Core features: measures and characterization for different languages

Published Version:
Vivas L., Montefinese M., Bolognesi M., Vivas J. (2020). Core features: measures and characterization for different languages. COGNITIVE PROCESSING, special issue 21(4), 651-667 [10.1007/s10339-020-00969-5].

Availability:
This version is available at: https://hdl.handle.net/11585/768988 since: 2020-11-23

Published:
DOI: http://doi.org/10.1007/s10339-020-00969-5

Core features: measures and characterization for different languages

Vivas, Leticia[1]; Montefinese, Maria[2,3]; Bolognesi, Marianna[4]; Vivas, Jorge[1]

[1]Basic and Applied Psychology and Technology Institute (IPSIBAT), UNMDP, CONICET, Psychology Faculty, Funes 3250, Mar del Plata (CP: 7600) Argentina

ORCID: Leticia Vivas 0000-0002-5295-8742

ORCID: Jorge Vivas 0000-0001-8648-2562

[2]Department of General Psychology, University of Padova, Via Venezia 12,

35131 Padova, Italy

[3]Department of Experimental Psychology, University College London, 26 Bedford Way WC1H 0AP London, United Kingdom

ORCID: Maria Montefinese 0000-0002-7685-1034

[4] Faculty of Modern Languages, University of Oxford, 47 Wellington Square OX1 2JF Oxford, United Kingdom

ORCID: Marianna Bolognesi 0000-0002-3292-8968

Abstract

According to the feature-based view of semantic representation, concepts can be represented as distributed networks of semantic features, which contribute with different weights to determine the overall meaning of a concept. The study of semantic features, typically collected in property-generation tasks, is enriched with measures indicating the informativeness and distinctiveness of a given feature for the related concepts. However, while these measures have been provided in several languages (e.g. Italian, Spanish and English), they have hardly applied comparatively across languages. The purpose of this paper is to investigate language-related differences and similarities emerging from the semantic representation of aggregated core features. Features with higher salience for a set of concrete concepts are identified and described in terms of their feature type. Then, comparisons are made between domains (natural vs. artefacts) and languages (Italian, Spanish, and English) and descriptive statistics are provided. These results show that the characterization of concrete concepts is overall fairly stable across languages, although interesting cross-linguistic differences emerged. We will discuss the implications of our findings in relation to the theoretical paradigm of semantic feature norms, as well as in relation to speakers' mutual understanding in multilingual settings.

Keywords: conceptual representation; semantic features; feature listing task

Introduction and theoretical background

The feature-based models assume that semantic representation of concepts can be described as a distributed set of features (Baroni et al. 2010; Bolognesi, 2017; Cree et al. 2006; Sartori, Polezzi, et al. 2005; Taylor et al. 2007; Vigliocco et al. 2009; Vigliocco et al. 2004). These models suggest that conceptual representations may consist of different types of features, such as perceptual, functional, associative, and encyclopaedic, which have been traditionally classified in several ways according to different possible taxonomies (e.g., Garrard et al 2001; Wu and Barsalou 2010). Different types of concepts can display therefore diverse distributions of feature types across semantic categories (tools, animals, fruits, etc.) and domains (natural or artefact). For example, in a pioneering study, Warrington and McCarthy (1987) reported category-specific impairments observed in patients who have suffered brain damage in specific brain areas. Such patients displayed selective impairments of visual and functional information, which are recognized to be crucial for the processing of living things and artefacts respectively. Subsequent empirical analyses conducted on normative data derived from a feature listing task have shown that different categories (i.e., foods, creatures, tools) are indeed characterized by different types of feature configurations. It is now accepted that the concepts of the artefacts category show a greater number of functional features, particularly the distinctive ones (Garrard et al., 2001). Instead, the food category is characterized by a greater number of features related to the senses of touch and taste, encyclopaedic features and a smaller number of properties denoting visual parts and surface. Furthermore, the concepts from the animal's category tend to have more visual–motion features and only few functional features (Cree and McRae, 2003).

Besides, the degree of contribution of each feature to the core meaning of the related concept (i.e. the semantic features that are key to the identification of the concept) varies. The identification of the salient features of a concept has important implications for the study of normal and impaired cognition (Ursino et al., 2018). There are several ways to measure this contribution. For example, distinctiveness, cue validity, relevance, and significance are measures that have shown to have an effect on several tasks, such as, confrontation naming (Catricalà, et al., 2015), naming-to-definition (Sartori, Lombardi, et al.2005) and feature verification (Montefinese et al. 2014a) in different languages.

These measures of salience that allow for the identification of the core features are generally extracted from semantic feature production norms. Numerous research groups of different countries and languages have analysed and elaborated feature production norms (Buchanan et al. 2013; Buchanan et al. 2019; De Deyne et al., 2008; Lenci et al. 2013; McRae et al. 2005; Moldovan et al. 2015; Montefinese et al. 2013a; Ruts et al., 2004; Vinson and Vigliocco, 2008; Vivas et al. 2017; Zannino et al. 2006). These norms are also particularly important for elaborating and testing theories about semantic representation. Devereux et al. (2014) suggest that these theories have notably improved since models based on information derived from the semantic production norms have been developed. This information not only refers to the identity of the features produced in the norms, but also to their statistical properties. The most common measures of feature salience and the exact procedure used to compute them will be described in the following paragraphs. Although all these measures indicate the salience of a feature to the concept meaning, they can capture different aspects of its semantic representation.

One of the most well-known measures is the feature distinctiveness. This is calculated as the inverse function of the number of concepts for which the feature appears across the dataset. Feature distinctiveness allows to identify and distinguish a concept from the others. A related measure is the so-called cue validity (Bourne and Restle, 1959; Rosch and Mervis, 1975). This is computed as the conditional probability of a concept given a feature, and represents how easily a feature "pops up" when thinking of a concept (Smith and Medin, 1981).

The production frequency is another measure whose value indicates the number of participants who list a feature in response to a given a concept in a property-generation task and it can be transformed into feature dominance, when it is calculated as the proportion of participants listing such feature for a given concept. This measure reflects the importance of a feature for a given concept and the accessibility of that feature when a concept is recalled from its name (Smith and Medin, 1981). More recently, Sartori and Lombardi (2004) proposed the concept of relevance. This variable is composed of two values: a local component, that is, the dominance and a global component, that is, the distinctiveness (both described above). In the last few years, Montefinese and colleagues (2014a) have proposed an alternative and more complete measure called semantic significance. Following a similar logic of the relevance measure, this variable includes the distinctiveness and the accessibility of a feature, which is a composite measure of dominance plus order of production values.In the same way, Vivas and colleagues (2014) have proposed a similar measure like accessibility, called relative weight. This measure indicates the weight of the feature (frequency plus order) as a function of all the data produced for that concept by a group of people. In this study we did not consider these latter two measures (significance and relative weight) because the information about the featural production order is not available in the English norms (McRae et al., 2005). This prevented the possibility to calculate the significance and the relative weight in cross-language comparisons that include the English language.

Up to now, some measures of feature informativeness have been described and their influence on different behavioural tasks has been illustrated. In this scenario, some fundamental questions about the stability of those core features in different languages arise: Does a given concept have the same core features in different languages? Do these measures have similar weight in conceptual representation across different languages and cultures?

These questions lead to the well-known debate about the relation between thought and language. The strong view of the linguistic relativity hypothesis states that language influences conceptual knowledge, leading to differences in the concept representation for speakers of different languages (Levinson, 1996). However, a more dynamic view (see Montefinese et al. 2019) of the linguistic relativity hypothesis proposed that language properties may influence semantic representation under specific circumstances such as in tasks involving a deep encoding of lexical properties of concepts (Vigliocco et al 2005) or tasks engaging a verbal encoding (Brysbaert et al. 1998).

Interestingly, Gentner and Goldin-Meadow (2003) affirm that the dominant position of the last few decades within cognitive psychology has been that “conceptual structure is relatively constant in its core features across cultures” (p. 5), suggesting a lack of variation of semantic representation across languages. In this vein, a recent comparative study of semantic feature production norms for English, Italian and German (Kremer and Baroni, 2011) pointed out that, although the idiosyncratic language differences exist, the pattern of distribution of feature types remains similar across the three languages. Likewise, Matl and colleagues (2003) propose that although language categorizations tend to vary cross-linguistically, conceptual categorizations tend to be more universal. In particular, speakers of English, Chinese, and Spanish showed substantially different patterns of naming for a set of common objects, but they categorized the objects by similarity in much the same way (Matl et al. 1999). Other authors agree in that although the conceptual representation is dynamic and its activation is context-dependent the words that we use to communicate with other speakers obviously refer to some common conceptual features that allow us to understand each other (White, Storms, Matl, & Verheyen, 2018; Yee & Thompson-Schill, 2016). Consequently, there should be stable and essential aspects of the concepts’ meaning that are shared between speakers and allow effective communication. These “core” features might be shared across languages. Martin (2016) on the proposal of the GRAPES model, calls these features conceptual “primitives”. He characterizes them as “object-associated properties that underpin our ability to quickly and efficiently identify objects at the basic-category level (…) regardless of the modality of presentation” (p. 2). For example, when we think about a cat we all probably think about an animal that miaow. So “is_an_animal” and “miaow” constitute core features of the concept. However, even those core features can show certain variability across cultures as it will be shown in the current paper.

The goal of our study is to test if equivalent words that are usually translated into one another across three languages (English, ENG; Spanish, SPA; and Italian, ITA) actually designate (concrete) concepts that are characterized by slightly different configurations of features. While comparing the feature configuration between languages it is possible to identify some common features that constitute the core meaning of the concept at issue as we will show. However, we hypothesize that there are additional elements of conceptual representation that would be part of the cultural hub of each language and that would not be shared across languages.

In particular, we hereby report an empirical study in which we address the following specific research questions:

1. To what extent are concepts expressed in different languages (ENG, SPA, ITA) comparable, when we look at the features produced to describe their content? Does the concept’s domain (natural/artefact) have an influence?
2. Is the pattern of similarity between concepts comparable across languages? Is the similarity-based categorization equivalent across languages?
3. Is featural composition the same across the three languages? Do the participants produce the same type of features?
4. Are feature informativeness measures consistent across language? Are they influenced by the concept’s semantic domain (artefact vs. natural)?

One of the novelties of the current paper is the variety of analysis proposed to answer these questions, as we will perform vector comparison techniques, qualitative analyses, social network analysis and statistical samples comparisons. We also include three of the most widely used languages: English, Spanish and Italian. The remainder of the paper is organised as follows: in Section 1 we show comparisons of feature sets between languages; in Section 2 we test cross-language consistency in conceptual similarity pattern and categorization; in Section 3 we show cross-language comparisons of the distribution of feature types; in Section 4 we establish cross-language comparisons for the feature informativeness measures considering the influence of domain.

<u>Section 1</u>

Methods

Materials

The stimuli and semantic measures used to perform the analyses were extracted from three semantic normative databases: Italian (Montefinese et al., 2013), English (McRae et al., 2005) and Spanish (Vivas et al., 2017) derived from a feature listing task. The Italian database was composed of 120 concepts, the English 541 and the Spanish 400. The instructions given to participants were similar across the three norms: participants were asked to list features that describe the concept, examples were provided, responses were collected on written forms and the task did not have a time limit. The posterior recording procedure to identify similar features (e.g. *legs, has_legs*) followed the same criteria across the three norms. The sample in the three cases was composed of university students.

For the purpose of feature comparison, we derived featural informativeness measures (McRae et al., 2005; Buchanan et al., 2013; 2019; Vivas et al., 2017) and conceptual similarity measures (Montefinese et al. 2015; Montefinese et al. 2018) for the shared concepts between norms. As a first step, common concepts between databases were identified. Italian and Spanish shared 65 concepts, Italian and English 63, and Spanish and English 214. The intersection of the three languages was 43 concepts. . For Section 1 we used the set of 43 concepts. All the materials reported are freely available to the scientific community for noncommercial use on Open Science Framework (OSF) repository (http://tiny.cc/4z656y).

Procedure

To explore if feature composition is similar across languages a feature vector comparison was performed. The vectors were constructed for the 43 common concepts across the three languages. Attributes from feature production database were included if they were produced by at least five participants, since those with a lower production rate could be considered non-normative (this is the usual criteria for semantic feature production norms). In order to build the concept's vectors the features were ordered by production frequency. Pairwise comparisons were performed through a cosine similarity analysis through the software Synonym Finder (http://iaai.fi.mdp.edu.ar:8080/sfweb)[1]. To do this calculation, this software uses a well-established geometric vector comparison technique in the usual n-dimensional Euclidean space from the angle formed between them, such that parallelism represents greater similarity while orthogonality indicates greater difference (Kintsch, 2001). The greater cosine values represent greater semantic similarity between concepts. Values were in the range between 0 and 1. As an initial analysis, the distribution of cosine values across the three pairs of languages was displayed. Percentile, minimum and maximum values were identified. As a second step, the analysis focused on differences between languages that could explain the low values in the cosine similarity analysis. First, a qualitative analysis was performed in order to detect the concepts that obtained lower similarity values. We identified the features with high salience in one language but low salience or even absent in the other language. Some examples are displayed to illustrate them. To test if there were differences in the quantity of features produced per concept across the three languages, we performed a one-way ANOVA to compare the number of features (with and without taxonomic features) across the three languages. Finally, a t-test was performed comparing the cosine values between domains for each pair of languages in order to study if the differences were mainly in natural or artefacts domains.

Results

For the 43 shared concepts between English and Spanish the median value was .71 and the range was between .39 to .93; for Italian and Spanish the median value was .65 and the range from .07 and .90; and for Italian and English the median was .59 and the range from .23 to .85. On Table 1 descriptive statistics of the distribution of values for each pair can be seen.

== Please, insert Table 1==

As can be observed, the majority of the concepts show a high level of correspondence between languages, indicating similar configuration of features between norms. The full cosine matrices can be found in the electronic repository on OSF as mentioned above.

As a next step, a qualitative analysis was performed with those concepts with low similarity values (lowest quartile). It could be observed that: 1) for certain concepts there are specific features that were produced in one culture with high production frequency but not in the other; 2) English tend to have more features per concept, and 3) some semantic categories seem to be more different in

[1] The interested reader can ask the corresponding author for access.

feature configuration than others. Consequently, a deeper analysis was performed in order to tackle these issues as will be detailed.

The following examples can be mentioned to illustrate the first point:

a) ENG/SPA comparison: for the concept barn, the cosine value obtained was .44. English participants produced the feature *is_red*, with a production frequency of 12. This is probably because they tend to be painted with that colour in North America. As that is not the case in Argentina, Spanish participants did not produce this feature. The features *found_in_the_country* and *found_on_farms* were produced in both languages but the measures of informativeness indicated their relative importance was very different between the two languages. The former obtained lower value of production frequency in English (10) than in Spanish (19), while the latter obtained the opposite relation, higher in English (24) than in Spanish (5). Another example can be mentioned is the feature *mate* for the concept kettle, where the cosine value between English and Spanish was .19. Indeed, for Spanish the feature *mate* took values of production frequency of 18. The reason is that the corresponding beverage is very popular on the Argentinean Spanish speaking culture, but unsurprisingly it did not appear in the English norms.
b) for the ITA/SPA comparison: many of the concepts belonged to the categories of clothes (hat, coat and jacket) and kitchenware (fork, spoon, pot and ladle). In general, Italian tend to produce more features for those concepts, particularly those referring to the variants the concept can show. For example, for the concept COAT Italian participants produced *is_long* (PF = 13), *has_buttons* (PF = 10), *is_short* (PF = 8), *made_of_different_materials* (PF = 7). Another example, from the category of kitchenware is the concept spoon where they produced *made_of_steel* (PF = 16), *made_of_plastic* (PF = 12), *made_of_silver* (PF = 7). Furthermore, for the natural concepts that obtained lower values of similarity the reason seems to be that Spanish participants tended to produce more taxonomic features such as *is_a_mammal*, *is_an_animal*, *is_quadruped*.
c) for the ITA/ENG comparison: for the concept closet the feature *is_small* was produced in the English norms while the feature *is_big* was produced in the Italian norms. The former obtained values of production frequency of 5, for the English database and didn't appear in the Italian database, while the latter obtained values of production frequency of 10 for the Italian database and didn't appear in the English database.

Regarding the number of features produced, it was observed that the English had significantly more features per concept than the other databases, either including taxonomic features (F(2,126) = 59.42; p < .001) or excluding them ($F_{(2,126)}$ = 91.35; p < .001) with a mean of 13.98 (SD = 3.25) with taxonomics and 15.60 (SD = 3.77) without, followed by Italian (M = 14.19; SD = 2.89; and M = 14.95; SD = 3.17) and Spanish (M = 6.98; SD = 2.19; and M = 8.72; SD = 2.66). The total number of features for the set of 43 concepts for the English database was of 671, for the Italian 643 and for the Spanish 375. As can be seen, the Argentinean participants were the ones that produced fewer features overall.

Moreover, for each pair of languages a two-tailed independent sample *t*-test was carried out with the set of 43 shared concepts to compare the mean cosine values between domains (artefacts vs natural). It was observed that there was no statistical difference between ENG/SPA ($t_{(41)}$ = -1.26; p = .21; d = .46) and ITA/ENG ($t_{(41)}$ = -1.04; p = .309; d = .40), but significant differences were observed in the pair ITA/SPA ($t_{(41)}$ = -2.45; p = .02; d = .91) indicating that natural domain concepts tended to show lower cosine values.

Discussion

In general, high similarities in feature lists across languages were observed. It is important to note here that it was not expected to find values of similarity near to 1 because we assume, as some other authors do (Yee and Thompson-Schill, 2016), that conceptual representations are not stable among speakers of a same language due to their different experience with that object, consequently even more differences are expected to be observed across languages. Besides, the degree of similarity was not equivalent between pairs of languages. The greater similarities were observed between Spanish and English, followed by Italian and Spanish and the least, Italian and English. This is counterintuitive because Italian and Spanish seem to have a priori similar cultural roots, while English and Spanish do not. However, there are some specific categories such as kitchenware and clothes, where Italians tend to produce more specific features in these two domains. This can be interpreted in light of the cultural differences between these two countries, and can be motivated by the relevance that Italians give to food and fashion, considered as cultural hallmarks of this country.

From the qualitative analysis it can be inferred that the differences observed across languages are a reflex of the cultural differences regarding the daily experience with the object at issue. Interestingly, even though we focused our analysis on the featural representations of concrete concepts, which denote tangible referents, we observed some interesting differences across languages. For example, the concept colander is richer in features in Italian and Spanish compared to English. We believe that this may be due to the extreme salient role that colanders play in the Italian and Argentinian culture and the life of its speakers, being used on a daily basis to drain pasta. Another example with the concept horse shows how geographical differences between countries can influence concept representation. Horses in Argentina are associated with the countryside, with pampas and large plains, while Italian speakers associate it with horse racing, probably because they are used to see those animals in that scenario. Consequently, although core properties tend to maintain stable across languages and allow speakers to translate terms and to communicate and understand each other, features that are more peripheral can change according to the experience with the object of reference as proposed by Yee and Thompson-Schill (2016). Variables such as the familiarity with the object or the way it is used on certain countries seem to have an impact on the features produced on each language. These variables also showed to be responsible for variability within age groups in the same language (White, Storms, Malt, & Verheyen, 2018). It is important to note that semantic feature production norms focus on the commonality across speakers but inter-subject variability also exist (Chaigneau, Canessa, Barra, & Lagos, 2018).The difference in the number of features across languages can also weakly bias the comparison. Although the weight of the non common features is very low for the vector comparison when they have low production frequency, it can be the case that Argentinean participants omitted features that could be common with the other two languages because they considered them irrelevant for the task. In spite of having the same instructions there is a clear difference in the way the Argentinean participants interpreted them or the level of motivation to respond. The only difference that we observed in the procedure of data collection is that McRae´s team send the forms to fill in to many participants and they received back only a percentage. As a consequence, those participants without sufficient motivation could opt for not sending the form back. In the case of que Argentinean sample, they did the activity in the context of a degree course requirement (they had to choose between filling the form or doing an alternative activity), so perhaps many participants chose to complete the form without being completely engaged with the task.

Regarding domain differences, it was observed that the concepts displaying lower similarity values between Italian and Spanish designated mainly natural things. This could be explained by the fact that this domain includes many taxonomic features, and, as we will analyse in depth in Section 3, Spanish-speaking participants tended to produce more taxonomic (i.e. categorical) features.

These results suggest that although some relevant features are shared between most of the selected concepts (as indicated by the high values of cosine similarity) allowing the corresponding labels (words that denote the concept) to be translated, there are also cultural particularities that differentiate the concept representation from one culture to another.

Section 2

Methods

Materials

The stimulus set was the same as for the Section 1.

Procedure

To investigate whether the semantic similarity between concepts within a language is consistent across the English, Italian and Spanish sets, we carried out the representational similarity analysis (see Montefinese and Vinson, 2017 for a similar procedure). In other words, first we derived the concept-concept dissimilarity matrices computed on normalized distance measures (i.e., cosine distances between the feature vectors of every concept pair) and then we performed Spearman's pairwise correlations between dissimilarity matrices of each language pair (ENG vs. ITA, ENG vs. SPA, ITA vs. SPA).

analyseTo perform the clustering we used the Johnson's Hierarchical Clustering (Johnson, 1967). This procedure is as follows: given a symmetric matrix representing similarities, the algorithm finds a series of nested partitions of the items. Then, the different partitions are ordered according to increasing levels of similarity. The algorithm begins with the identity partition (in which all items are in different clusters). After that, the most similar pairs of items are joined and they are considered a single entity. In this way, the algorithm continues until all items have been joined into a single cluster. We considered a value of association of .30 to visualize the clusters. The software selected to plot the results was NetDraw (Borgatti, 2002). This visualization software is part of the Social Network Analysis software UCINET 6 (Borgatti et al. 2002).

Additionally, the clustering coefficient was calculated for the three networks. This number represents a measure of the extent to which nodes tend to cluster together (Dubossarsky, De Deyne, and Hills, 2017). The weighted overall clustering coefficient is the weighted mean of the clustering coefficient of all the actors, each one weighted by its degree.

Results

Representational similarity analysis showed a high degree of similarity between conceptual representations derived from the different languages. Specifically, the similarity between concepts in Italian population was strongly correlated with that from other semantic norms (for English: $r = .52$, $t_{(61)} = 27.057$, $p < .001$; for Spanish: $r = .55$, $t_{(65)} = 30.36$, $p < .001$), showing a good consistency across different languages. Spanish and English showed a higher variability between the two languages, as revealed by the lower but still high correlation ($r = .57$, $t_{(212)} = 105.59$, $p < .001$).

These results were corroborated by the cluster analysis. The corresponding graphs for each language can be seen below (Figures 1-3) as well as for the three languages (Figure 4). In general, this analysis revealed a high correspondence in the feature similarity-based categorization of concepts across the three languages.

== Please, insert Figure 1 ==

== Please, insert Figure 2 ==

== Please, insert Figure 3 ==

== Please, insert Figure 4 ==

In particular, the emerging clusters seem to be characterized by semantically coherent classes of concrete concepts. The categories clothes, animals, furniture, means of transportation and kitchenware are present in the three languages. However, they show some slight differences within the semantic categories. For example, in English and Italian air vehicles (airplane and helicopter) constitute a separate cluster, detached from the cluster of ground transportation. In Spanish all the vehicles are clustered together, and they are also connected to the class of concepts denoting animals by the link between truck and elephant. This is probably due to the fact that both share the features *is_large* and *is_heavy*. Something like this occurred in the links between ladle and chair, table and door, in italian language. Furniture shared features with ladle like made_of_wood and has_a_handle.

Another example is the cluster of concepts that denote tools (artefacts) that while in Spanish is isolated, in Italian is connected to the class of concepts denoting furniture. In English the artefacts are isolated, like in Spanish, with knife and scissors constituting a separate cluster that could be named ‘cutting tools’, ‘sharp tools’, or ‘tools that could be used as weapons’.

The clustering procedure for the three joined languages show the same clusters as the separate clustering by language. Only concepts like ashtrays, lamp, clock, ruler and pencil that were left alone on the individual languages clustering, now where clustered together with the same concept in the other two languages. As we said in Section 1, the core properties tend to maintain stable across languages and allow speakers to translate terms and to communicate and understand each other. The slight nuances of meanings that concepts have in each person and in each culture affect the values of similarity and the consequent clustering procedure, but do not impact the hard core of the meaning of each concept.

Overall, Spanish seems to have more defined clusters. On the contrary, the Italian and English clusters tend to be more spread (106.114 and 101.975 respectively), as can be seen for the categories of animals, clothes and means of transportation. Consequently, the English language has the less dense network.

Discussion

Overall, we found that representation of similarity between concepts is not affected by language differences, suggesting that this measure is stable across languages. This result has been supported by the cluster analysis performed on the similarity between concepts for each language. In particular, we observed that semantic clustering was similar between languages, showing that concepts tend to be organized within the same categories as Matl and colleagues (1999) proposed. However, the differences in the salience of some key features was reflected in some cluster variations across languages. The difference in the dispersion of clusters, particularly between Spanish and Italian can probably be due to the use of taxonomic features. This kind of attribute was produced with high frequency by Spanish participants and would determine the fusion of the clusters and consequently, the high value of clustering coefficient. On the contrary, Italian participants were less likely to produce taxonomic features. This could be the reason why there was not a single cluster for animals where these kind of concepts could be organized. In fact many of them, like cow, elephant or horse were left alone in the network. In the same way some clothes were left outside the clusters and means of transportation were established in two separated clusters.

Regarding the clusters produced by the three languages together, it showed the same behavior as each particular language graph and, fundamentally, grouped together equivalent concepts across languages.

Section 3

Methods

Materials

The 43 common concepts across the three semantic norms were selected for this analysis (McRae et al., 2005; Montefinese et al., 2013a,b; Vivas et al., 2017).

Procedure

In order to identify similarities and differences in the type of features produced in each norm a chi-squared analysis was carried out. Garrard and colleagues'(2001) taxonomy was used to code features which includes the following categories: sensorial (including information from all sensory modalities, eg. is_red), function (those that describe an action, activity or use of an item, eg. flyes), encyclopaedic (those describing associative relations, such as for bag the feature *found_in_grocery_stores*) and categorizing (i.e. superordinate category, eg. is_an_animal). This analysis was performed for the full set of shared concepts and then by domain.

Results

The feature type proportion across languages was analysed. It was observed that in the three languages sensorial features prevail, followed by function, encyclopaedic and categorizing. However, significant differences between proportions were observed across languages (C = .18; p < .001). Italian participants tend to produce a higher proportion of sensorial features than the other two languages, while Spanish tend to use more categorizing features (Figure 5).

== Please, insert Figure 5 ==

Additionally, a comparison across languages but split by domain was performed. Significant differences in feature type were again observed in both domains, natural (C = .17; p < .001) and artefacts (C = .12; p < .001) indicating that there was no domain effect on the feature type differences across languages.

Discussion

The most salient difference between languages was that categorizing features tended to appear in Spanish participants, while they were less likely to appear in Italians. Besides, in this last language sensorial features tend to prevail. These results are intriguing because as instructions, concepts and populations were equivalent it is hard to find a satisfactory explanation. Besides, they are contrary to the results obtained by Kremer and Baroni (2011), who found that Italian participants tended to produce more categorizing features than German and English participantes. However, a direct comparison with that paper must be taken with caution as the instructions, coding scheme and concepts were different. More evidence should be obtained in order to be able to give a satisfactory explanation of this result.

Furthermore, considering the similarities observed, all languages showed higher proportion of sensorial features. It is recognized that the sensorial data is a very relevant source of information to develop conceptual representations, especially for the concrete concepts, as in the case of the current study (see Hoffman et al. 2018; Meteyard et al. 2012). Thus, it is very likely that most of the features that compose conceptual representation of concrete concepts correspond to this kind of information. Functional features and sensorial features constitute what Borghesani and Piazza (2017) call motor-perceptual semantic features and are extracted through our physical interaction with the object. While the other kind of features, encyclopaedic and categorizing are not acquired through direct sensory-motor experience but derived from language or extracted by a posterior integration of motor-perceptual features. Categorizing features require some cognitive operations that allow to establish rules to decide the membership of each concept to a category, while encyclopaedic are mainly extracted from declarative information, consequently acquired through language. As these last kinds of features are mainly defined by language (Borghessani and Piazza, 2017) it is very likely that they tend to show more cross-linguistic variability.

Section 4

Methods

Materials

The stimuli set were the shared concepts between pairs of languages as described in Section 1.

Procedure

We averaged each informativeness measure (i.e., dominance, distinctiveness, cue validity and relevance) across all the features of each common concept for each language. Note that although the values of the informativeness measures have been derived from the databases of the available norms (McRae et al., 2005; Montefinese et al., 2013a,b; Vivas et al., 2017), due to slight differences in the computation of the relevance across norms, we calculated new relevance values for the features of all the three norms. In particular, the semantic relevance value for each feature-concept pair was calculated as $k_{ij} = x_{ij} \times \log10(I/I_j)$, where x_{ij} indicates the dominance of a feature j in a concept i and represents the local component of k_{ij}, I indicates the total number of concepts in the corresponding norms (i.e., 541 concepts in McRae et al., 2005; 120 concepts in Montefinese et al., 2013; 400 concepts in Vivas et al., 2017), I_j indicates the number of concepts in which the feature j appears, and $\log10(I/I_j)$ represents the global component of k_{ij}. Next, to test the generalizability of the featural informativeness measures, we compared each measure for the concepts shared between each language pair (ENG vs ITA, ENG vs SPA, ITA vs SPA), by using Pearson's correlations. Next, we tested domain-based differences across the different languages on informativeness measures by means of a series of two-tailed independent *t*-tests. The false discovery rate (FDR) correction was applied at $p = 0.05$, with the procedure described by Benjamini and Hochberg (1995), to correct for multiple comparisons. To investigate better the significant domain-based differences across languages, we carried out by-items factorial analysis of variance (ANOVA) on the mean values of each informativeness measure with language (1) ENG vs. ITA; 2) ENG vs. SPA; 3) ITA vs. SPA) and semantic domain (artefact vs. natural). Post hoc Tukey's tests were used when necessary.

Results

All correlations are shown in Table 2, and indicate that almost all the featural informativeness measures were significant and generalizable across different languages. Both distinctiveness and cue validity presented modest correlations between norms (distinctiveness range: .46 to .57 and cue validity range: .48 to .58). Both dominance and relevance measures presented more variability, as revealed by lower correlations (dominance range: .11 to .48 and relevance range: .32 to .45). Note that for dominance, the correlation between the English and Spanish languages resulted in an unexpected lower, and not significant, linear relationship ($N = 214$; $r = .11$; $p = .113$).

== Please, insert Table 2 ==

We then tested language-related differences in the featural salience measures for each semantic domain (artefact vs natural). To this end, we performed a series of two-tailed *t*-tests for independent samples. Table 3 shows the statistics and the cross-language comparisons for artefact and natural domains. This analysis revealed that when English and Spanish languages were contrasted with the Italian language the concepts of both domains were different for dominance and relevance (see Table 3). This difference was due to the higher mean values of dominance and relevance for English and Spanish compared with the Italian language for the concepts belonging to the artefact and natural domains (see Table 3 for the descriptive statistics). However, the ANOVAs on dominance and relevance values did not confirm this result. Indeed, they revealed no significant language by domain interaction (all $F\text{s}_{(1,1561)} > .53$, $p\text{s} > .47$, $\eta^2_p\text{s} > .001$).

Instead, when the English language was contrasted with the Italian language, distinctiveness and cue validity showed a significant difference for the natural domain compared to the artefact domain. It is worth noting that the ANOVAs on distinctiveness and cue validity corroborated this result with a significant interaction between language (ENG vs. ITA) and domain (artefact vs. natural) ($F\text{s}_{(1,1769)} < 5.89$, $p\text{s} < .01$, $\eta^2_p\text{s} > .01$) (see Table 3). This two-way interaction was due to the fact that the significant effect of language ($F\text{s}_{(1,1769)} < 6.81$, $p\text{s} < .01$, $\eta^2_p\text{s} > .01$) was modulated by the semantic domain. Indeed, post-hoc analyses revealed that this interaction was driven by the higher values of distinctiveness and cue validity for the Italian compared to the English in the natural domain ($p\text{s} < .02$). In contrast, no difference emerged between English and Italian for the artefact domain ($p\text{s} > .99$).

A different picture emerged for the contrast between the English and Spanish languages. Indeed, results showed that no significant domain-related difference emerged between the English and Spanish concepts for dominance, distinctiveness and cue validity, whereas the relevance dimension was significantly different in both artefact ($t_{(127)} = 38.11$; $p < .001$; $d = 3.16$) and natural ($t_{(83)} = 33.43$; $p < .001$; $d = 3.29$) domains. However, this pattern of results was not confirmed by the ANOVAs, suggesting that this result should be taken with caution. Indeed, no significant two-way interaction with language (ENG vs. SPA) and domain factors has been revealed ($F_{(1,4837)} = .28$, $p = .60$, $\eta^2_p\text{s} < .01$).

== Please, insert Table 3 ==

Discussion

In this section we assessed the consistency of the salience measures across three languages: English, Italian and Spanish by means of correlational analysis, t test comparisons and analysis of variance. The overall pattern of results showed that almost all the salience measures were consistent across languages, revealing a good generalizability across different languages. In line with a moderate view of the linguistic relativity theory (e.g. Slobin, 1996; Montefinese et al., 2018; Vigliocco et al., 2005), this result confirms that semantic measures that are more concept-related, rather than word-related, are not usually influenced by language differences. Along the same line of reasoning, previous studies (Brysbaert et al. 2014; Montefinese et al. 2014b) reported high cross-language correlations for other concept-related measures, like concreteness and imageability, across Dutch, English, Italian, Spanish and German. In contrast, word-related measures, as the familiarity, showed a lower correlation because are more language specific and consequently vary considerably from one language to another (Eilola and Havelka, 2010; Montefinese et al., 2014b). Note that the relevance measure showed a similar trend across languages, although it was only marginally significant. This result can be due to the fact that, although core features tend to remain stable across languages, some cultural differences could affect the informativeness of certain features in conceptual representation of a given language (as shown in the previous sections).. As we mentioned above, this measure is related to the number of participants who list a feature for a given concept and it is indirectly related to the number of features produced by the participants. For this reason, this non significant correlation may be driven by the concepts with a very low number of features for Spanish language. Indeed, when we remove from the correlation the common concepts with 2-3 features (e.g., the concept lantern has 12 features in English and only 2 in Spanish) the correlation is significant ($N = 208$; $r = .16$; $p = .02$). More importantly, it should be noted that for the common concepts with the Italian, in Spanish there was no concept with 2-3 features. However, although significant, the correlation is very low. Therefore, we think that other factors could affect the correlation between Spanish and English for the dominance measure. Further investigation is needed to clarify this aspect.

We also observed that salience measures for both the artefact and natural domains are distributed in a similar way across the three languages, confirming the stability of semantic representation across languages. Interestingly, we found a modulation of the language type (English vs. Italian) on the semantic domain when distinctiveness and cue validity were considered. In particular, we found a significant difference in the distinctiveness and cue validity measures between English and Italian only for the natural domain. This result could be due to differences between the two original databases. Specifically, as compared to the English norm, the Italian one includes a higher number of concrete concepts belonging to the categories of natural domain, but which in turn, are composed of a smaller number of concepts. As a consequence, it is less likely for features to be shared across concrete concepts, because the concepts belonging to the same category usually share a greater number of features.

Further significant cross-language differences have been found by means of the *t* tests comparisons regarding relevance. However, these results should be taken with caution because they have not been replicated by the analysis of variance and they merits further investigation.

General discussion

The aim of this study was to identify equivalences across three languages (Italian, English and Spanish) in feature composition obtained through a feature-listing task for a set of concrete concepts. The research questions addressed by this study have the singularity of covering both the common aspects that constitute the core features for the three languages, as well as addressing the idiosyncratic aspects of each of them. To the best of our knowledge, studies that carried out cross-linguistic comparisons making use of semantic feature norms are scarce (Kremer and Baroni, 2011; Vivas et al. 2019, submitted). Kremer and Baroni compared Italian and German focusing particularly on the kind of features produced and some salience measures. In Vivas et al.'s paper, they compared Spanish and English performing a vector comparison similar to the one presented here, but they did not include Italian and did not consider salience measures. As we intended to show in this paper, the information provided by feature listing tasks can be very informative regarding concept's representation. However, caution must be taken when comparing sets of feature norms from different languages extracted from different experimental settings. In the current paper, by using different analytical approaches, our results showed that informativeness and similarity measures can provide very valuable information about the importance of specific features to the core meaning of the concepts in spite of interlinguistic variability.

In particular, results from the feature vector comparison analysis suggest that the three languages tend to have more similarities than differences between the feature compositions. However, some differences emerged across languages. These can be due to differences in the context of use of the objects for each language/culture of the speakers that produced the features. These differences are expected if we consider that conceptual representation is built from our daily experience with objects and that the semantic features are an abstraction from verbal and non-verbal information we receive (Hoffman et al., 2018; Yee and Thompson-Schill, 2016), as well as from the language we speak (e.g., Winawer et al., 2007; Casasanto 2008). Thus, the differences between languages are a logical consequence of the different environments, uses and habits with those objects. As people from different cultures have different interactions with the same concept (as the example of the concept horse mentioned above) and encounter it on different contexts (in the case of horse a field, a farm, a racetrack), it is very likely that this would be reflected in feature composition. In spite of that, a set of common core features is essential to understand each other and to allow a concept to be translated to another language. Also, as Yee and Thompson-Schill (2016) suggested, "the significant commonalities in human experience (and especially within a given culture) would lead different individuals' representations (and their labels) to be similar enough for most practical purposes" (p. 1024).

There were also observed some differences regarding the number of features produced in the feature-listing task, with English participants producing more features. Although the feature collection was very similar across the three norms, we suspect that there are some cultural differences in the way participants were committed to the task that was reflected in the number of features produced.

Moreover, a chi-square analysis on the proportion of feature type produced by each language showed that all languages produced most sensorial features. This kind of information is acquired via our direct physical experience with the object at reference and it is an essential part of the concept representation. According to *attribute-based models* (Borghesani and Piazza, 2017), sensorial, jointly with functional features, constitute motor-sensorial features which are a fundamental element of concepts representation. Additionally, many studies have demonstrated that these features became active when we think about the object (e.g. Martin, 2007). Concept representation also include other kinds of features like encyclopaedic and categorizing (superordinate category), but in the case of concrete objects sensorial features are more abundant due to the access to direct experience and manipulacion of those objects.

These across-language similarities were also supported by correlational analyses and *t*-test comparisons, showing that feature informativeness measures are consistent across-languages (and semantic domains). Some language-dependent effects we found could be attributed to methodological differences between norms (see Discussion section of the Section 4). These results are in line with a study comparing semantic descriptions across German, Italian and English participants (Kremer and Baroni, 2011). Indeed, the authors showed that distinctiveness, cue validity and production frequency were far from being significantly different across languages.

We performed additional analyses, which have stressed furtherly the commonalities in concept representations across languages. For example, representational analysis revealed that the degree of resemblance between concepts is represented in a similar way across-languages. In addition, results from a cluster analysis and the calculation of the density coefficient for each language network showed that the organization of concepts in groups was very similar across languages, thus supporting Malt and colleagues' proposal (1999) that there are categorization similarities across languages. Although some authors propose that the selection of the properties that are more relevant in order to assign an object to a category is modulated by language (Zannino et al. 2015), it seems that the three languages considered in the current paper (all of them Western languages) lead to similar categorizations. However, some differences were observed such as the diversity in dispersion, reflected in density coefficient values. As we mentioned above, the difference in the dispersion of clusters, particularly between Spanish and Italian, can probably be due to the use of taxonomic features. Spanish speakers produced those features with higher frequency and that would contribute to establish more cohesive groups of concepts.

Overall, although some of our results are in line with recent proposals within the framework of the linguistic relativity theory suggesting that language can bias conceptual representation, some others indicate that there is a part of concept representation (the core features) that is stable across languages. Indeed, although the feature listing task is foremost a verbal task, and it has been found in previous studies language-dependent effect for verbal tasks, we found that this task mostly tap into language-independent representations of concepts.

Specifically, our results have shown a strong support for the idea that there are some core features that are shared across individuals and languages and constitutes the core meaning. We understand that shared meaning, is not limited to denotative aspects of communication, but it plays a crucial role in reciprocal understanding. Communication between people also involves sharing contextual aspects of categorization and possible inferences that emerge from the properties of the features, and producing similar deductions and analogies that are based on the common semantic features that characterize the concepts. That is why core features must be shared between speakers of a language to allow communication and also across languages allowing translations to be made. However, there are also features that are very relevant in one language due to the daily use of the concept or other cultural variables but that are not shared across languages. These features that are not shared with other cultures, constitute a quite stable aspect of the concept representation for the corresponding language and not for the others (that is why they appeared in the feature norms). Besides, there are also idiosyncratic features, that constitute individual's representations, that are not shared even within the culture, and that can change across time. That is why we agree with Yee and Thompson-Schill (2016) in that concepts representations are dynamic, experience-based and that different features can be activated depending on the contextual needs.

Object meaning does not constitute an emergent product of its intrinsic properties. It is not an entity that has to be discovered as an inherent property. On the contrary, the meaning is an emergent from the weighted sum of all the relations that the subject have had on his/her history with the object at issue. Thus, the meaning constitutes a social and historical construction. This construction implies the semiotic regulation that is the result of the interaction with others and includes and interpretation process performed by the actors involved. Consequently, the meaning of a symbol cannot be seen as a static entity, not even permanent, due to the fact that people and communities modify it according to the circumstances, needs, interests, convenience and ideology. Nevertheless, we also consider that core features are essential components of the concept representation. They allow the external referent to be identified and categorized and make the object distinguishable from others. These kind of features are shared within and between cultures and make translation and communication possible. Our results endorse this idea by showing commonalities in feature composition across Spanish, Italian and English and equivalences in feature informativeness measures.

There are some final considerations that have to be made regarding the analysis we performed and the material we used. It is important to consider that we had the limitation that concepts included in the three norms were not the same across the three languages so we worked with a limited sample (43 common across the three databases). Also, the conditions in which data collection was made for the original feature listing tasks, although the same instructions were given, was not identical. This can constitute a bias on the way people respond regarding kind and amount of features.. Additionally, hard work has to be performed in every semantic feature production norm that is the unification of the label used to express the same feature. For example, the feature *has a handle* can be expressed as has_a_handle, has_handle, with_handle and the researcher must select one individual label to express the idea. A similar processing has to be done when comparing norms across languages. Although consensus criteria are applied, the task is very artisanal. It is important to mention that there is a consensus work in progress across the research groups that have been working on semantic norms to establish some protocols or guidelines to collect feature norms in order to avoid these methodological differences (a concrete example can be seen in this same issue of the journal on the paper from Buchanan, De Deyne, and Montefinese, 2019). Finally, it shall be observed that within the semantic feature norms paradigm, words are typically presented to participants in isolation, that is, without a context or situation that can help them disambiguate subtle semantic differences. Additionally, the features that emerge as relevant on one task (with higher values of informativeness) perhaps are not produced as more relevant in other tasks and contexts. However, this hypothesis is beyond the aim of the current paper Further studies have to be performed in order to determine if core features are always activated when thinking of a concept although they are not openly express in certain task or the feature importance can change according to the task demands (see Lebois, Wilson-Mendenhall and Barsalou, 2014 and Vivas, Manoiloff, García, Lizarralde & Vivas, 2019 for two alternative perspectives).

Conclusions

The current article is part of the special topic titled "Eliciting Semantic Properties: Methods and Applications". The goal of this special topic is to investigate and extend the use of property listing tasks and semantic feature norming studies, by illustrating their breadth of applications, as well as highlighting the theoretical and methodological issues faced by this paradigm, which may limit its usefulness.

Our paper addressed the issue of cross-language comparison within the semantic feature paradigm, and investigated the extent to which the existing semantic features databases are comparable to one another. This study illustrates some interesting techniques that can be applied in order to compare standard norms collected through a feature-listing task. The results showed strong similarities in feature composition and concept organization across three languages. With our study, we therefore contributed to highlight the applicability of such paradigm across languages, and at the same time we provided some observations that can be used to formulate further research questionshypotheses within this theoretical framework. For example, given the subtle differences between conceptual representations reported in a qualitative manner in our study, it might be interesting to investigate, in further studies, whether native speakers and foreign language learners produce similar features in response to the same concepts, expressed in the target language. Another possible path of research could be that of comparing the cross-linguistic comparability of feature-based conceptual representations for concrete vs. abstract concepts, given that the latter type of concepts appears to be more tightly related to language-specific factors.

Additionally, the results of this study illustrate the importance to use local semantic features norms to obtain values of semantic variables to avoid cultural bias in the measures selected.

Compliance with Ethical Standards

Funding: This work was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 702655 and by the University of Padua (SID 2018) to MM, by the National Agency of Scientific and Technological Promotion

(Argentina), grant PICT 2015-0983 and the National University of Mar del Plata to LV and JV, and by the Creative Multilingualism research program to which MB is affiliated, which is funded by the Arts and Humanities Research Council as part of the Open World Research Initiative.

Ethical Approval: This article does not contain any studies with human participants performed by any of the authors. The materials used were already published semantic norms.

Conflict of Interest: The authors declare that they have no conflict of interest.

References

Baroni, M, Barbu, E, Murphy, B, and Poesio, M (2010) Strudel: A distributional semantic model based on properties and types. Cogn. Sci., 34:222–254. https://doi.org/10.1111/j.1551-6709.2009.01068.x

Benjamini, Y, and Hochberg, Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological), 57(1):289-300.

Bolognesi M (2017) Using semantic features norms to investigate how the visual and verbal modes afford metaphor construction and expression. Lang Cogn, 9(3):525-552. https://doi.org/10.1016/S0022-5371(78)90119-6

Borgatti, SP (2002) NetDraw: Graph Visualization Software. Harvard: Analytic Technologies

Borgatti, SP, Everett, MG and Freeman, LC (2002) Ucinet 6 for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Borghesani,V, Piazza, M (2017) The neuro-cognitive representations of symbols: the case of concrete words. Neuropsychologia, 105:4-17. https://dx.doi.org/10.1016/j.neuropsychologia.2017.06.026

Brysbaert, M, Stevens, M, De Deyne, S, Voorspoels, W, and Storms, G (2014) Norms of age of acquisition and concreteness for 30,000 Dutch words. Acta Psychol., 150:80-84. https://doi.org/10.1016/j.actpsy.2014.04.010

Brysbaert, M, Fias, W, Noel, MP (1998) The Whorfian hypothesis and numerical cognition: Is “twenty-four” processed in the same way as “four-and-twenty”? Cognition, 66:51-77. https://dx.doi.org/10.1016/S0010-0277(98)00006-7

Buchanan, E. M., De Deyne, S., & Montefinese, M. (2019). A practical primer on processing semantic property norm data. *Cognitive processing*, 1-13.

Buchanan, EM, Holmes, JL, Teasley, ML, and Hutchison, KA (2013) English semantic word-pair norms and a searchable Web portal for experimental stimulus creation. Behav Res 45, 3:746-757. https://doi.org/10.3758/s13428-012-0284-z

Buchanan, EM, Valentine, KD, and Maxwell, N (2019) English Semantic Feature Production Norms: An Extended Database of 4,436 Concepts. Behav Res. https://doi.org/10.3758/s13428-019-01243-z

Bourne, LE Jr, and Restle, F (1959) Mathematical theory of concept identification. Psychol. Rev., 66(5):278-296. https://psycnet.apa.org/doi/10.1037/h0041365

Casasanto, D. (2008). Who's afraid of the big bad Whorf? Crosslinguistic differences in temporal language and thought. Language learning, 58, 63-79. https://doi.org/10.1111/j.1467-9922.2008.00462.x

Catricalà E, Della Rosa PA, Plebani V, Perani D, Garrard P, Cappa SF (2015) Semantic feature degradation and naming performance. Evidence from neurodegenerative disorders. Brain Lang. 147, 58–65. https://doi.org/10.1016/j.bandl.2015.05.007

Chaigneau, S, Canessa, E, Barra, C, Lagos, R (2018) The role of variability in the property listing task. Beh Res, 50:972-988. https://doi.org/10.3758/s13428-017-0920-8

Cree, GS, and McRae, K (2003) Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). JEP: General, 132:163-201. http://dx.doi.org/10.1037/0096-3445.132.2.163

Cree, GS, McNorgan, C, and McRae, K (2006) Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. J Exp Psychol Learn Mem Cogn, 32(4):643. http://dx.doi.org/10.1037/0278-7393.32.4.643

De Deyne, S, Verheyen, S, Ameel, E, Vanpaemel, W, Dry, MJ, Voorspoels, W, and Storms, G (2008) Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. Behav Res, 40:1030–1048. https://doi.org/10.3758/BRM.40.4.1030

Devereux, BJ, Tyler, LK, Geertzen, J, and Randall, B (2014) The Centre for Speech, Language and the Brain (CSLB) concept property norms. Behav Res, 46(4):1119-1127. https://doi.org/10.3758/s13428-013-0420-4

Dubossarsky, H, De Deyne, S, and Hills, TT (2017) Quantifying the structure of free association networks across the lifespan. Dev. Psychol., 53(8):1560-1570. https://psycnet.apa.org/doiLanding?doi=10.1037%2Fdev0000347

Eilola, TM, and Havelka, J (2010) Affective norms for 210 British English and Finnish nouns. Behav Res, 42(1):134-140. https://doi.org/10.3758/BRM.42.1.134

Garrard, P, Lambon Ralph, MA, Hodges, JR, and Patterson, K (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. Cogn Neuropsychol, 18:125-174. http://dx.doi.org/10.1080/02643290042000053

Gentner, D, and Goldin-Meadow, S (2003) Whither Whorf. In D Gentner and S Goldin-Meadow (Eds.), Language in Mind: Advances in the study of language and thought (3-14). Cambridge, MA: MIT Press.

Hoffman, P, McClelland, JL, and Lambon Ralph, M (2018) Concepts, Control, and Context: A Connectionist Account of Normal and Disordered Semantic Cognition. Psychol. Rev., 125:293-328. http://dx.doi.org/10.1037/rev0000094

Johnson, S C (1967). Hierarchical Clustering Schemes. Psychometrika, 32:241-253. https://doi.org/10.1007/BF02289588

Kintsch, W. (2001) Predication. Cogn. Sci.,25:173–202. https://doi.org/10.1207/s15516709cog2502_1

Kremer, G and Baroni, M (2011) A set of semantic norms for German and Italian. Behav Res, 43, 1:97-109. https://doi.org/10.3758/s13428-010-0028-x

Lebois LA, Wilson-Mendenhall CD, Barsalou LW (2014) Are Automatic Conceptual Cores the Gold Standard of Semantic Processing? The Context-Dependence of Spatial Meaning in Grounded Congruency Effects. Cogn Sci. 39(8):1764-801. https://doi.org/10.1111/cogs.12174

Lenci, A, Baroni, M, Cazzolli, G, y Marotta, G (2013) BLIND: a set of semantic feature norms from the congenitally blind. Behav Res. 45, 4:1218–1233. https://doi.org/10.3758/s13428-013-0323-4.

Levinson, SC (1996) Language and space. Annu. Rev. Anthropol., 25:353-382. https://doi.org/10.1146/annurev.anthro.25.1.353

Malt B, Sloman S, Gennari S, Shi M, Wang Y (1999) Knowing versus naming: similarity and the linguistic categorization of artifacts. J Mem Lang., 40:230–262. https://doi.org/10.1006/jmla.1998.2593

Malt, B, Sloman, S, and Gennari, S (2003) Universality and language specificity in object naming. J Mem Lang., 49:20-42 http://dx.doi.org/10.1016/S0749-596X(03)00021-4

Martin, A. (2007). Conceptual representation in the brain. Annual Review of Psychology, 58, 25-45. https://doi.org/10.1146/annurev.psych.57.102904.190143

Meteyard, L, Rodríguez Cuadrado, S, Bahrami, B and Vigliocco, G. (2010) Coming of age: A review of embodiment and the neuroscience of semantics. Cortex 48(7):788-804. · https://doi.org/10.1016/j.cortex.2010.11.002

McRae, K, Cree, G, Seidenberg, M and McNorgan, C (2005) Semantic feature production norms for a large set of living and nonliving things. Behav Res, 37:547-559. https://doi.org/10.3758/BF03192726

Moldovan, CD, Ferré, P, Demestre, J, and Sánchez-Casas, R (2015) Semantic similarity: Normative ratings for 185 Spanish noun triplets. Behav Res, 47:788-799. https://doi.org/10.3758/s13428-014-0501-z

Montefinese, M, Ambrosini, E, Fairfield, B, and Mammarella, N (2013a) Erratum to: Semantic memory: A feature-based analysis and new norms for Italian. Behav Res, 45:462-462. https://doi.org/10.3758/s13428-012-0263-4

Montefinese, M, Ambrosini, E, Fairfield, B and Mammarella, N (2013b) Semantic memory: A feature-based analysis and new norms for Italian. Behav Res, 45, 2:440-461. https://doi.org/10.3758/s13428-012-0291-0

Montefinese, M, Ambrosini, E, Fairfield, B, and Mammarella, N (2014a) Semantic significance: a new measure of feature salience. Mem. Cogn., 42(3), 355-369. https://doi.org/10.3758/s13421-013-0365-y

Montefinese, M, Ambrosini, E, Fairfield, B, and Mammarella, N (2014b) The adaptation of the affective norms for English words (ANEW) for Italian. Behav Res, 46(3):887-903. https://doi.org/10.3758/s13428-013-0405-3

Montefinese, M, Ambrosini, E, and Roivainen. E (2018) No grammatical gender effect on affective ratings: evidence from Italian and German languages. Cogn Emot. 33(4):848-854. https://doi.org/10.1080/02699931.2018.1483322

Montefinese, M, Zannino, GD., and Ambrosini, E (2015) Semantic similarity between old and new items produces false alarms in recognition memory. Psychol. Res., 79(5):785-794. http://dx.doi.org/10.1007/s00426-014-0615-z

Montefinese, M, Vinson, D, and Ambrosini, E (2018) Recognition memory and featural similarity between concepts: the pupil's point of view. Biol. Psychol., 135:159-169. https://doi.org/10.1016/j.biopsycho.2018.04.004

Montefinese M, and Vinson D (2017) Resemblance among similarity measures in semantic representation. In G. Gunzelmann, A. Howes, T. Tenbrink, and E. J. Davelaar (Eds.), Proceedings of the 39th Annual Conference of the Cognitive Science Society (p. 3787). Austin TX: Cognitive Science Society.

Rosch E and Mervis C (1975) Family Resemblances: Studies in the Internal Structure of Categories, Cogn. Psychol 7: 573-605 https://doi.org/10.1016/0010-0285(75)90024-9

Ruts, W, De Deyne, S, Ameel, E, Vanpaemel, W, Verbeemen, T, and Storms, G (2004) Dutch norm data for 13 semantic categories and 338 exemplars. Behav Res, 36:506-515. http://dx.doi.org/10.3758/BF03195597

Sartori, G, Lombardi, L, and Mattiuzzi, L (2005) Semantic relevance best predicts normal and abnormal name retrieval. Neuropsychologia 43:754–770. http://dx.doi.org/10.1016/j.neuropsychologia.2004.08.001

Sartori, G, Polezzi, D, Mamelia, F, and Lombardi, L (2005) Feature type effects in semantic memory: An event related potentials study. Neurosci. Lett. 390:139–144. https://doi.org/10.1016/j.neulet.2005.08.015

Sartori, G and Lombardi, L (2004) Semantic relevance and semantic disorders. J. Cogn. Neurosci., 16:439-452. https://doi.org/10.1162/089892904322926773

Slobin, DI (1996) From "Thought and Language" to "Thinking for Speaking". In: JJ Gumperz and SC Levinson (Eds.) Rethinking Linguistic Relativity. Cambridge (MA): Cambridge University Press, pp. 70-96.

Smith, EE, Shoben, EJ, and Rips, LJ (1974) Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 81:214-241. http://dx.doi.org/10.1037/h0036351

Smith, EE, and Medin, DL (1981) Categories and concepts. Cambridge, MA: Harvard University Press.

Taylor, KI, Moss, HE and Tyler, LK (2007) The conceptual structure account: a cognitive model of semantic memory and its neural instantiatios. In: J Hart y M Kraut (Eds.) Neural basis of semantic memory. Cambridge University Press.

Ursino, M, Cupini C, Cappa S, Catricalà E (2018). A feature-based neurocomputational model of semantic memory. Cognitive Neurodynamics. https://doi.org/10.1007/s11571-018-9494-0

Vinson, DP, and Vigliocco, G (2008) Semantic feature production norms for a large set of objects and events. Behavior Research Methods, 40 (1):183-190. https://doi.org/10.3758/BRM.40.1.183

Vigliocco, G, Vinson, DP, Lewis, W, and Garrett, MF (2004) Representing the meanings of object and action words: The featural and unitary semantic system hypothesis.Cogn. Psychol, 48(4):422–488. http://dx.doi.org/10.1016/j.cogpsych.2003.09.001

Vigliocco, G, Vinson, DP, Paganelli, F, and Dworzynski, K (2005) Grammatical gender effects on cognition: Implications for language learning and language Use. JEP: General, 134(4):501–520. http://dx.doi.org/10.1037/0096-3445.134.4.501

Vigliocco G, Meteyard L, Andrews M, Kousta S (2009) Toward a theory of semantic representation. Lang Cogn.1:219–247. https://doi.org/10.1515/LANGCOG.2009.011

Vivas, J, Lizarralde, F, Huapaya, C, Vivas, L y Comesaña, A (2014) Organización reticular de la memoria semántica. Natural Finder y Definition Finder, dos métodos informatizados para recuperar conocimiento. Encontros Bibli 19, 40:235-252. http://dx.doi.org/10.5007/1518-2924.2014v19n40p235

Vivas, J, Vivas, L, Comesaña, A, García Coni, A, and Vorano, A (2017) Spanish semantic feature production norms for 400 concrete concepts. Behav Res 49, 3:1095–1106. https://doi.org/10.3758/s13428-016-0777-2

Vivas, L, Manoiloff, L, García, A, Lizarralde, F, Vivas, J (2019) Core semantic links or lexical associations: Assessing the nature of responses in word association tasks. J. Psycholinguist. Res. 48:243-256. https://doi.org/10.1007/s10936-018-9601-8

Warrington, EK, and McCarthy, RA (1987) Categories of knowledge further fractionation and an attempted integration. Brain 110:1273-1296. https://doi.org/10.1093/brain/110.5.1273

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., and Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. Proc. Natl. Acad. Sci. U.S.A. 104, 7780–7785. https://doi.org/10.1073/pnas.0701644104

Wu, L and Barsalou, W (2010) Perceptual simulation in conceptual combination: Evidence from property generation. Acta Psychol (Amst) 132:173–189. doi: 10.1016/j.actpsy.2009.02.002

Yee, E and Thompson-Schill, S (2016) Putting concepts into context. Psychon Bull Rev, 23:1015–1027. https://doi.org/10.3758/s13423-015-0948-7

Zannino GD, Perri R, Pasqualetti P, Caltagirone C, Carlesimo GA (2006) Analysis of the semantic representations of living and nonliving concepts: a normative study. Cogn Neuropsychol 23:515–40. https://doi.org/10.1080/02643290542000067

Zannino GD, Caltagirone C, Carlesimo GA (2015) The contribution of neurodegenerative diseases to the modelling of semantic memory: A new proposal and a review of the literature. Neuropsychologia. 75:274-90. https://doi.org/10.1016/j.neuropsychologia.2015.06.023