

ÁDÁM GALAC

Eötvös Loránd University, Intercultural Linguistics Doctoral Programme
adam.galac@gmail.com

Ádám Galac: Basic-level multimodal perception verbs in French, Spanish, and Hungarian: a contrastive corpus study of Fr. *sentir*, Sp. *sentir*, and H. *érez*
Alkalmazott Nyelvtudomány, XXI. évfolyam, 2021/2. szám, 62–79.
doi:http://dx.doi.org/10.18460/ANY.2021.2.005

Basic-level multimodal perception verbs in French, Spanish, and Hungarian: a contrastive corpus study of Fr. *sentir*, Sp. *sentir*, and H. *érez*¹

This paper presents a corpus-based contrastive analysis of French *sentir*, Spanish *sentir*, and Hungarian *érez* ‘feel’. These basic-level multimodal perception verbs (i.e. verbs encoding multiple sensory modalities) have many semantic features in common, but they operate in different linguistic and sociocultural contexts, and Hungarian *érez* differs from its Romance counterparts concerning its etymology as well. Based on 500–500 random concordance lines extracted from the TenTen corpora, I have examined the verbs’ semantic and formal properties, taking into account not only their primary contextual meanings but also features such as epistemicity, figurativity, the degree of grammaticalization, and the constructions they occur in. By doing so, I have extended the seminal study of Enghels and Jansegers (2013) – who looked into the degree of equivalence between French *sentir*, Spanish *sentir*, and Italian *sentire* – to a genealogically and typologically different language and to novel aspects of analysis. The quantitative results highlight important differences in usage and also point to promising future perspectives in the investigation of multimodal perception verbs.

Keywords: *sentir*, *érez*, multimodal perception verbs, contrastive corpus study

1. Introduction

This paper presents a contrastive examination of three highly polysemous basic-level perception verbs: French *sentir*, Spanish *sentir*, and Hungarian *érez*. These three verbs, which are perhaps best rendered as ‘feel’ in English, can be considered as equivalents of one another in many contexts: not only can they refer to (external) tactile, olfactory and gustatory sensory perception, but they are crucial in the linguistic conceptualization of internal sensations (interoception, proprioception), emotions, and certain cognitive states as well. In this way, they can be regarded as the multimodal¹ perception verbs par excellence in the corresponding languages (cf. Fernández Jaén, 2012: 472), constituting a kind of categorization that does not conform to the five senses folk model (cf. Winter, 2019: 11–15), but instead highlights the interrelationships between perception, bodily sensations, emotion, and cognition.

¹ This research was funded by the ÚNKP-20-3 New National Excellence Program of the Hungarian Ministry for Innovation and Technology.

A corpus-based semantic analysis of French *sentir*, Spanish *sentir*, and Italian *sentire* was carried out by Enghels and Jansegers (2013), who looked at the degree of equivalence of the three cognates in parallel translation corpora, on the one hand, and in comparable monolingual corpora, on the other. On the basis of four main semantic categories – general physical perception, a specific modality of perception (hearing, taste, vision, touch, smell), emotional perception, and cognitive perception (Enghels & Jansegers, 2013: 975) –, they found important differences in the proportion of the different semantic clusters in the three languages. Specifically, Italian *sentire* is more distant from its French and Spanish counterparts due to its dominant auditory meaning, French *sentir* is more frequently used in the context of cognitive perception and knowledge, and Spanish *sentir* refers more typically to emotional perception (Enghels & Jansegers, 2013: 974; 986). They also point out that the boundaries between the semantic categories are fuzzy, as there are many ambiguous and creative contextual uses that cannot be classed incontestably (Enghels & Jansegers, 2013: 985).

The present analysis extends this investigation to Hungarian *érez*, a multimodal perception verb exhibiting a similar semantic structure but embedded in a genealogically and typologically different language, and another sociocultural environment. Yet while Enghels and Jansegers mainly focus on the degree of equivalence between three cognate verbs, the primary aim of this study consists in exploring the behavior of basic-level multimodal perception verbs in different linguistic and cultural environments. To this end, it also widens the scope of analysis, taking into consideration various semantic and formal aspects besides the basic contextual meanings. These are presented in more detail in Sections 3 and 4.

The structure of the paper is as follows: Section 2 gives an overview of the theoretical backdrop against which this examination has been carried out. Section 3 describes the data sources and the principles of annotation along with the limitations of such an investigation. Section 4 presents the results and looks for their possible motivations. Finally, Section 5 concludes the paper with some final remarks and future research possibilities.

2. Theoretical background

2.1 Embodiment, perception, and cultural cognition

In Western culture, the dominant way of modeling human beings was for a considerable time characterized by a strict distinction between the body and the mind, conceiving these two as fundamentally separate entities that belong to different levels of existence. However, at the end of the 20th century, an increasing number of findings have begun to challenge this view, leading to a model that emphasizes the embodied nature of human cognition.² In particular, behavioral experiments, brain imaging, and neuropsychology have accumulated evidence

that our sensorimotor and perceptual systems shape our cognitive processes not only indirectly through our sensory experiences but also directly through shared mechanisms and simulation (cf. Bergen, 2012; Bergen, 2015). As a consequence, human reasoning and language inevitably emerge from the body (cf. Lakoff & Johnson, 1999).

This theoretical perspective assigns sensory linguistics a more central role than earlier approaches, as the examination of perceptual language can yield valuable insights into the interrelationship between bodily sensations and mental processes (the latter either unconsciously emerging from the former or encompassing a somewhat less unconscious cultural conceptualization of it). In this context, multimodal perception verbs like *sentir* and *érez* constitute an especially interesting case, since they class a great variety of bodily, emotional and cognitive experiences into one linguistic category and thus reflect a similar cultural conceptualization of these experiences.

For linguistic conceptualizations are cultural constructs, and an analysis like the present one must take into account the fact that cognition and language are shaped by the cultural context they are embedded in. Indeed, as Sharifian (2008: 116) points out, cognition can be viewed as the property of cultural groups, and language and conceptualization “emerge at the cultural level of cognition”. Consequently, culture provides an important counterpoint to body-based effects and cannot be excluded from semantic analyses. The rich polysemy of the studied verbs is in fact a culturally provided basis for a similar conceptualization of the wide range of sensations they can refer to.

2.2 Semantic premises

In compliance with the views of contemporary (cognitive) semantics, the present analysis regards linguistic meaning as consisting of dynamic categories based on prototypes, i.e. central senses that can be extended to various contextual readings in a flexible way (cf. Rosch, 1978; Geeraerts, 2010: 182–192). If certain clusters of contextual readings become conventionalized, they may form local prototypes besides the original global one, leading to a complex hierarchical semantic structure (cf. Langacker, 1990: 266; Györi, 2002: 151–152). These local prototypes may become distinct enough to yield an obvious case of polysemy,³ but very often there is a continuum between them that makes it extremely difficult to decide whether the lexical item should be interpreted as polysemous or semantically vague (cf. Geeraerts, 2010: 196–199). While the former is emphasized by conceptual network models such as the radial network model (Brugman, 1988, originally 1981; Lakoff, 1987; Lemmens, 2016), the latter has been described through metaphors like “a seamless fabric of meaning potential” (Cruse, 1982: 79; cited after Geeraerts, 2010: 199) or “a fog” (Geeraerts, 2015: 244) with the areas around the prototypes thicker than those in between.

The dynamic model of semantic structure also suggests that semantic change forms an integral part of a linguistic item's meaning potential and cannot be strictly separated from it. What really exists is continuous change, and synchronic analyses and dictionaries only provide snapshots of a process which, according to Győri (2002: 150; 160), can be broken down into three main stages: pragmatic ambiguity, polysemy, and changed meaning. A crucial factor here is again culture (or cultural cognition), which is responsible for the selection and conventionalization of particular (contextual) instances of change – instances that the given language community finds relevant and useful enough to preserve (cf. Croft, 2000; Győri, 2002). As Sharifian (2008: 122–123) puts it,

Inherent within the system of every language are categories, schemas, conceptual metaphors and propensities for certain perspectives that reflect cultural cognitions of those who have spoken the language over the history of its existence.

In light of the above, it seems to be worth looking at the etymologies of the studied verbs, which turn out to be surprisingly different in spite of the synchronic similarities regarding their meaning potential.

The Classical Latin etymon of the Romance cognates, *sentīō*, used to refer to general physical and cognitive perception (cf. Hertegonne, 2014: 23–44; Jansegers & Gries, 2020: 148), but it did not express emotional content. Its closest out-of-context translations in English are perhaps ‘perceive’, ‘experience’, ‘notice’, ‘understand’, ‘think’, ‘deem’, ‘judge’ (the opinion-sense emerging as a consequence of mental perception).⁴ However, the restriction of its sensory scope (the exclusion of visual and in French also of auditory meanings), the emergence of an emotional local prototype, and the profiling of particular domains (e.g. the olfactory domain in French or the emotional one in Spanish) have led to a reinterpretation of the semantic structure of this verb and to the modification of its global prototype (see also Galac, 2020: 140–141). It is not the aim of the present study to give a detailed overview of the history of *sentīō* from Latin to Romance (readers are referred to Fernández Jaén, 2012; Hertegonne, 2014; Jansegers, 2017; Jansegers & Gries, 2020), but one final point should be mentioned as it will be of high relevance for this analysis: the disappearance of the Latin synthetic passive voice and the development of reflexive verbs in the Romance languages (cf. Hertegonne, 2014: 8–9).

In contrast, Hungarian *érez* has probably developed its current polysemy from a single tactile meaning, related to the verb *ér* ‘touch; reach, arrive’ (TESz., 1: 785–786; EWUng., 1: 330). This is not without parallels, as touch often functions as a source domain for both general sense perception and emotional perception (Sweetser, 1990: 37–38). Unfortunately, early Hungarian data is very scarce, and the first attestation of the word is already a semantically complex one: in the Old Hungarian Lamentations of Mary (Ómagyar Mária-siralom, the oldest existing Hungarian poem, dating from around 1300), it refers to pain caused by

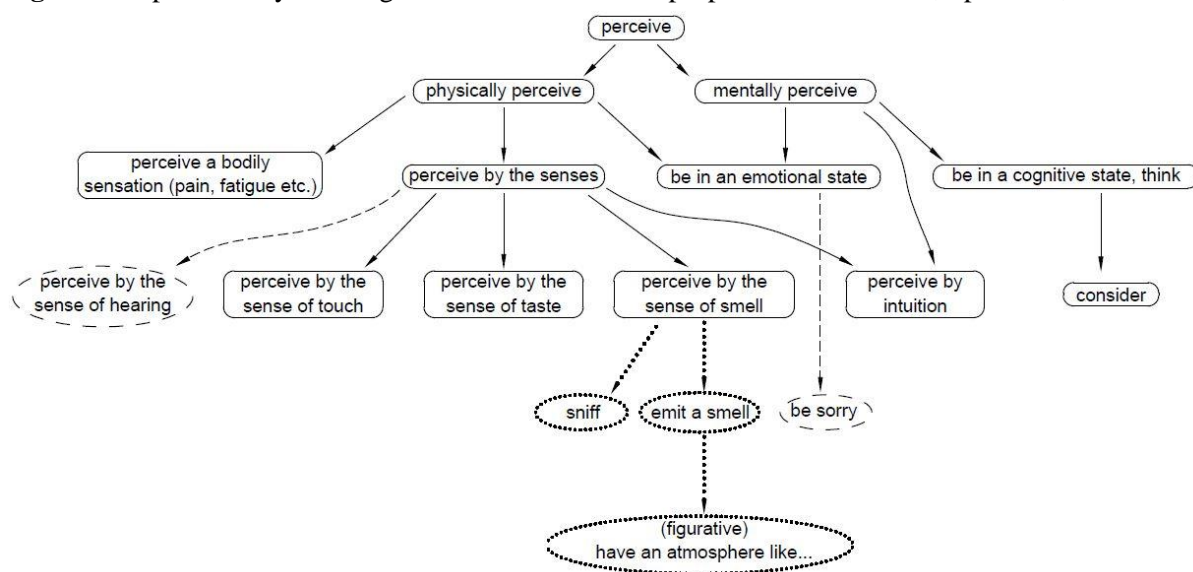
“the dagger of sorrow”, suggesting that the verb could already refer to both physical and emotional perception, thus allowing the use of this metaphor (1).

- (1) *en erz-em ez bu-thur-u-th*
 Glossing: I feel-1SG this sorrow-dagger-POSS.3SG-ACC
 Translation: ‘I feel this dagger of sorrow’
 Source: <http://magyar-irodalom.elte.hu/gepesk/kkor/028.htm>
 (22.6.2021)

In sum, the studied verbs have followed different paths to arrive at the same relative position in different contexts: they have become the primary basic-level multimodal perception verbs in the corresponding languages, with considerable overlap in their semantic structure and communicative functions. The following sections present a contrastive synchronic corpus-study with the aim of drawing a more detailed picture of the linguistic behavior of each verb – but before moving on to this empirical examination, let us look at a preliminary overview of the possible senses of the three verbs, given in Figure 1.

Of course, such an overview is only a simplification – both the theoretical considerations discussed above and the great variation in the categorizations of different dictionaries suggest that we are dealing with a more complex issue than what can be depicted in a two-dimensional graph. Still, it seems useful to have a preliminary working model showing the connections between the main domains of experience that are coded by the examined verbs. The bubbles drawn with solid lines contain the senses (perhaps prototypes of different levels) that are common to all three verbs. The dotted lines stand for meanings found only in the case of Fr. *sentir* (olfactory ones), while the dashed lines mark what is particular to Sp. *sentir* (auditory uses and the discourse marker *lo siento* ‘I am sorry’). Hungarian *érez* does not seem to have a clearly distinguishable sense that is lacking in its Romance counterparts. The arrows point from more general to more specific domains, thus drawing a kind of taxonomy of the domains featured in the graph. Yet it must be emphasized again that this overview is only a starting point for a more in-depth analysis, and it does not reflect the cognitive architecture of the verbs’ semantic features, nor their diachronic development (which is summarized above).

Figure 1. A preliminary working model of the semantic properties of Fr. *sentir*, Sp. *sentir*, and H. *érez*



3. Methodology

3.1 The data

The present analysis is based on data extracted from the TenTen Corpus Family, a family of large general-purpose corpora in more than forty languages, all with a size of several billion words. They are collected automatically from a large number of web sources filtered for “linguistically valuable content” “according to the same criteria and can be regarded as comparable corpora” (<https://www.sketchengine.eu/documentation/tenten-corpora/>, 5.3.2021). These features make them ideal sources for the cross-linguistic investigation of everyday language use, even if we have to keep in mind that they consist only of written texts created (mostly) in digital contexts. An overview specifying the size of the examined corpora and the total number of occurrences of Fr. *sentir*, Sp. *sentir* and H. *érez* is given in Table 1.

Table 1. Corpora used for the analysis

Language	Name of the corpus	Total number of words	Number of hits for the lemma <i>sentir</i> / <i>érez</i>
French	French web corpus 2017 (frTenTen17)	5.7 billion words	815,206 (119.08 per million tokens)
Spanish	Spanish web corpus 2018 (esTenTen18)	17.5 billion words	4,899,563 (241.28 per million tokens)
Hungarian	Hungarian web corpus 2012 (huTenTen12)	2.5 billion words	921,326 (291.38 per million tokens)

3.2 The annotation

The empirical examination consisted in the manual analysis of 500 randomly generated concordances (occurrences in context) of each verb. Although I had some main ideas about what tags to use (cf. the semantic categories in Figure 1), I adopted a bottom-up approach and started the annotation with an open mind, looking for the tags that best describe the data and that cover phenomena on various levels of analysis. Thus, the first part of the work was characterized by a circular process with continuous revision and reanalysis that finally resulted in the following descriptive system.

The semantic properties of an occurrence in context are described along three main axes: primary meaning component, additional meaning, and degree of grammaticalization.

The primary meaning components broadly correspond to the bubbles of Figure 1 – for the most part, they refer to a sensory domain that is activated during the perceptual experience. There are separate tags for each modality of external sensory perception, one for thermoception (perceiving temperature), and a further one called “general external perception” for cases where no specific sensory medium is implied (2). Internal physical sensations are classed as “pain”, “physical”, and “general physical”, the latter two distinguishing between particular bodily sensations (e.g. itching or feeling one’s muscles, cf. 3) and physical states affecting the whole body (e.g. feeling tired or ill). Leaving the realm of physical experiences, the label “general mental” was used for cases that mingle emotion and cognition in an inseparable way (4), while “emotional” and “cognitive” refer to instances where one of these components is clearly the dominant one (5–6). Contextually vague cases (without a direct object or any other specification) are marked as “general_absolute” (7). Finally, there are extensions that incorporate different perspectives on the perceptual event: agentive and percept olfactory uses of Fr. *sentir*, on the one hand,⁵ and the Spanish discourse marker *lo siento* ‘I am sorry’, on the other.

- (2) *Ha érezzük, hogy csúszunk, akkor engedjük el kicsit a féket,
hogy újra forogjanak a kerekek, aztán újra próbáljunk fékezni.*

Translation: ‘If you feel you are slipping, release the brake slightly to get the wheels turning again, then try braking again.’

Source: huTenTen12 (wita.hu)

- (3) *j’ai senti mes abdominaux travailler en profondeur*

Translation: ‘I felt my abdominal muscles working in depth’

Source: frTenTen17 (marieclaire.fr)

- (4) *Minden egyes nap, mikor dolgozni mentem, csodásan éreztem magam.*

Translation: ‘Every single day when I went to work, I felt wonderful.’

Source: huTenTen12 (vasarnapihirek.hu)

- (5) *sentí como la culpa me carcomía por dentro*

Translation: ‘I felt the guilt gnawing at my insides’

Source: esTenTen18 (activoforo.com)

- (6) *Les gens sentaient que le parti socialiste n’était pas derrière moi, et du coup l’opinion m’aidait.*

Translation: ‘People felt that the socialist party was not behind me, and suddenly the public opinion was helping me.’

Source: frTenTen17 (segorama.fr)

- (7) *la fluidez con que las imágenes, los símbolos, las formas de pensar y sentir circularon entre la cultura de las elites y la de las clases populares*

Translation: ‘the fluidity with which images, symbols, and ways of thinking and feeling were circulating between the culture of the elites and that of the popular classes’

Source: esTenTen18 (caia.org.ar)

Besides these primary meanings, two additional components seem to be highly relevant for Fr. *sentir*, Sp. *sentir*, and H. *érez*. The first one, labeled as “consider”, marks that the perceptual experience is paired with some kind of judgment or opinion on the part of the perceiver (8), while the second one, tagged as “epistemic”, refers to cases where the perceptual experience conveys some further information that goes beyond it in scope and complexity (9). Interestingly, these two components are often difficult to separate, because many contexts do not make it clear whether the additional information comes from the perceiver in the form of a subjective opinion or judgment, or from the external world as a piece of further information that is made accessible by the perceptual experience – these ambiguous cases were marked as “consider/epistemic” (10). A third type of additional meaning – actually a specific subcategory of epistemicity – is labeled as “intuition” and refers to inexplicable hunches (11), while a fourth one is connected with figurativity and is most typical of the percept olfactory uses of Fr. *sentir* (12).

(8) *Ahora tenía que dejar el espacio necesario para que éste fuera quien contara lo que sintiera necesidad.*

Translation: ‘Then I had to leave the necessary space for him to be the one to tell what he deemed necessary.’

Source: esTenTen18 (expectopatronum-rpg.com)

(9) *Votre compagne ne peut que se sentir aimée...*

Translation: ‘Your partner cannot but feel that you love her...’

Source: frTenTen17 (online.fr)

(10) *A l'inverse, 36% des personnes interrogées ne se sentent que rarement ou jamais en insécurité dans les transports en commun d'Ile-de-France.*

Translation: ‘Conversely, 36% of respondents rarely or never feel unsafe on public transportation in the Paris region.’

Source: frTenTen17 (ifop.fr)

(11) *Nevetgélve és jókedvűen érkeztek, ám mikor megpillantották az ácsorgó négyest, érezték, hogy nincs rendben valami.*

Translation: ‘They arrived laughing and in a good mood, but when they saw the group of four standing there, they got a feeling that something was not right.’

Source: huTenTen12 (csodaidok.hu)

(12) *Ça sentait déjà le vieux Zemmour rance dans la façon dont les médias accueillait le nouveau livre de Houellebecq.*

Translation: ‘It already smelled like rancid old Zemmour in the way the media welcomed Houellebecq's new book.’

Source: frTenTen17 (c-g-a.org)

The third semantic dimension concerns the degree of grammaticalization of the verbs in question, as they sometimes appear in semantically bleached positions. This is particularly true of reflexive forms, which can function as simple copulas (compare the examples in Table 2 and their reformulations with the corresponding equivalent of ‘to be’), but non-reflexive verbs can also behave as light verbs, especially in prefabricated expressions (e.g. Sp. *sentir celos* ‘to be jealous’, H. *szükségesnek érez valamit* ‘to deem something necessary’).

Table 2. Examples for copulative uses of Fr. *sentir*, Sp. *sentir*, and H. *érez*

Reflexive forms used as copulas	Reformulations with ‘to be’
Fr. <i>se sentir à l’aise</i> ‘to feel comfortable, at ease, relaxed’	<i>être à l’aise</i> ‘to be at ease, relaxed’
Sp. <i>me sentiría muy cabreado</i> ‘I would feel furious’	<i>estaría muy cabreado</i> ‘I would be furious’
H. <i>nyűgösnek éreztem magam</i> ‘I felt cranky/irritable’	<i>nyűgös voltam</i> ‘I was cranky/irritable’

In addition to these semantic properties, I also attempted to capture the syntactic behavior of the studied verbs by looking at the constructions they appear in. I annotated the following: reflexive uses; verbal complements (direct object, adjective, adverb, subordinate clause etc.); causative constructions (13); object-oriented conceptualization (14); infinitive construction expressing potentiality (15). Moreover, I tagged some recurring contexts that might be considered characteristic of the studied verbs: for instance, constructions meaning ‘to feel good/bad’ (cf. example 4), or uses referring to the felt identity of the subject (e.g. *se sentir femme* ‘to feel like a woman’, *sentirse cubano* ‘to feel Cuban’).

- (13) *Tratando de capturar y retener los preciosos momentos que le hacen sentir conectado y amado.*

Translation: ‘Trying to capture and retain the precious moments that make him feel connected and loved.’

Source: esTenTen18 (ucg.org)

- (14) *Es una sensación única, difícil de explicar, pero se siente increíble.*

Translation: ‘It is a unique feeling, difficult to explain, but it feels incredible.’

Source: esTenTen18 (wapa.pe)

- (15) *140-ig jók vagyunk, 150 körül már érezni a lég ellenállását, 160 fölött már hangos*

Translation: ‘Up to 140 we are good, around 150 you can feel the air resistance, above 160 it gets loud’

Source: huTenTen12 (totalcar.hu)

3.3 Limitations

Like every scientific endeavor, this study also has its limitations that should be taken into consideration in order to get a clear picture of the reliability of its outcomes. First, due to the fact that semantic landscapes are continuous and the

boundaries between the prototypical senses are often fuzzy, it was not always evident which tag should be used. An especially thorny issue was the part of the fog (following Geeraerts' metaphor, see 2.2) around and between the prototypes "general external perception" and "general mental perception": while the former refers to the perception of something in the outside world through an unspecified sensory medium, the latter focuses on the inner mental activity of the perceiver. Yet the two are often combined with one another to yield a kind of complex intuitive perception of an external fact through mental processing (cf. example 11). In these cases, I tried to determine which of the two is the dominant one, and included the other one as a secondary component.

A second caveat issues from the fact that the data was processed by one researcher alone, which aggravates the degree of subjectivity of the results. It would be desirable to verify these in a research group to reduce the amount of subjective decisions as much as possible.

And third, it would also be desirable to examine a larger sample with the same methodology, as well as samples extracted from other types of corpora, in order to obtain more precise data and uncover corpus- and genre-specific differences.

4. Results

Table 3 gives a general overview of the number of occurrences found for each primary contextual meaning (in token value and percentage). As the corpus query also yielded some false results (28 for French, 33 for Spanish, and 4 for Hungarian), the percentages are calculated by the number of occurrences that do contain what I intended to search for, i.e. verbal forms of Fr. *sentir*, Sp. *sentir*, and H. *érez*.

Table 3. Contextual meanings of Fr. *sentir*, Sp. *sentir*, and H. *érez*

Contextual meaning (primary component)	Fr. <i>sentir</i>		Sp. <i>sentir</i>		H. <i>érez</i>	
	tokens	percent	tokens	percent	tokens	percent
general internal physical perception	16	3.4%	17	3.6%	22	4.4%
specific internal physical perception	27	5.7%	32	6.9%	35	7.1%
pain	2	0.4%	6	1.3%	5	1.0%
temperature	3	0.6%	4	0.9%	3	0.6%
touch	11	2.3%	9	1.9%	17	3.4%
taste	4	0.8%	2	0.4%	3	0.6%
smell	50	10.6%	2	0.4%	8	1.6%
hearing	-	-	9	1.9%	-	-
general external perception	53	11.2%	35	7.5%	43	8.7%
general mental perception	247	52.3%	220	47.1%	235	47.4%
cognitive	35	7.4%	24	5.1%	73	14.7%
emotional	23	4.9%	95	20.3%	49	9.9%
be sorry	-	-	4	0.9%	-	-
general absolute	1	0.2%	8	1.7%	3	0.6%
TOTAL	472	100.0%	467	100.0%	496	100.0%

As it can be seen, most domains are present in similar proportions, but one can spot major differences as well. It is important to note, though, that the reason for these differences does not necessarily reside in the semantic structure of the examined verbs – it may also originate from cultural factors, as some domains may be verbalized more frequently in certain cultural environments than in others.

First, the olfactory domain is much more prevalent in the case of Fr. *sentir*: it is found over six times more frequently than for H. *érez*, and over 26 times more frequently than for Sp. *sentir*. This is primarily due to the fact that Fr. *sentir* has several olfactory senses, some of which go beyond the standard event structure of *sentir* and *érez*. Besides the so-called experiencer type of conceptualization, referring to pure perceptual experiences (e.g. English *to hear*), it can also encode the agentive type, a similar perspective but with a voluntary perceiver (e.g. English *to listen*), as well as the percept one, where the subject of the verb is the object of perception (e.g. English *to sound*; cf. Gisborne, 2010: 4–8). Moreover, all three meanings of Fr. *sentir* can be used figuratively (see Table 4).

Table 4. Olfactory meanings of Fr. *sentir*

experiencer		agentive		percept	
literal	figurative	literal	figurative	literal	figurative
11	2	1	1	19	16

Second, references to tactile perception seem to be slightly more typical of H. *érez*: though this may be connected to its etymology (the verb’s original meaning was probably a tactile one, see 2.2), discourse features may also play a role in its motivation, since a considerable portion of its occurrences (7 out of 17) occur in pornographic contexts.⁶

Third, Sp. *sentir* is used much more frequently for the verbalization of emotions than its French and Hungarian counterparts (over 20.3% as opposed to 4.9% for Fr. *sentir* and 9.9% for H. *érez*). This marked emotional sense was also found by Jansegers and Gries (2020: 148) as “the most frequent sense of the verb in present-day Spanish” and led Jansegers (2017: 142–144) to conclude that it is the most prototypical sense of Sp. *sentir* (cf. also its derivation ‘be sorry for something’, which is much more dominant in spoken language than suggested by our 4 occurrences found in written contexts).

And fourth, the cognitive dimension stands out in Hungarian (14.7%), much more than in French (7.4% as opposed to Spanish *sentir*’s 5.1%), although Enghels and Jansegers (2013: 986) observed that Fr. *sentir* is more typically a cognitive verb than Sp. *sentir* or It. *sentire*:

French *sentir* most dominantly – but certainly not exclusively – covers the field of cognitive (but often intuitive) perception, meaning ‘to think’ or sometimes even ‘to know’.

However, the additional meaning components (shown in Table 5) support this observation as Fr. *sentir* has the highest percentage of epistemic uses (27.8%) and references to intuitive perception (7.0%), suggesting that it has a strong connotation of gaining knowledge. It is closely followed by H. *érez*, while Sp. *sentir* has somewhat lower proportions in this respect (especially concerning intuition). On the other hand, the component “consider” is much more present in Spanish and Hungarian than in French. As to figurativity, we get similar proportions in the three languages if we subtract the figurative olfactory meanings of Fr. *sentir* (19 instances altogether, see Table 4).

Table 5. Additional meaning components of Fr. *sentir*, Sp. *sentir*, and H. *érez*

Additional meaning component	Fr. <i>sentir</i>		Sp. <i>sentir</i>		H. <i>érez</i>	
	tokens	percent	tokens	percent	tokens	percent
consider	28	5.9%	47	10.1%	55	11.1%
epistemic	131	27.8%	72	15.4%	113	22.8%
consider / epistemic	108	22.9%	75	16.1%	103	20.8%
intuition	33	7.0%	7	1.5%	27	5.4%
figurative	25	5.3%	4	0.9%	3	0.6%

Table 6 shows the extent to which Fr. *sentir*, Sp. *sentir*, H. *érez* and their corresponding reflexive forms are used in (partially) grammaticalized ways. Non-

reflexive forms can function as light verbs (predominantly in Spanish and Hungarian, not so typically in French), while reflexive ones can be used as copulas (above all in Spanish and French): the most cases are found in Spanish (135 instances or 28.9% altogether), implying that Sp. *sentir* might be the most grammaticalized of the three verbs. Fernández Jaén, who notes that the sense ‘be sorry’ is the most grammaticalized one in Spanish (Fernández Jaén, 2012: 457–458), also points out that in phrases like *Raquel se siente sola* ‘Raquel feels alone’, “it is evident that *sentirse* has undergone an almost complete semantic bleaching” (Fernández Jaén, 2012: 441).⁷ The examined data abound in such or even more obvious cases, like *sentirse atraído por algo/alguien* ‘to be attracted by sg/to sy’ or *sentirse orgulloso* ‘to be proud’.

Table 6. Grammaticalized uses of Fr. *sentir*, Sp. *sentir*, and H. *érez*

Partially grammaticalized uses	Fr. <i>sentir</i>		Sp. <i>sentir</i>		H. <i>érez</i>	
	tokens	percent	tokens	percent	tokens	percent
light verb (non-reflexive)	8	1.7%	58	12.4%	35	7.1%
copula (reflexive)	56	11.9%	77	16.5%	8	1.6%
TOTAL	64	13.6%	135	28.9%	43	8.7%

Finally, Table 7 displays the most salient constructional and contextual differences that were found in connection with the three verbs. Reflexive uses account for broadly half of the instances in French and Spanish, while their proportion is only 26.8% in Hungarian. On the other hand, H. *érez* introduces much more subordinate clauses than its Romance counterparts (35.7% as opposed to 10.6% in French and 13.9% in Spanish) – and some of these subordinations are further nuanced by the cataphoric markers *úgy* and *azt*, which add a degree of subjectivity to the typically epistemic meaning (*érzi, hogy* ↔ *úgy érzi, hogy; azt érzi, hogy* ‘feel that’). A third interesting difference concerns what could be called indefinite objects – demonstrative or interrogative pronouns that only hint at the feeling in question without further detail –, the use of which seems to be much more restricted in French than in Spanish and Hungarian.

Three constructions are completely absent either in the Hungarian or the Romance data. The first one is causation: while French and Spanish have semantically flexible analytic constructions to express causation (Fr. *faire sentir*, Sp. *hacer sentir* ‘to make feel’), Hungarian does not. There is a synthetic form derived with the help of the causative morpheme *-tat/-tet* – *éreztet valamit valakivel* –, but its meaning is restricted to the mental domain (‘to imply something to someone’), and it is not found by the corpus query because it is considered as a separate lexeme. The second one, labelled as object-oriented (or percept) conceptualization, is connected with a special use of reflexive forms in Romance: it puts the object of perception in the role of the grammatical subject and adds an element of potentiality to the meaning of the verb. In French and

Spanish, it can be formulated either as a simple reflexive (Fr. *cela se sent*, Sp. *eso se siente* ‘it can be felt, it can be perceived’) or combined with causation (Fr. *cela se fait sentir*, Sp. *eso se hace sentir* ‘it makes itself felt, it makes itself perceived’). Hungarian has a separate lexeme for this meaning as well, derived from *érez*: *éreződik* ‘it can be felt, it can be perceived’ (the differences between these constructions would be worth a more in-depth study in the future). The third construction, on the other hand, is reserved for Hungarian: it is the predicative use of the infinitive *érezni* ‘to feel, to perceive’ in the same sense as the above (‘it can be felt, it can be perceived, it is perceptible’), probably originating from the ellipsis of *lehet* ‘it is possible’.

At the discourse level, one specific construction can be highlighted, as it occurs with much higher frequency in Hungarian (12.7%) than in French (7.0%) and Spanish (4.7%). It is a specific use of the reflexive with an adverbial complement, meaning ‘to feel good / bad / marvelous / awful etc.’. In all three languages, the positive and mental domains are prevalent, but there are some references to negative moods and to (positive or negative) bodily states as well.

Table 7. A selection of relevant constructions with Fr. *sentir*, Sp. *sentir*, and H. *érez*

Constructions	Fr. <i>sentir</i>		Sp. <i>sentir</i>		H. <i>érez</i>	
	tokens	percent	tokens	percent	tokens	percent
reflexive	244	51.7%	217	46.5%	133	26.8%
subordinate clause	50	10.6%	65	13.9%	177	35.7%
indefinite object	5	1.1%	34	7.3%	34	6.9%
causative	30	6.4%	31	6.6%	-	-
object-oriented (percept) conceptualization	11	2.3%	19	4.1%	-	-
infinitive (<i>érezni</i>)	-	-	-	-	5	1.0%
‘feel good/bad’	33	7.0%	22	4.7%	63	12.7%

5. Conclusion

This paper has presented an empirical examination with a view to mapping out the semantic and formal particularities of the primary basic-level multimodal perception verbs in French, Spanish, and Hungarian: Fr. *sentir*, Sp. *sentir*, and H. *érez*. Following the footsteps of Enghels and Jansegers (2013) who carried out a similar contrastive analysis of Fr. *sentir*, Sp. *sentir*, and It. *sentire*, the present study hopes to demonstrate that not only cognates should be studied in this way, but that it is worthwhile to look into *feel*-type verbs in other languages as well. So far, both linguistics and the philosophy of perception have focused mainly on visual (and sometimes auditory) perception, though vision is probably not a typical sensory modality at all (cf. Lycan, 2000): studies on the other sensory channels (the so-called “lower senses”, proprioception etc.) may lead to new

insights on the nature of perception itself and on its conceptualizations in different cultures.

The present analysis is an exploratory step in this direction. By looking at data extracted from the TenTen corpora, it has highlighted important aspects of similarity and variation in the behavioral profile of the studied verbs. Essentially, it has shown that mental (cognitive and emotional) perception is referred to at least about twice as frequently as physical perception in the examined sources (see Table 8). The Spanish and Hungarian data are very close in this respect (here this proportion is 1 : 2.9 and 1 : 2.6, respectively), but French is biased in favor of physical perception because of its 50 olfactory occurrences (among which 19 are used figuratively and thus should be actually counted as instances of mental perception).

Table 8. The two principal domains of perception expressed by Fr. *sentir*, Sp. *sentir*, and H. *érez*

Type of perception	French		Spanish		Hungarian	
	tokens	percent	tokens	percent	tokens	percent
physical perception	166	35.2%	116	24.8%	136	27.4%
mental perception	305	64.6%	343	73.4%	357	72.0%

In sum, the cross-linguistic differences found in this examination may result from: 1. the inherent semantic characteristics of the verbs (e.g. the agentive and percept olfactory meanings of Fr. *sentir*); 2. structural features of a language (e.g. fewer reflexive forms in Hungarian because the reflexive construction is not as multi-faceted as in the Romance languages); 3. cultural and discourse features (e.g. a greater proportion of ‘feel good/bad’ in the Hungarian data).

The study also points out that there is still much to do in the contrastive examination of Fr. *sentir*, Sp. *sentir*, and H. *érez*, as well as in the exploration of perceptual language in general. First and foremost, a thorough analysis of the correlations between the different tags used in the present study is needed in order to make generalizations and delineate paths for future research. Specifically, a comparison of reflexive and non-reflexive occurrences would be particularly revealing, but it would also be important to examine the relations between the primary and the additional meanings (epistemic, figurative etc.). An analysis of recurrent patterns of use (similar to that of ‘feel good/bad’) would enrich our knowledge about the typical functions of the studied verbs in their sociocultural contexts, while looking at H. *megérez* and Fr. *ressentir* would provide valuable information on the role of aspect in the conceptualization of perceptual experiences and on partial intralinguistic synonymy, respectively. A further possibility would be to include an axiological tag that differentiates between positive, neutral, and negative sensations and enables investigations in this respect, too. And one could also widen the scope of the comparison and examine

multimodal perception verbs in other languages or verbs referring to other types of perception.

References

- Bergen, B. K.** (2012). *Louder Than Words. The New Science of How the Mind Makes Meaning*. New York: Basic Books.
- Bergen, B. K.** (2015). Embodiment, simulation, and meaning. In: Riemer, N. (ed.) *The Routledge Handbook of Semantics*. London & New York: Routledge. 142–157.
- Brugman, C.** (1988). *The Story of Over: Polysemy, Semantics and the Structure of the Lexicon*. New York: Garland.
- Croft, W.** (2000). *Explaining Language Change. An Evolutionary Approach*. Harlow: Longman.
- Cruse, D. A.** (1982). On lexical ambiguity. *Nottingham Linguistic Circular* 11. pp. 65–80.
- Engels, R. & Jansegers, M.** (2013). On the crosslinguistic equivalence of *sentir(e)* in Romance languages: A contrastive study in semantics. *Linguistics* 51/5. pp. 957–991. <https://doi.org/10.1515/ling-2013-0034>
- EWUng = Benkő Loránd et al.** (eds., 1993–1995). *Etymologisches Wörterbuch des Ungarischen* 1–2. Budapest: Akadémiai Kiadó.
- Fernández Jaén, J.** (2012). *Semántica cognitiva diacrónica de los verbos de percepción física del español*. Tesis doctoral. Alicante: Universidad de Alicante.
- Galac Á.** (2020). Semantic change of basic perception verbs in English, German, French, Spanish, Italian, and Hungarian. *Argumentum* 16. pp. 125–146. <https://doi.org/10.34103/ARGUMENTUM/2020/9>
- Geeraerts, D.** (2010). *Theories of Lexical Semantics*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198700302.001.0001>
- Geeraerts, D.** (2015). Sense individuation. In: Riemer, N. (ed.) *The Routledge Handbook of Semantics*. London & New York: Routledge. 233–247.
- Gisborne, N.** (2010). *The Event Structure of Perception Verbs*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199577798.001.0001>
- Győri G.** (2002). Semantic change and cognition. *Cognitive Linguistics* 13/2. pp. 123–166. <https://doi.org/10.1515/cogl.2002.012>
- Hertegonne, E.** (2014). *Étude diachronique des propriétés sémantico-syntaxique du verbe de perception sentire/sentir du latin jusqu'au français contemporain*. MA Thesis. Gent: Universiteit Gent.
- Jansegers, M.** (2017). *Hacia un enfoque múltiple de la polisemia. Un estudio empírico del verbo multimodal « sentir » desde una perspectiva sincrónica y diacrónica*. Berlin & Boston: De Gruyter. <https://doi.org/10.1515/9783110476972>
- Jansegers, M. & Gries, S. Th.** (2020). Towards a dynamic behavioral profile: A diachronic study of polysemous *sentir* in Spanish. *Corpus Linguistics and Linguistic Theory* 16/1. pp. 145–187. <https://doi.org/10.1515/cllt-2016-0080>
- Lakoff, G.** (1987). *Women, Fire and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press. <https://doi.org/10.7208/chicago/9780226471013.001.0001>
- Lakoff, G. & Johnson, M.** (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.
- Lycan, W. G.** (2000). The slighting of smell. In: Bhushan, N. & Rosenfeld, S. (eds.) *Of minds and molecules: New philosophical perspectives on chemistry*. Oxford: Oxford University Press. 273–289.
- Rosch, E.** (1978). Principles of categorization. In: Rosch, E. & Lloyd, B. B. (eds.) *Cognition and Categorization*. Hillsdale, NJ: Lawrence Erlbaum, 27–48.
- Sharifian, F.** (2008). Distributed, emergent cultural cognition, conceptualisation and language. In: Frank, R. M., Dirven, R., Ziemke, T. & Bernárdez, E. (eds.) *Body, Language and Mind. Vol 2. Sociocultural Situatedness*. Berlin & New York: Mouton de Gruyter, 109–136.

- Sweetser, E.** (1990). *From Etymology to Pragmatics. Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511620904>
- TESz = Benkő L et al.** (eds., 1967–1976). *A magyar nyelv történeti-etimológiai szótára 1-3*. Budapest: Akadémiai Kiadó.
- Viberg, A.** (1984). The verbs of perception: a typological study. In: Butterworth, B., Comrie, B. & Östen, D. (eds.) *Explanations for Language Universals*. Berlin: De Gruyter, 124–162. <https://doi.org/10.1515/9783110868555.123>
- Winter, B.** (2019). *Sensory Linguistics. Language, perception and metaphor*. Amsterdam, NL & Philadelphia, PA: John Benjamins Publishing Company. <https://doi.org/10.1075/celcr.20>
- Yu, N.** (2015). Embodiment, culture, and language. In: Sharifian, F. (ed.) *The Routledge Handbook of Language and Culture*. London & New York: Routledge, 227–239.

¹ In this paper, the term “multimodal” is used in the context of perception verbs that can refer to more than one sensory modality. It should not be confused with other possible uses, e.g. Forceville’s multimodal metaphor.

² It is important to note that the idea of the embodied mind is not without precursors, as it was already proposed by the philosopher Giambattista Vico (1668–1744), as well as by Maurice Merleau-Ponty and Jean Piaget (for a historical overview see Yu, 2015).

³ But not too distinct, for then they may get detached from the network and become the global prototypes of new lexical items – a process that Győri (2002: 152) terms as “prototypicalization”.

⁴ Cf. the digitized *Latin Dictionary* by Charlton T. Lewis & Charles Short for more detail and examples: <http://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0059:entry=sentio> (22.06.2021)

⁵ For a typology of perception verbs (experiencer, agentive, and percept / object-oriented / copulative), see Viberg (1984), Gisborne (2010: 4–8), or Galac (2020: 127).

⁶ From a cultural linguistic point of view, and also as a further aspect on the composition of the TenTen corpora, it may be worth noting that while the Hungarian data contained 21 and the French data 19 pornographic sources, in the case of Spanish it was only 4.

⁷ Translation by the author. The original text: “es evidente que *sentirse* se ha desesemantizado casi por completo en estos contextos” (Fernández Jaén, 2012: 441).