento* is the heir of an intrinsically negative prosody related to *sentir(se)* from its origins. In other words, two different paths converge in the rise as a DM 'lo siento': It received both a cognitive and an emotional input, by taking a cognitive form and an emotional meaning.

In this way, the dynamic behavioral profile visualizes the 'Multiple inheritance' of constructions (Goldberg 2003; Trousdale 2013) and the existence of 'Multiple Source Constructions' (cf. De Smet et al. 2013). Indeed, as Van de Velde et al. (2013: 473) state, although most case studies nowadays admit that change does not affect individual lexemes, but entire constructions, they tend to focus on just one construction, "drawing straight lines between a construction and a single historical ancestor". In our case study, we find no such direct lineage from *sentir* towards *lo siento*, but different constructions converging in a change towards the pragmatic marker – the change is not linear, but rather cumulative. In other words, the bird's eye view adopted in the dynamic BP allows us to represent and interpret Cxzn as a holistic process, not as an isolated phenomenon, it allows us to recover a bigger picture within the development of such a particular construction as *lo siento*.

Finally, it is worth noticing that the contemporary corpus suggests an even more recent diversification within *lo siento* itself towards other interpersonal functions: In its use, *lo siento* is not limited to the expression of regret, but also seems to have extended its function towards a kind of adversative marker announcing a negative or contrary opinion to that of the interlocutor, as illustrated in the following example:

- (35) - He tocado en la puerta de su despacho hace un rato, por si había venido y podía dedicarme media hora. Ya le dije que quería hablar con usted más despacio y pedirle consejo. ¿Cuándo podrá? [...].  
- Si quiere, ahora mismo -respondió con apresuramiento.  
- No, **lo siento**, pero no tengo más remedio que irme ya. He telefonado y me están esperando.  
- Bueno, entonces cuando usted quiera. [CREA: Salvador Caja, 2002]  
'I have knocked on your office door a while ago, in case you had arrived and you could devote half an hour to me. I already told you that I wanted to talk to you more slowly and ask your advice. When do you have time? [...] - If you want, right now, he answered in a hurry. -No, *I'm sorry*, but I already have to go now. I have called and they are already waiting for me. -OK, whenever you want then'.

This particular use of *lo siento* can thus be described in terms of (*counter*)*expectation* or *adversativity* (Mortier & Degand 2009: 303), much like some uses of *I'm sorry* or *Excuse me* in English. The adversative interpretation also becomes evident from the frequent collocations with negative adverbs such as *no* or the adversative conjunction *pero* ('but' as in (35)). This suggests thus the existence of at least two types of *lo siento*: In the first type, the speaker expresses his empathy regarding the pain or negative experience of the interlocutor because of some past event, that is, it is based on real facts. On the contrary, when *lo siento* is used with an adversative, counter-expectation function, the speaker anticipates a possible negative experience for the interlocutor (Jansegers 2017). This difference has been formalized



by Aijmer (1996: 99-100) by introducing the distinction between retrospective vs. anticipatory apologies: “the retrospective apology is remedial, supportive (face-saving) and self-demanding [...] when *sorry* has anticipatory function, it can be analyzed as a polite preface or discourse marker with a softening or disarming function”.

Similarly, adopting this function, *lo siento* is used to soften a possible face-threatening action, for example, when the speaker needs to convey some bad news. As such, it has become a practical tool to cancel a meeting or simply to say ‘no’:

- (36) ¿Me da la revista en la que salen las fotos que le han robado a la Obregón? [...] **Lo siento**, se han agotado - dice el quiosquero. [CREA: Press, 2001]  
‘Could you give me the magazine that published the pictures that they stole from Obregón? [...] *I’m sorry*, they are sold out, says the newsagent’.

Note that, contrary to example (35) where the adversative meaning is mainly conveyed by the collocation with the adversative conjunction *pero*, this example demonstrates that *lo siento* can also appear in adversative contexts without the explicit presence of *pero*.<sup>10</sup>

On a more theoretical level, this multifunctional use of *lo siento* may shed some light on the nature of networks in Cxzn alluded to above. Besides the generally recognized vertical inheritance links, the emergence of a new adversative, counter-expectation subschema suggests the need to expand the nature of networks and take into account other kinds of links, such as horizontal networks (Van de Velde 2014; see also Traugott forthcoming for the nature of networks in Cxzn). That way, the adversative DM are horizontally linked with the apologetic DM. This is an interesting line of research that merits future study.

#### 4. Concluding remarks

By means of a corpus-based dynamic BP analysis, this study has examined the diachronic evolution of the polysemy of the Spanish verb *sentir* (‘to feel’), which has led to a number of significant insights situated at both the methodological and the more qualitative, theoretical level of analysis.

Methodologically, our study presents two improvements with regard to previous applications of the BP approach. First of all, it presents the first application of the BP approach to historical data. Whereas the majority of existing studies in quantitative historical semantics have so far focused on the onomasiological axis of semantic change and are based on observed frequencies or percentages of collocational co-occurrence, our study focuses on the semasiological axis of variation and leverages more complex historical data that are based on the similarities of vectors.

Second, rising to one of the major challenges for the diachronic study of meaning change – the visualization of dynamic relationships within and between senses over time – this study also provides an extension of the methodological apparatus of the BP approach by complementing the traditional HAC analysis with a dynamic BP approach derived from MDS maps. Not only has the MDS analysis been proven to be valid since it yields results that are compatible with previous findings, it even seems to go beyond the previous HAC analysis: Whereas HAC potentially suggests misleading discreteness and mutual exclusivity of (elements within) meaning clusters, the use of MDS allows for a visualization of the continuity of the semantic space. In other words, MDS is not only compatible with HAC, but also more precise.

Theoretically, the study also offers a comprehensive perspective on the process of Cxzn. Starting from a broad definition of construction as a form-function pairing (rather than

form-meaning in the strictest sense), Cxzn should be considered a multifaceted process going beyond the (morpho)syntactic level and including discourse phenomena. Based on that idea, we applied the dynamic BP approach to the Cxzn of the DM *lo siento* ('I'm sorry'). By examining the DM within the bigger picture of the changing polysemic profile of the verb over time, the dynamic BP approach offers both a holistic and a more detailed perspective on, and interpretation of, the change. More precisely, our analysis suggests that two different paths converged in the rise of the DM *lo siento*: Instead of a direct lineage from *sentir* towards *lo siento*, different constructions converged in a change towards the DM by taking a cognitive form and an emotional meaning. In addition, in more recent uses, there seems to be more functional diversification within the DM itself, extending from apologetic towards adversative uses. As such, the dynamic behavioral profile sheds some light on the nature of networks in Cxzn: It not only visualizes "multiple inheritance" of constructions (Goldberg 2003; Trousdale 2013) and the existence of multiple source constructions (De Smet et al. 2013), but also suggests the need to expand the nature of networks and take into account other kinds of links, such as horizontal networks (Van de Velde 2014; Traugott forthcoming).

Finally, it should be noted that the present paper provides a more general exploration of how BP vectors give rise to structure in the MDS plots and how those change over time, but it goes without saying that future work of this kind might also choose to focus on a more fine-grained analysis of how particular senses' BP vectors change over time (in a way that is comparable to Divjak and Gries's 2006 analysis of BP vectors using *t* and *z*-scores as a follow-up of their cluster analysis).

## Notes

---

<sup>1</sup> The extreme discrepancy for the number of middle voice uses (*sentirse*) between the 21st century and other periods is striking. There are different explanations for this, methodological as well as linguistic. As for the former, this may be a sampling effect, although it is unlikely that it is due to the quite small number of examples from oral data (which make up not even 7% of the data for the 21st century). As for the latter, we can only speculate at this point that this could point towards a relatively recent grammaticalization process of the pseudo-copular construction. Indeed, according to the *Nueva gramática de la lengua española* (RAE-ASALE 2009: §38.1d), pseudo-copular verbs derive from full lexical verbs through processes of grammaticalization. The grammaticalization process giving rise to pseudo-copular verbs has been called *copularization* (cf. Hengeveld 1992; Stassen 1997), and comes with a series of diagnostics or so-called parameters of grammaticalization (e.g. Lehmann (1995 [1982] and Hopper 1991), one of which is the mechanism of *extension* (or context generalization), i.e. use in new contexts (Heine 2003: 579). A specific type of extension is the distributional extension within the construction itself, or *host-class expansion*, by which the range of collocations is expanded (cf. Himmelmann 2004; Traugott 2011). The quantitative boom of *sentirse* cases in the 21st century reflected in Table 1 could thus be an indication of a fairly recent grammaticalization process: It suggests that, in a first time, the predicative complement slot with *sentirse* preferably combined with a specific lexical element before other elements appeared in this position. Teasing apart what exactly is responsible for that frequency difference merits more detailed analysis, but is beyond the scope of the present paper, since the quantitative methods employed here are based on percentages per verbs.

<sup>2</sup> A list of all ID tags as well as the different senses distinguished and their English paraphrase is provided in Appendix 1.

---

<sup>3</sup> An example of an infinitival DO is illustrated in sentences of the type: [...] *sentimos despedirnos de este gran hombre* [...] (‘[...] , we regret to say goodbye to this great man and painter [...]).

<sup>4</sup> The consulted dictionaries are: el *Diccionario de la Lengua Española* (DRAE), el *Diccionario de Uso del Español* (DUE), el *Diccionario del Español Actual* (DEA) y el *Gran Diccionario de la Lengua Española* (GDLE) for the synchronic data and el *Nuevo tesoro lexicográfico de la lengua española* (NTLLE) for the diachronic data. The translation corpus (approx. 2.5 million words) contains source texts written in a non-Romance language and their translations in Spanish, French, and Italian. The comparable corpus consists of three monolingual corpora in the same languages, where each corpus contains 500 examples of *sentir* randomly selected from literary and press texts. See Enghels & Jansegers 2013 for more details. Although the general semantic categories of physical, cognitive and emotional perception were all found in the standard dictionaries, the translation and comparable corpus study helped to refine these general categories which led to the fine-grained senses distinguished in the ID tags of the present study (cf. Appendix 1).

<sup>5</sup> Cf. Hilpert 2013: 67 for an exhaustive list of MDS applications.

<sup>6</sup> More precisely, from the 22 examples expressing physical experience in the middle voice use in the 13th century, 9 are explicitly negative (referring to physical pain), 5 are positive and 8 are neutral physical perception.

<sup>7</sup> We use the term *semantic prosody* as defined by Louw (1993: 157) as “a consistent aura of meaning with which a form is imbued by its collocates”.

<sup>8</sup> It should be noted that the terminology is not universally agreed upon. Whereas Hilpert (2013: 16) talks about *constructional change* as “a conventionalized form-meaning pair of a language, altering it in terms of its form, its function, any aspect of its frequency, its distribution in the linguistic community, or any combination of these”, Traugott & Trousdale (2013) distinguish between *constructionalization* (the development of form<sub>new</sub>-meaning<sub>new</sub> pairs, i.e. constructions) and *constructional changes*. The latter are understood as changes to features of constructions, such as semantics (e.g. *wif* ‘woman’ > ‘married woman’) or morphophonology (e.g. *had* > ‘d). Such changes precede or follow constructionalization. In the present section, we use the general term *constructionalization* as in Traugott & Dasher and as a synonym for Hilpert’s *constructional change*.

<sup>9</sup> This process of meaning change towards greater subjectivity has been associated with the mechanism of subjectification. Given the topic of this paper, we interpret the concept of *subjectivity* here from a diachronic perspective after the example of Traugott and contrary to Langacker’s synchronic interpretation of the phenomenon, who conceptualizes subjectivity in light of the construal of the event from the speaker’s viewpoint (cf. Langacker 1985; De Smet & Verstraete 2006 for discussion). Traugott (1995: 2, 1999: 3) defines subjectification as a mechanism that causes meanings to “shift toward greater subjectivity [and] become increasingly associated with speaker attitude”. That is, subjectification typically makes explicit the attitude of the speaker. More recently, this semantic-pragmatic change has also been correlated with changing distributional and morphosyntactic patterns (cf. among others Stein & Wright 1995; Torres Cacoullos & Schwenter 2007). For example, the subjectification of verbal DM has been correlated with syntactic reduction (Brinton 2008). This is exactly what we see in the development of the expression *lo siento* where *sentir* loses some of its prototypical morphosyntactic properties as a verb which is reflected both in the fixation of the form (1st person singular, present tense) and the reduction of its argument structure (restricted to the clitic *lo*).

<sup>10</sup> In these cases, *lo siento* functions as an intersubjective pragmatic marker that has absorbed the adversative implication of its collocations. This shows that the semantic change of *sentir*

not only exemplifies the process of subjectification explained above, but also a more recent development towards intersubjectification, i.e. “a mechanism whereby meanings become more deeply centered on the addressee” (Traugott 1999: 3).

### Appendix 1. List of ID tags and ID tag levels (21st century)

|                                | TYPE OF ID TAG                                      | ID TAG   | ID TAG LEVEL  |
|--------------------------------|---|--|---|
| <b>VERB</b>                    | morphosyntactic properties                          | tense  | present vs. past vs. future vs. infinitival form  |
|                                |   | mood   | indicative vs. subjunctive vs. conditional vs. imperative vs. N/A   |
|                                |   | infinitival form   | infinitive vs. gerund vs. N/A   |
|                                |   | person   | 1 vs. 2 vs. 3 vs. N/A   |
|                                |   | number   | singular vs. plural vs. N/A   |
|                                |   | voice  | active vs. passive vs. impersonal vs. middle  |
|                                |   | grammatical aspect   | perfective vs. imperfective   |
|                                |   | presence of causative V  | yes vs. no  |
|                                |   | infinitive   | <i>sentir, sentirse</i>   |
|                                | semantic properties                                 | general sense  | general physical perception vs. specific perception modality vs. cognitive perception vs. emotional perception vs. ambiguous physical P-cognitive P vs. ambiguous physical P-emotional P vs. ambiguous cognitive P-visual P vs. ambiguous synaesthesia vs. ambiguous vs. metalinguistic use |
| specific sense                 |   | 30 levels of ID tags (cf. Table below “Senses as labeled in the dendrogram and their paraphrase in English”) |   |
| <b>ARGUMENT STRUCTURE (AS)</b> |   | basic AS   | full lexical AS vs. weakened AS vs. absolute use  |
|                                | formal characteristics of S                         | lexical presence S   | with vs. without lexical subject  |
|                                |   | form S   | NP vs. proper name vs. pronoun vs. relative pronoun vs. clause vs. N/A  |
|                                | semantic characteristics S                          | semantic role S  | preceptor vs. experiencer vs. stimulus vs. ambiguous vs. N/A  |
|                                |   | semantic type S [nature ±animate; ±concrete]   | animate_human vs. animate_human_collective vs. animate_animal vs. animate_collective vs. inanimate_concrete vs. inanimate_abstract vs. inanimate_body part vs. clause_event vs. clause_state vs. N/A  |
|                                | formal characteristics of DO                        | presence DO  | with vs. without DO   |
|                                |   | form DO  | NP vs. pronoun vs. relative pronoun clitic vs. clause vs. gerund vs. infinitive vs. N/A   |
|                                |   | NP ± determinant   | NP with determinant vs. NP without determinant vs. N/A  |
|                                | semantic characteristics of DO                      | referent DO  | person vs. concrete entity vs. abstract entity vs. situation/event vs. ambiguous vs. N/A  |
|                                |   | type of clitic DO  | clitic of the reflexive passive vs. impersonal clitic vs. reflexive clitic vs. non-reflexive clitic vs. N/A   |
|                                |   | type of non-reflexive clitic   | referential clitic vs. non-referential clitic vs. N/A   |
|                                | characteristics of the predicative complement of DO | presence predicative complement of DO  | with vs. without predicative complement of DO   |
|                                |   | predicative complement of DO: form   | adjective vs. adverb vs. clause introduced by <i>como</i> vs. NP vs. personal pronoun vs. N/A   |
|                                | characteristics of the predicative complement of S  | presence predicative complement of S   | with vs. without predicative complement of S  |
|                                |   | predicative complement of S: form  | adjective vs. adverb vs. clause introduced by <i>como</i> vs. NP vs. personal pronoun vs. clause with gerund vs. clause with infinitive vs. prepositional NP vs. N/A  |

|           |  |   |   |
|-----------|--|---|---|
|           | characteristics of the prepositional object (PO) | presence PO                               | with vs. without PO   |
| ADJUNCTS  | formal characteristics of the adjuncts           | presence of adverbial adjunct             | with vs. without adverbial adjunct                                      |
|           |  | type of adverbial adjunct: form           | adverb vs. PP vs. NP vs. N/A  |
|           | semantic characteristics of the adjuncts         | semantic role adverbial adjunct           | space vs. goal vs. cause time vs. manner vs. quantity vs. agent vs. N/A |
| DISCOURSE | scope  | predicational autonomy                    | yes vs. no  |
|           | collocations                                     | presence adversative/negative conjunction | yes vs. no  |
|           |  | presence vocative                         | yes vs. no  |

### Senses as labeled in the dendrogram and their paraphrase in English

| Spanish label in dendrogram     | English paraphrase   |
|---------------------------------|--|
| EMO.encontrarse                 | be in an emotional state   |
| FIS_GEN.encontrarse.METAF       | be in a general physical state: metaphorical use                 |
| AMBIG.FIS.EMO                   | ambiguous between physical and emotional perception              |
| FIS_GEN.encontrarse             | be in a general physical state                                   |
| COGN.considerar                 | cognitive perception: consider, judge                            |
| EMO.experimentar.METAF          | experience an emotional perception: metaphorical use             |
| EMO.experimentar                | experience an emotional perception                               |
| AMBIG.EMO.COGN                  | ambiguous between emotional and cognitive perception             |
| FIS_GEN.experimentar.corp.METAF | experience a general physical bodily sensation: metaphorical use |
| FIS_ESP.AUD.METAF               | specific physical auditory perception: metaphorical use          |
| FIS_ESP.TACT                    | specific physical tactile perception                             |
| FIS_GEN.experimentar.term       | general physical perception: thermal sensation                   |
| FIS_ESP.AUD                     | specific physical auditory perception                            |
| FIS_GEN.experimentar.corp       | experience a general physical bodily sensation                   |
| AMBIG.FIS                       | ambiguous cases of physical perception                           |
| FIS_ESP.GUST.METAF              | specific physical gustative perception: metaphorical use         |
| FIS_ESP.TACT.METAF              | specific physical tactile perception: metaphorical use           |
| EMO.lamentar                    | emotional perception: regret                                     |
| COGN.presentir                  | cognitive perception: have a presentiment                        |
| COGN.pensar                     | cognitive perception: think                                      |
| COGN.intuir                     | cognitive perception: intuit                                     |
| COGN.crear.opinar               | cognitive perception: believe, opine                             |
| COGN.darsecuenta                | cognitive perception: realize                                    |
| FIS_GEN.manifestarse            | general physical perception: appear, show up                     |
| EMO.capacidad.experimentar      | ability to perceive: emotional perception                        |
| FIS_GEN.capacidad.percibir      | ability to perceive: general physical perception                 |
| AMBIG                           | ambiguous cases  |
| AMBIG.FIS.COGN                  | ambiguous between physical and cognitive perception              |
| COGN.comprender                 | cognitive perception: understand                                 |
| FIS_ESP.OLF                     | specific physical olfactory perception                           |

## References

- Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. New York: Addison Wesley Longman.
- Atkins, B. T. S. (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.
- Atkins, B. T. S. (1992). Tools for computer-aided corpus lexicography: the Hector project. In F. Kiefer, G. Kiss, & J. Pajzs (Eds.), *Papers in Computational Lexicography: Complex '92* (pp.1-60). Budapest: Hungarian Academy of Sciences.
- Barðdal, J., Smirnova, E., Sommerer, L., & Gildea, S. (Eds.). (2015). *Diachronic construction grammar*. Amsterdam: John Benjamins.
- Berez, A. L., & Gries, St. Th. (2009). In defense of corpus-based methods: A behavioral profile analysis of polysemous get in English. In S. Moran, S.T. Darren & M. Scanlon (Eds.), *Proceedings of the 24th Northwest Linguistics Conference. University of Washington Working Papers in Linguistics* (pp. 157-166). Seattle, WA: Department of Linguistics.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- Boas, H. C. (2013). Cognitive construction grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 233-252). New York: Oxford University Press.
- Bolinger, D. (1968). Entailment and the meaning of structures. *Glossa*, 2, 119-127.
- Boogaart, R., Coleman T., & Rutten, G. (2014). *Extending the Scope of Construction Grammar*. Berlin: Mouton De Gruyter.
- Brinton, L. J. (2008). *The Comment Clause in English: Syntactic Origins and Pragmatic Development*. Cambridge: Cambridge University Press.
- Corominas, J., & Pascual, J. A. (1980-1991). *Diccionario crítico etimológico castellano e hispánico*. Madrid: Gredos. [DCELC]
- Croft, W. A., & Timm, J. (2013). Using Optimal Classification for multidimensional scaling analysis of linguistic data.  
<http://www.unm.edu/~wcroft/MDSfiles/MDSforLinguists-UserGuide.pdf>  
(Accessed 10/27/2016).
- Croft, W. A., & Poole, K. T. (2008). Inferring universals from grammatical variation: Multidimensional scaling for typological analysis. *Theoretical Linguistics*, 34(1), 1-37.
- De Smet, H., Ghesquière, L., & Van de Velde, F. (Eds.). (2013). *On multiple source constructions in language change*. Special issue of *Studies in Language*, 37 (3).
- De Smet, H. & Verstraete, J-C. (2006). Coming to terms with subjectivity. *Cognitive Linguistics* 17, 365-392.
- Divjak, D. (2010). *Structuring the lexicon. A clustered model for near-synonymy*. Berlin: Mouton de Gruyter.
- Divjak, D., & Gries, St. Th. (2006). Ways of trying in Russian: Clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory*, 2(1), 23-60.
- Enghels, R. & Jansegers, M. (2013). On the cross-linguistic equivalence of *sentir(e)* in Romance languages: a contrastive study in semantics, *Linguistics* 51(5), 957-991.
- Fernández Jaén, J. (2012). *Semántica cognitiva diacrónica de los verbos de percepción física del español*, PhD dissertation, Alicante: Universidad de Alicante.
- Firth, J.R. (1957). *Papers in Linguistics, 1934-1951*. London: Oxford University Press.

- Giacalone Ramat, A., Mauri, C., & Molinelli, P. (Eds.). (2010). *Synchrony and diachrony: A dynamic interface*. Amsterdam: John Benjamins.
- Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7 (5), 219-224.
- Gries, St.Th. (2006). Corpus-based methods and cognitive semantics: The many senses of *to run*. In St.Th. Gries & A. Stefanowitsch (Eds.), *Corpora in cognitive linguistics: Corpus-based approaches to syntax and lexis* (pp. 57-99). Berlin: Mouton de Gruyter.
- Gries, St.Th. (2010a). Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. *The Mental Lexicon*, 5(3), 323-346.
- Gries, St.Th. (2010b). BehavioralProfiles 1.01. A program for R 2.7.1 and higher.
- Gries, St.Th., & Divjak, D. (2009). Behavioral profiles: A corpus-based approach to cognitive semantic analysis. In V. Evans & S. Pourcel (Eds.), *New directions in cognitive linguistics* (pp. 57-75). Amsterdam: John Benjamins.
- Gries, St.Th., & Divjak, D. (2010). Quantitative approaches in usage-based cognitive semantics: Myths, erroneous assumptions, and a proposal. In D. Glynn & K. Fischer (Eds.), *Quantitative methods in cognitive semantics: Corpus-driven approaches* (pp. 333-353). Berlin: Mouton de Gruyter.
- Gries, St.Th., & Otani, N. (2010). Behavioral profiles: A corpus-based perspective on synonymy and antonymy. *ICAME Journal*, 34, 121-150.
- Hanegreefs, H. (2008). *Los verbos de percepción visual. Un análisis de corpus en un marco cognitivo*. PhD dissertation, Leuven: Katholieke Universiteit Leuven.
- Harris, Z. S. (1954). Distributional Structure. *Word*, 10, 146-162.
- Heine, B. (2003). Grammaticalization. In: B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 575-601). Oxford: Blackwell.
- Hengeveld, K. (1992). *Non-verbal predication: theory, typology, diachrony*. Berlin: Mouton de Gruyter.
- Hilpert, M. (2011). Dynamic visualizations of language change. Motion charts on the basis of bivariate and multivariate data from diachronic corpora. *International Journal of Corpus Linguistics*, 16 (4), 435-461.
- Hilpert, M. (2013). *Constructional change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press.
- Hilpert, M. (2012). Diachronic collostructional analysis: How to use it and how to deal with confounding factors. In K. Allan & J. A. Robinson (Eds.), *Current methods in historical semantics* (pp. 133-160). Berlin: Mouton De Gruyter.
- Hilpert, M. (2016). Change in modal meanings: Another look at the shifting collocates of *may*. *Constructions and Frames*, 8 (1), 66-85.
- Himmelman, N. (2004). Lexicalization and grammaticization: opposite or orthogonal? In: W. Bisang, N. P. Himmelmann & B. Wiemer (Eds.), *What Makes Grammaticalization? A Look from its Fringes and its Components* (pp. 21-42). Berlin: Mouton de Gruyter.
- Hopper, P.J. (1991). On Some Principles of Grammaticization. In E. C. Traugott & B. Heine (Eds.), *Approaches to grammaticalization* (pp. 17-36), Amsterdam: John Benjamins.
- Hopper, P.J., & Traugott, E. C. (2003). *Grammaticalization*. Cambridge: Cambridge University Press.
- James, A. R. (1983). Compromisers in English: a cross-disciplinary approach to their interpersonal significance. *Journal of Pragmatics*, 7 (2), 191-206.
- Jansegers, M. (2017). *Hacia un enfoque múltiple de la polisemia. Un estudio empírico del verbo multimodal 'sentir' desde una perspectiva sincrónica y diacrónica*. Berlin: Mouton De Gruyter.

- Jansegers, M., & Enghels, R. (2013). De verbo de percepción a marcador de disculpa: la evolución diacrónica del verbo *sentir* en español, *Revue de Linguistique Romane*, 305, 139-166.
- Jansegers, Marlies, Vanderschueren, C., & Enghels, R. (2015). The polysemy of the Spanish verb *sentir*: a Behavioral Profile analysis, *Cognitive Linguistics*, 26 (3), 381-421.
- Jucker, A. H., & Ziv, Y. (1998). Discourse Markers: Introduction. In A. H. Jucker & Y. Ziv (Eds.), *Discourse Markers: Descriptions and Theory* (pp. 1-12), Amsterdam: John Benjamins.
- Kövecses, Z. (2008). Metaphor and Emotion. In R. W. Gibbs (Ed.), *The Cambridge handbook of Metaphor and Thought* (pp. 380-396), Cambridge: Cambridge University Press.
- Kurath, H. (1921). *The semantic sources of the words for the emotions in Sanskrit, Greek, Latin, and the Germanic languages*. Chicago: University of Chicago Press.
- Langacker, R.W. (1985) Observations and speculations on subjectivity. In J. Haiman (Ed.), *Iconicity in Syntax* (pp. 109-150), Amsterdam: John Benjamins.
- Lehmann, Christian (1995 [1982]). *Thoughts on grammaticalization*. Munchen: Lincom Europa.
- Levshina, N. (2011). *Doe wat je niet laten kan: A usage-based analysis of Dutch causative construction*, PhD dissertation, Leuven: Katholieke Universiteit Leuven.
- Levshina, N. (2015). *How to do Linguistics with R. Data exploration and statistical analysis*. Amsterdam: John Benjamins.
- Louw, B. (1993). Irony in the text or insincerity in the writer? The diagnostic potential of semantic prosodies, in: M. Baker, G. Francis, E. Tognini-Bonelli, (Eds.), *Text and Technology. In honour of John Sinclair* (pp. 157–177) , Amsterdam: John Benjamins.
- Majid, A., Boster, J. S., & Bowerman, M. (2008). The cross-linguistic categorization of everyday events: A study of cutting and breaking. *Cognition*, 109 (2), 235-250.
- Maldonado, R. (1999). *A media voz. Problemas conceptuales del clítico se*. Mexico D.F.: Instituto de Investigaciones Filológicas, UNAM.
- Moliner, M. (1998). *Diccionario de uso del español*. Madrid: Gredos. [DUE]
- Mortier, L., & Degand, L. (2009). Adversative discourse markers in contrast. The need for a combined corpus approach. *International Journal of Corpus Linguistics*, 14 (3), 301-329.
- Norrick, N.R. (1979). The lexicalization of pragmatic functions. *Linguistics*, 17, 671-685.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1997). *A comprehensive grammar of the English language*. London: Longman.
- Real Academia Española (2001). *Diccionario de la lengua española*, 22<sup>a</sup> ed. Madrid: Espasa. [DRAE]
- Real Academia Española (2001). *Nuevo tesoro lexicográfico de la lengua española*. Madrid: Espasa. Online version: <http://buscon.rae.es/ntlle/SrvltGUILoginNtlle>. [NTLLE]
- Real Academia Española & Asociación de Academias de la Lengua Española (2009). *Nueva gramática de la lengua española. Morfología y sintaxis*. Madrid: Espasa. Online edition: [www.rae.es](http://www.rae.es) [RAE-ASALE].
- Sagi, E., Kaufmann, S., & Clark, B. (2012). Tracing semantic change with Latent Semantic Analysis. In K. Allan & J. Robinson (Eds.), *Current Methods in Historical Semantics* (pp. 161-183). Berlin: Mouton De Gruyter.
- Sánchez, A. (1985). *Diccionario de Uso: Gran Diccionario de la lengua española*. Madrid: Sociedad General Española de Librería. [GDLE]
- Seco, M., Andrés, O. & Ramos, G. (1999). *Diccionario del español actual*. Madrid: Aguilar. [DEA]
- Sinclair, J. (Ed.). (1987). *Collins Cobuild English Language Dictionary*. London: HarperCollins.



- Stassen, L. (1997). *Intransitive Predication*. Oxford: Clarendon Press.
- Stein, D. & Wright, S. (Eds.) (1995). *Subjectivity and Subjectivisation: Linguistic Perspectives*. Cambridge: Cambridge University Press.
- Stenström, A. B. (1995). Some remarks on comment clauses. In B. Aarts & C. F. Meyer (Eds.), *The verb in contemporary English* (pp. 290-301). Cambridge: Cambridge University Press.
- Sweetser, E. (1990). *From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.
- Szmrecsanyi, B. (2011). *The geolinguistics of grammatical variability in traditional British English dialects: A large-scale frequency-based study*. Habilitation thesis, Freiburg: Freiburg University.
- Szmrecsanyi, B., & Kortmann, B. (2009). Vernacular universals and angloversals in a typological perspective. M. Filppula, J. Klemola & H. Paulasto (Eds.), *Vernacular Universals and Language Contacts: Evidence from varieties of English and beyond* (pp. 33-53), London: Routledge.
- Szmrecsanyi, B., & Wolk, C. (2011). Holistic corpus-based dialectology, *Brazilian Journal of Applied Linguistics/Revista Brasileira de Linguística Aplicada*, 11(2), 561-592.
- Torres Cacoullos, R. & Schwenter, S. A. (2007). Towards an operational notion of subjectification. In: R. T. Cover & Y. Kim (Eds.), *Proceedings of the 31st Annual Meeting of the Berkeley Linguistics Society: General Session and Parasession on Prosodic Variation and Change* (pp. 347-358). Berkeley : Berkeley Linguistics Society.
- Traugott, E.A. (1995). The role of the development of discourse markers in a theory of grammaticalization. Paper presented at ICHL XII, Manchester 1995. <http://web.stanford.edu/~traugott/ect-papersonline.html>
- Traugott, E.A. (1999). From subjectification to intersubjectification. Paper presented at the Workshop on Historical Pragmatics, Fourteenth International Conference on Historical Linguistics, Vancouver, July 1999.
- Traugott, E. C. (2011). Grammaticalization and mechanisms of change. In: H. Narrog & B. Heine (Eds.), *The Oxford Handbook of Grammaticalization* (pp. 19-30). Oxford: Oxford University Press.
- Traugott, E.C. (2014). Toward a constructional framework for research on language change. *Cognitive Linguistic Studies*, 1, 3-21.
- Traugott, E.C. (Forthc). Modeling language change with constructional networks.
- Traugott, E.C., & Trousdale, G. (2013). *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Trousdale, G. (2013). Multiple inheritance and constructional change. *Studies in Language*, 37 (3), 491-514.
- Van de Velde, F. (2014). Degeneracy: The maintenance of constructional networks. In R. Boogaart, T. Coleman & G. Rutten (Eds.), *Extending the Scope of Construction Grammar* (pp. 141-179). Berlin: Mouton De Gruyter.
- Van de Velde, F., De Smet, H. & Ghesquière, L. (2013). On multiple source constructions in language change. *Studies in Language*, 37 (3), 473-489.
- Wheeler, E. S. (2005). Multidimensional scaling for linguistics. In R. Koehler, G. Altmann, & R.G. Piotrowski (Eds.), *Quantitative Linguistics: An international handbook* (pp. 548-553). Berlin: Mouton De Gruyter.
- Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort*. Cambridge, M.A: Addison-Wesley Press.