

ORIGINAL ARTICLE

# Semantic feature salience matters: new values of significance and relative weight for the Spanish feature norms for young and older adults

Leticia Vivas<sup>1</sup>, Matías Yerro Avincetto<sup>1,2</sup>, Sofía Romanelli<sup>3</sup>, Francisco Lizarralde<sup>4</sup> and Jorge Ricardo Vivas<sup>1</sup>

<sup>1</sup>IPSIBAT (UNMDP-CONICET), Facultad de Psicología, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina, <sup>2</sup>Facultad de Ingeniería, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina, <sup>3</sup>Instituto de Humanidades y Ciencias Sociales (INHUS, UNMDP/CONICET), Facultad de Humanidades, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina and <sup>4</sup>Facultad de Ingeniería, Universidad Nacional de Mar del Plata, Mar del Plata, Argentina

**Corresponding author:** Leticia Vivas; Email: [lvivas@mdp.edu.ar](mailto:lvivas@mdp.edu.ar)

(Received 26 July 2023; revised 22 May 2024; accepted 15 August 2024)

## Abstract

To determine the semantic importance of certain features has been a topic of interest for many semantic memory models. Diverse measures have been proposed in the last years. Semantic significance, the latest measure proposed by Montefinese and colleagues, is both sophisticated and comprehensive. Given the cultural and linguistic variability of semantic measures, this study presents values of significance and relative weight for the Spanish-speaking population corresponding to 400 concrete concepts. First, we presented data for both young and older adults. Second, we assessed the effect of significance on response times in two speeded feature verification tasks. Third, we compared significance values with the existing Italian significance norms. To evaluate the effect of significance to predict response times, two speeded verification tasks (Experiment 1 and 2) were carried out, selecting a total of 130 concepts for analysis. In Experiment 1, subjects were presented with a feature followed by a concept, while in Experiment 2, the order of stimulus presentation was reversed (i.e., the concept was presented before the feature). An independent linear mixed model showed that significance was a good predictor of response latencies in Experiments 1 and 2. Moreover, results revealed a strong positive correlation between the Spanish and Italian significance values. Findings are discussed in terms of recent theories of semantic cognition.

**Keywords:** Relative weight; semantic features; semantic significance; Spanish norms

## Introduction

Feature models of semantic memory assume that concept representation is composed of a set of features or properties (Jackson et al., 2015; Martin, 2016; Soshi et al., 2017). However, the degree of contribution of these features to the definition of different concepts is not homogeneous. An important number of features constitute the representation of a concept. These features are derived from verbal and nonverbal experience, both individual and culturally mediated, always involved in certain context and circumstances, that the person has had with the object to which the concept refers. Some of these features are closer to the core meaning of concepts because they are more relevant to identify those concepts. These properties are crucial for facilitating communication among individuals within a particular community. For example, when the concept *giraffe is mentioned*, speakers of the same language typically think of an animal with a long neck. Moreover, research has demonstrated that these features are shared even across different languages (Vivas et al., 2020). Additionally, there are other properties that members of a community may share but are not essential for defining the concept (e.g., *eats leaves for giraffe*). And there are even other features which are idiosyncratic and specific to an individual's particular representation. This last group includes creative (e.g., *uses a scarf*) and bizarre features (e.g., *I can climb it to get to the roof*), as well as misconceptions (e.g., *it's oviparous*). These three types of features can be seen as continuous, ranging from the more central to the more peripheral feature of a concept's meaning (for further discussion on this topic, see Vivas et al., 2021). The current article will focus on those features closer to the core meaning, which are fundamental for mutual comprehension.

Several measures have been proposed to characterize the importance of a feature for the central meaning of a concept. Cue validity was the first measure assessed when Rosch and Mervis (1975), Rosch (1978) studied the internal structure of categories. Cue validity was proposed and is useful to calculate the conditional probability that an object belongs to a category if it has a particular characteristic. Into a features norm it is calculated as the production frequency of the feature divided by the sum of the production frequencies of that feature for all the concepts in which it appears (McRae et al., 2005). For example, the feature *flies* has a production frequency of 25 in the English norms and a total production frequency of 712, so the corresponding value is .035. The highest value would be 1 if the feature is only produced for that concept. More recently, cue validity has been shown to be a valuable indicator in different processes, for instance, in the processing of words and non-words during reading (Tiwari et al., 2020), in changes in attentional selection (Lou et al., 2022), and to predict the impact of the emotional content of images (Denefrio et al., 2017).

Lately, attention has focused on feature distinctiveness, those characteristics that are unique to a particular concept. Feature distinctiveness has been proposed as a good measure to examine category-specific semantic deficits (Devlin et al., 1998), to characterize prototypicality (Garrard et al., 2001), and to study its influence on language acquisition, visual lexical decision, and semantic decision (Siew, 2020; 2021). It is calculated as "the inverse of the number of concepts in which the feature appears in the norms" (McRae et al., 2005, p.552). For example, the feature *flies*

appears in 46 concepts in the English norms, so the calculus would be  $1/46$  obtaining a value of .022 of distinctiveness. Here again a value of 1 would mean that the feature is highly distinctive for that concept, as it does not appear in other concepts. This concept was also central to theories explaining category-specific semantic deficits in neurological patients. For instance, Conceptual Structure Account (CSA) (Taylor et al., 2007) proposes that living and nonliving things differ in the degree to which their distinctive and shared features are correlated with other features of the concept. Distinctive features of living things tend to be weakly related to other features of the concept (e.g., *has stripes*), making them more vulnerable to the effects of brain damage. In contrast, distinctive features of nonliving things tend to be highly correlated with other important features (e.g., *used for cutting* and *has a handle*). This idea was further discussed in a paper by Duarte and colleagues (Duarte et al., 2009), who studied this hypothesis in Alzheimer's disease (AD). They observed that distinctive features appear to be affected regardless of the domain (living and nonliving things). This same result was obtained by Catricalà and colleagues (Catricalà et al., 2015). Additionally, Duarte et al. (2009) noted that in moderate stages of the disease, distinctive features of living things tend to be more affected.

More recently, *relevance* was proposed as a more elaborated measure combining *dominance* (i.e., production frequency in feature production norms) with *distinctiveness* (Sartori & Lombardi, 2004). According to these authors, the former is considered the local component, indicating the importance of the feature for the concept, while the latter, known as the global component, constitutes an index of the importance of that feature for the rest of the concepts included in the norms' database. As an example, the feature makes honey has a production frequency (dominance) of 27 in the Spanish norms and a distinctiveness value of 1 for the concept bee leading to a value of 233.384 according to the relevance formulae. Mechelli et al. (2006) suggest that semantic relevance explains category effects in medial fusiform gyri. Additionally, this measure has proven to be the best predictor of name retrieval accuracy in a naming-to-description task. Sartori et al. (2005) showed that, for both normal participants and those with AD with semantic impairment, the semantic relevance of the concept description predicts response accuracy in name retrieval better than distinctiveness and dominance of the same description. However, it is important to note that semantic significance had not been defined at that time and, as we will see later, this measure adds a small adjustment to semantic relevance. More recently, evidence has shown that damage to relevant features in AD and the semantic variant of Primary Progressive Aphasia tends to be highly correlated with failures on picture naming tasks (Catricalà et al., 2015).

The newest variable proposed to measure feature importance is *significance*. This metric was developed by Montefinese and colleagues (Montefinese et al., 2013) and posteriorly, Montefinese et al. (2014) tested the predictive value with speeded verification tasks. It constitutes a combination of *accessibility* (*dominance* plus order of production in feature norms) and *distinctiveness*. As can be seen, this parameter adds the order of production to the values already included in *relevance* (dominance and distinctiveness). The order of production provides additional and pertinent information regarding the weight of the feature for that concept in a particular language community. For example, for the concept *eagle* the feature "*flies*" tends to

be produced earlier, while “*lays eggs*” generally appears in the fourth or fifth position in the participant’s production in feature norms. The above-mentioned authors have provided evidence that *significance* is a good predictor of subjects’ verification latency in a feature verification task (i.e., subjects have to judge whether a particular feature is related to a certain concept).

Although some of these measures are relatively stable across languages (Vivas et al., 2020) differences can also be observed in several concepts (Perniss et al., 2012; Vivas et al., 2020). For example, the concept *turkey* (“pavo” in Spanish) has a very different meaning for an Argentinean individual compared to an American English individual, as for the latter it is related to Thanks giving Day. This is the reason why local norms must be developed and, particularly, this is the main aim of the current paper: to present values of semantic significance for the Spanish-speaking population. Besides, in order to compare semantic significance in relation to alternative measures of feature importance, we analyzed its effect on participant’s response speed in two feature verification tasks (feature-concept and concept-feature) and compared its effect with the current gold standard measure, *relevance*. In the third place, a comparison was performed with the Italian significance values.

### Semantic significance calculation

To achieve the main aim of the current paper, the values of semantic significance for the entire database of 400 concepts from the Argentinean Spanish feature production norms (Vivas et al., 2017) comprising 3071 features were calculated (see Supplementary Material). The formula employed here is quite similar to the one used by Montefinese et al. (2014, p. 358). The main difference lies in the calculus of *accessibility*. While they calculated the *centile order of production plus dominance* (see Montefinese et al., 2013, p.445), our calculus included a measure called *relative weight* (RW). To obtain that measure, the following procedure is performed: each participant (which will be labeled with subscript  $j$ ) has produced a list of  $n_j$  features to describe the concept. Each of these features is given a weight determined by its position in the vector of features, divided by the length of the vector. Hence, the first feature will have a weight equal to 1 (i.e.,  $n_j/n_j$ ), the second one  $(n_j-1)/n_j$  and the  $i$ -th feature, starting from the one with weight 1, will have a weight of  $(n_j-(i-1))/n_j$ . After this process, for each feature that was produced by more than one participant, a sum is performed over the respective weights and a new vector is thus formed (again, ordered from higher to lower values). The resulting vector is then normalized with respect to the standard Euclidean norm, obtaining a new vector whose entries are real numbers between 0 and 1. This procedure is performed by the Definition Finder software (Vivas et al., 2014), which can be downloaded from the [OSF](#).

The final formula used to calculate significance is the following:

$$\text{Sig} = 10 * \text{RW} * \text{LN}(400 / \text{CPF})$$

where RW refers to relative weight, LN to natural logarithm, and CPF to the number of concepts in which the feature appears. Four hundred refers to the full set of concepts available in the norms. The values of *relevance* and *distinctiveness* were extracted from the Spanish semantic feature production norms both for young

(Vivas et al., 2017) and older adults (Vivas et al., 2022) following the formulae of Sartori et al. (2005) and McRae et al. (2005), respectively. We also included other salience measures already published in Vivas, J. et al. (2017) and Vivas, L. et al. (2022) (see Supplementary Material). It is worth noting that, as we already had data from Spanish semantic feature production norms for both young and older adults, we decided to calculate semantic significance for both groups. We believe this could offer a valuable resource for researchers requiring centrality measures for the elderly too, given the documented differences in semantic memory between both populations (Mirasso et al., 2022; White et al., 2018; Yoon et al., 2004). However, the speeded verification task which will be elaborated on in the following paragraphs was exclusively performed with young adults due to their greater availability for data collection.

### **Comparison with semantic relevance**

To achieve the secondary aim of the current paper, we studied the potential of the *significance* measure to predict response times through a feature verification experiment. We compared this new measure to another well-established measure: *semantic relevance* (Sartori & Lombardi, 2004). These experiments were carried out solely with younger adults.

The complete experimental procedure consists of two experiments consisting of speeded feature verification tasks. In Experiment 1, participants performed the task with the presentation of a feature followed by a concept (feature-concept condition). In Experiment 2, the order was reversed, with the concept followed by a feature (concept-feature condition). These two alternatives were proposed because it is likely that the activation spreading differs when the task starts with a feature (e.g., “flies,” which can refer to either a living or a nonliving object) compared to when it starts with a concept (e.g., *plane*) (Ramscar et al., 2010; Ursino et al., 2013).

## **Experiment 1**

### **Method**

#### *Participants*

One hundred and twenty participants (96 females and 24 males; mean age = 24.7 years;  $SD = 5.22$ ) took part in Experiment 1. They were all undergraduate students at the National University of Mar del Plata or young professionals. All subjects reported having normal or corrected-to-normal vision. Eighty-five percent of them reported being right-handed. The presence of neurological or psychiatric diseases was an exclusion criterion. Moreover, they provided written informed consent before participating in the study. This study adhered to the Helsinki principles (World Medical Association, 2013) and was approved by the Ethical Committee of the National University of Mar del Plata.

#### *Materials*

Stimuli and semantic measures were selected from the Spanish semantic feature production norms (Vivas et al., 2017). From the full set of 400 concepts, we selected

a subset of 130 according to the following criteria: a) the concept was not a compound word; b) it was not polysemic; c) it belonged to a well-defined semantic category (i.e., easily identified by most people); d) the categories had more than four exemplars; and finally e) they were included in McRae et al.'s (2005) set of concepts (just in case some author would be interested in future comparisons).

The stimuli were the same for both experiments and consisted of 130 concrete concepts belonging to 11 categories (animals, tools, fruits, musical instruments, furniture, buildings, clothing, vegetables, utensils, accessories, and vehicles). They were all concrete (mean concreteness = 4.78, min = 4.39, max = 4.97 according to the Argentinian norms by Manioff and colleagues (Manioff et al., 2010)). The mean familiarity in a scale from 1 to 5 was 2.79 (our data ranged from 1.31 to 4.91) (Manioff et al., 2010) and the mean subjective lexical frequency on a scale from 1 to 5 was 2.68 (our data ranged from 1.08 to 5) (Martínez-Cuitiño 2015). Each of the 130 concepts was paired with four features: one core feature, one partially shared feature, one idiosyncratic feature, and a non-related feature. The procedure to operationalize these levels can be seen in Vivas et al (2021). So, 520 combinations were obtained. Four lists were elaborated in order to provide participants trials with an acceptable number of items. In order to do that concept-feature pairs were distributed taking into account that each concept appeared only once in each list and distributing equivalently the kind of features (core, partially shared, idiosyncratic, and non-related). Each contained 120 experimental concept-feature pairs plus 10 concept-feature pairs for the training session. The final lists can be seen in OSF.

*Procedure.* Subjects were placed at 60 cm in front of an LCD 17 computer screen (1280 × 960 pixels). Tasks were displayed using E-prime 2.0. Subjects were instructed to press the right button if the feature was reasonably true for that concept and the left button if it was not. Responses were considered correct if the subject responded “true” for those related to the concept. Filler trials (i.e., unrelated features) were discarded.

Trials began with a set of 10 training items to ensure that the participants were familiarized with the task, followed by the 120 experimental items. Each trial consisted of a feature presented for 2000 ms, followed by a fixation cross (inter-stimulus interval—ISI) for 500 ms, and then a concept that remained on the screen for 4000 ms or until the participant's response. The inter-trial interval (ITI) was 1500 ms (see Figure 1). Features were presented in black, and concepts were presented in blue. Both were written in lowercase fonts (Verdana, 24 pt), on a white background.

### **Statistical analysis**

The following analyses were performed using R version 4.4.0 (R core team, 2021). Reaction times (RTs) greater than 3000 ms and lower than 500 ms were removed as well as erroneous responses (14.08 % for Experiment 1). The number of data points considered for Experiment 1 (feature-concept) was 6186. Each numeric variable was centered and scaled, except for RTs. Following Baayen and Milin (2010), RTs were log-transformed (Ln\_RT) to meet normality assumptions ( $Z = 1.057$ ;  $p = .214$  for Experiment 1).

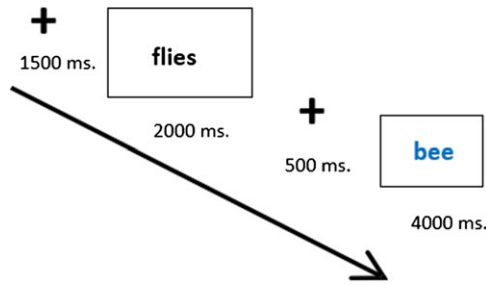


Figure 1. Example of items presentation.

As the data from the experiments violate the independence assumption due to both participants and concepts grouping the data points, simple regressions could not be applied. Therefore, the next step was to perform independent linear mixed model analyses for each measure to identify the most explanatory variable, considering  $\text{Ln\_RT}$  as the dependent variable. As the correlation between significance and relevance is  $r = 0.93$ , independent analyses were performed for significance and relevance to avoid collinearity. The formula used for each linear mixed model was

$$\text{LogRT} \sim \text{VAR} + (1|\text{Subj}) + (1|\text{Conc})$$

where “VAR” is replaced with Significance or Relevance, according to their specific model. Subjects and Concepts were considered as random factors, and their intercepts were taken into account by the models. To compare the explanatory power of relevance and significance in RTs between models, the Akaike information criterion (AIC) was used.

## Results

The results for both models (relevance and significance) for Experiment 1 (feature-concept) are shown in Table 1. Both models show statistically significant fixed effects (intercept and significance/relevance) based on the t-values. The significance model’s intercept has an estimate of 2.9832, and the significance term has an estimate of  $-0.0037$ . The relevance model’s intercept has an estimate of 2.9897, and the relevance term has an estimate of  $-0.0006$ . While the significance and relevance terms are statistically significant, the R-squared values imply that other factors not included in the models may be contributing to the variability in the outcome variable. In this case, individual RTs and complex feature-concept pairs may be contributing with added variability into the results.

To assess the difference in predictive power between significance and relevance, model comparisons were made using the AIC. Table 2 contains the comparative values of the model. While the AIC criterion might favor the significance model, the R-squared values from Table 1 suggest that both models explain a relatively small proportion of the variance. Therefore, the differences between models are not big enough to evaluate one predictive variable as better than the other. This suggests that both significance and relevance have similar predictive power in explaining RTs in Experiment 1.



### **Discussion of experiment 1**

Results indicated that both measures have the same predictive power for the feature-concept condition. Our results are in line with those obtained by Montefinese *et al.* (2013) in that both studies observed an effect of semantic significance in the feature verification task. However, they observed a superiority effect of significance compared to relevance, which was not observed in our experiment. Methodological differences in the calculation of significance between the two studies may explain this discrepancy, and these differences will be discussed in detail in the general discussion.

## **Experiment 2**

### **Method**

#### *Participants*

A group of 120 people participated in Experiment 2 (87 females and 33 males; mean age = 24.43 years;  $SD = 5.25$  years). They were all undergraduate students at the [removed for review] or young professionals. All subjects reported having normal or corrected-to-normal vision. The presence of neurological or psychiatric diseases was an exclusion criterion. Moreover, they provided written informed consent before participating in the study. This study followed Helsinki principles (World Medical Association, 2013).

### **Materials**

The materials for Experiment 2 were identical to those used in Experiment 1, but in reversed order.

#### *Procedure*

The procedure for Experiment 2 was identical to Experiment 1, except that the concept was presented prior to the feature description.

### **Statistical analysis**

The statistical analysis for Experiment 2 was the same as Experiment 1. For Experiment 2, 8.58 % of the responses were removed for the reasons mentioned above. The number of data points considered for Experiment 2 (Concept-Feature) was 6582.

Each numeric variable was centered and scaled, except for RTs. Following Baayen and Milin (2010), RTs were log-transformed ( $\text{Ln\_RT}$ ) to meet normality assumptions ( $Z = 0.696$ ;  $p = .718$  for Experiment 2).

### **Results**

Table 3 contains the model estimates for both variables, relevance and significance, for Experiment 2 (concept-feature). Both the significance model and relevance model show statistically significant results for the intercept and the concept-feature interaction term (significance/relevance) based on their t-values. In the significance



**Table 1.** Model Estimates for Experiment 1 (Feature-Concept)

| Significance Model  |             |                     |          |             |          |          |
|---------------------|-------------|---------------------|----------|-------------|----------|----------|
| Random Effects      |             |                     |          |             |          |          |
| Concept (Intercept) |             | Subject (Intercept) |          |             | Residual |          |
| Variance            | Std.Dev.    | Variance            | Std.Dev. | Variance    | Std.Dev. |          |
| 0.0022              | 0.0476      | 0.0032              | 0.0571   | 0.0112      | 0.1061   |          |
| Fixed Effects       |             |                     |          |             |          |          |
|                     | Estimate    | Std. Error          | t value  | correlation | IC 2.5%  | IC 97.5% |
| (Intercept)         | 2.9832      | 0.0079              | 374.59   | –           | 2.9675   | 2.9988   |
| Significance        | –0.0037     | 0.0001              | –21.45   | –0.322      | –0.0041  | –0.0034  |
| R2                  | Conditional |                     |          | Marginal    |          |          |
|                     | 0.06        |                     |          | 0.37        |          |          |
| Relevance Model     |             |                     |          |             |          |          |
| Random Effects      |             |                     |          |             |          |          |
| Concept (Intercept) |             | Subject (Intercept) |          |             | Residual |          |
| Variance            | Std.Dev.    | Variance            | Std.Dev. | Variance    | Std.Dev. |          |
| 0.0021              | 0.0467      | 0.0032              | 0.0571   | 0.0113      | 0.1066   |          |
| Fixed Effects       |             |                     |          |             |          |          |
|                     | Estimate    | Std. Error          | t value  | correlation | IC 2.5%  | IC 97.5% |
| (Intercept)         | 2.9897      | 0.0081              | 369.2    | –           | 2.9738   | 3.0056   |
| Relevance           | –0.0006     | 0.00003             | –20.1    | 0.378       | –0.0007  | –0.0005  |
| R2                  | Conditional |                     |          | Marginal    |          |          |
|                     | 0.05        |                     |          | 0.36        |          |          |

**Table 2.** Model comparison between predictive power of significance and relevance for experiment 1

| Experiment 1 (Feature-Concept) |      |         |         |        |          |
|--------------------------------|------|---------|---------|--------|----------|
| Model                          | npar | AIC     | BIC     | logLik | deviance |
| 1-Sig                          | 5    | –9625.6 | –9591.9 | 4817.8 | –9635.6  |
| 2-Rel                          | 5    | –9573.4 | –9539.7 | 4791.7 | –9583.4  |

model, the intercept has an estimated value of 3.0421 with a standard error of 0.0105. The significance term itself has an estimated value of –0.0030 and a standard error of 0.0002. Similarly, the relevance model shows a significant intercept (estimate: 3.0400, standard error: 0.0001) and a significant relevance term (estimate: –0.0005, standard error: 0.0000). However, the conditional R-squared values for both models are low (0.03 for significance model and 0.02 for relevance model),

**Table 3.** Model Estimates for Experiment 2 (Concept-Feature)

| Significance Model  |             |                     |          |             |          |          |
|---------------------|-------------|---------------------|----------|-------------|----------|----------|
| Random Effects      |             |                     |          |             |          |          |
| Concept (Intercept) |             | Subject (Intercept) |          | Residual    |          |          |
| Variance            | Std.Dev.    | Variance            | Std.Dev. | Variance    | Std.Dev. |          |
| 0.00173             | 0.04164     | 0.00995             | 0.09977  | 0.01313     | 0.11463  |          |
| Fixed Effects       |             |                     |          |             |          |          |
|                     | Estimate    | Std. Error          | t value  | correlation | IC 2.5%  | IC 97.5% |
| (Intercept)         | 3.04212     | 0.01051             | 289.5    | –           | 3.02147  | 3.06278  |
| Significance        | –0.00303    | 0.00018             | –16.7    | –0.259      | –0.00339 | –0.00267 |
| R2                  | Conditional |                     |          | Marginal    |          |          |
|                     | 0.03        |                     |          | 0.48        |          |          |
| Relevance Model     |             |                     |          |             |          |          |
| Random Effects      |             |                     |          |             |          |          |
| Concept (Intercept) |             | Subject (Intercept) |          | Residual    |          |          |
| Variance            | Std.Dev.    | Variance            | Std.Dev. | Variance    | Std.Dev. |          |
| 0.00179             | 0.0424      | 0.00995             | 0.0997   | 0.0132      | 0.1151   |          |
| Fixed Effects       |             |                     |          |             |          |          |
|                     | Estimate    | Std. Error          | t value  | correlation | IC 2.5%  | IC 97.5% |
| (Intercept)         | 3.04E+00    | 1.07E-02            | 284.98   | –           | 3.02247  | 3.0644   |
| Relevance           | –4.89E-04   | 3.377E-05           | –14.49   | –0.302      | –0.00055 | –0.00042 |
| R2                  | Conditional |                     |          | Marginal    |          |          |
|                     | 0.02        |                     |          | 0.47        |          |          |

indicating that these models explain a relatively small portion of the outcome variable's variance.

To assess the difference in predictive power between significance and relevance, model comparisons were made using the AIC as in Experiment 1. Table 4 contains the comparative values of the model. The differences between models are not large enough to evaluate one predictive variable as better than the other. This indicates that both significance and relevance demonstrate comparable predictive power in explaining RTs in Experiment 2 (concept-feature). While the AIC criterion seems to show that the significance model has a better fit, the R2 (as shown in Table 3) may be implying that this model is only marginally better than the relevance model.

### **Discussion of experiment 2**

In Experiment 2, we once again observed the same effect as in Experiment 1. Both measures, significance and relevance, demonstrated an effect on the concept feature condition of the feature verification task; however, we did not observe a superiority

**Table 4.** Model comparison between predictive power of significance and relevance, for experiment 2

| Experiment 2 (Concept-Feature) |      |         |         |        |          |
|--------------------------------|------|---------|---------|--------|----------|
| Model                          | npar | AIC     | BIC     | logLik | deviance |
| 1-Sig                          | 5    | -9143.3 | -9109.3 | 4576.6 | -9153.3  |
| 2-Rel                          | 5    | -9076.5 | -9042.5 | 4543.2 | -9086.5  |

**Table 5** Comparison between Spanish and Italian significance values

|         | Mean   | SD    | Min-Max      | 2.5th Percentile | 97.5th Percentile |
|---------|--------|-------|--------------|------------------|-------------------|
| Italian | 14.087 | 6.191 | 1.923-39.301 | 3.704            | 28.985            |
| Spanish | 10.093 | 7.366 | 1.09-58.54   | 2.4              | 30.30             |

of semantic significance. Combining frequency and order of production for each participant and in a population requires additional work on the raw data. However, the order of elicitation of a trait represents, in some way, the connectivity that exists between a concept and said reference trait. For this reason, we expect significance to be a better variable to evaluate the effect of feature saliency in different tasks. However, our findings showed that significance did not prove to be a better predictor of feature verification than relevance. Perhaps, surely, priming does not work in a similar way in production tasks and in property verification tasks, and this processing system is responsible for not allowing us to see, in this task, the predictive capacity of significance, which it undoubtedly exists at production time. Cognitive and neuropsychological differences have been pointed out in the distinction between identification and production tasks as part of implicit memory processes (Prull & Spataro, 2017); however, it is a topic that has not yet been clearly elucidated.

#### *Comparison with Italian measures*

Finally, to validate the significance values for the Spanish norms, a correlation analysis comparing these values to those for the Italian norms (Montefinese et al., 2014) was carried out. To this end, common concepts between the norms were identified ( $N = 65$ ), and a total of 245 shared features were selected. The results showed a strong positive correlation between them [ $r = .592$ ;  $p < .001$ ]. Descriptive values for Spanish and Italian norms are presented in Table 5.

## **General discussion**

One of the main contributions of the current study is to provide values of semantic significance for the Spanish-speaking population of both young and older adults, for the set of 400 concrete concepts included in the Spanish feature production norms for both young (Vivas et al., 2017) and older adults (Vivas et al., 2022). This provides the scientific community with new empirically obtained measures of feature importance.

As regards the feature-concept experiments, the analyses indicated that both variables, relevance and significance, explained RTs to the same extent. These results

indicate that semantic significance is a good predictor of response latencies in the feature verification task. However, contrary to the findings of Montefinese *et al.* (2014), we did not observe a superiority of this measure over relevance, considering the delta R<sup>2</sup> between both models in both experiments. Similarly, in the concept-feature experiment, similar results were obtained, indicating that significance is a good predictor of RT but not superior to relevance. One explanation for these results could be the differences between semantic significance calculations, as we employed a different formula from Montefinese. Additionally, semantic significance was calculated based on the total number of concepts in each case. It is relevant to highlight that Montefinese's norms included 120 concepts, while Vivas' norms included 400. Therefore, it is likely that the significance values will differ to some extent. Also, the variables included in the mixed effects models in each study were not exactly the same, adding another potential source of difference.

Besides, the significance values for the Spanish norms were found to be equivalent to those reported for the Italian population, showing significant correlations between shared features. Additionally, descriptive values, including mean, standard deviation, minimum, maximum, and extreme percentile values, were similar for both norms. It is worth observing here that both norms comprised a different number of concepts: 400 for Spanish and 120 for Italian. Moreover, both studies differed in the number of participants per concept. In this sense, the observed correlations might have been stronger in the absence of these differences.

Finally, some limitations should be acknowledged. First, this experiment was not originally designed to measure the effect of significance but rather to identify differences in response latencies based on the continuous degrees of meaning representation. Therefore, the presence of different degrees of strength in the link between feature and concept (e.g., idiosyncratic features with a very weak relation to the concept) may have delayed response times because the decision requires additional effort. Future research should address this issue by proposing feature-concept pairs that elicit univocal responses, such as true and false features that vary in terms of their degree of significance.

Secondly, it is relevant to acknowledge that significance is a relative measure of feature salience, similar to many other variables derived from feature norms. It is calculated considering the number of concepts included within those norms (in our case, 400), so it is relative to this specific corpus of concepts. Indeed, the degree of distinctiveness of a feature, which is included in the calculus of significance, may vary if the set of concepts is enlarged. It should be noted that semantic feature norms only include a subset of all possible concepts found in a language. Therefore, adding more concepts to the norms will modify the values of these measures. For example, within the category of tools, if the majority of exemplars are cutting tools, adding other types of tools to the norms will likely affect the values of cutting for the different tools.

## Conclusion

In the current study, we extend the use of a new measure of feature salience, semantic significance, originally proposed for the Italian population, to the Spanish-speaking population. Unlike previous measures such as production frequency,

distinctiveness, and order of production, this measure incorporates additional information to calculate the importance of a feature. Therefore, it is expected to be a better variable for assessing the effect of feature salience in different tasks.

However, our findings indicate that significance did not prove to be a better predictor of feature verification than relevance. Further experiments should be conducted to validate its explanatory capacity in different tasks.

**Replication package.** Replication data and materials for this article can be found at <https://osf.io/kyjb3/>.

## References

- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12–28.
- Catricalà, E., Della Rosa, P. A., Plebani, V., Perani, D., Garrard, P., & Cappa, S. F. (2015). Semantic feature degradation and naming performance. Evidence from neurodegenerative disorders. *Brain and Language*. <https://doi.org/10.1016/j.bandl.2015.05.007>
- Denefrio, S., Simmons, A., Jha, A., Dennis-Tiwary, T. A. (2017). Emotional cue validity effects: the role of neurocognitive responses to emotion. *PLoS One*, 12(7), e0179714.
- Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage. *Journal of Cognitive Neuroscience*, 10(1), 77–94.
- Duarte, L. R., Marquié, L., Marquié, J. C., Terrier, P., & Ousset, P. J. (2009). Analyzing feature distinctiveness in the processing of living and non-living concepts in Alzheimer's disease. *Brain and Cognition*. <https://doi.org/10.1016/j.bandc.2009.04.007>
- Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, 18(2), 125–174.
- Jackson, R. L., Hoffman, P., Pobric, G., & Ralph, M. A. L. (2015). The nature and neural correlates of semantic association versus conceptual similarity. *Cerebral Cortex*. <https://doi.org/10.1093/cercor/bhv003>
- Lou, H., Lorist, M. M., Pilz, K. S. (2022). Effects of cue validity on attentional selection. *Journal of Vision* 11, 22(8), 15.
- Manoiloff, L., Artstein, M., Canavoso, M. B., Fernández, L., & Seguí, J. (2010). Expanded norms for 400 experimental pictures in an Argentinean Spanish-speaking population. *Behavior Research Methods*, 42(2), 452–460.
- Martin, A. (2016). GRAPES—Grounding representations in action, perception, and emotion systems: how object properties and categories are represented in the human brain. *Psychonomic Bulletin and Review*. <https://doi.org/10.3758/s13423-015-0842-3>
- Martínez-Cuitiño, M., Barreyro, J. P., Wilson, M., & Jaichenco, V. (2015). Nuevas normas semánticas y de tiempos de latencia para un set de 400 dibujos en español. *Interdisciplinaria*, 32(2), 289–305.
- McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, 37(4), 547–559.
- Mechelli, A., Sartori, G., Orlandi, P., Price, C. J. (2006). Semantic relevance explains category effects in medial fusiform gyri. *Neuroimage*, 15, 30.
- Mirasso, V., Inveninato, M., Savastano, R. & Vivas, L. (2022). Caracterización de las representaciones semánticas en personas mayores. *Investigaciones en Psicología*, 25(2), 37–45.
- Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2013). Semantic memory: a feature-based analysis and new norms for Italian. *Behavior Research Methods*, 45(2), 440–461.
- Montefinese, M., Ambrosini, E., Fairfield, B., & Mammarella, N. (2014). Semantic significance: a new measure of feature salience. *Memory and Cognition*, 42(3), 355–369.
- Perniss, P., Vinson, D., Seifart, F., & Vigliocco, G. (2012). Speaking of shape: the effects of language-specific encoding on semantic representations. *Language and Cognition*, 4(3), 223–242.
- Prull, M. & Spataro, P. (2017). Editorial: the role of the distinctions between identification/production and perceptual/conceptual processes in implicit memory: findings from cognitive psychology, neuroscience and neuropsychology. *Frontiers in Psychology*, 8, 1129.

- R Core Team** (2021). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ramscar, M., Yarlett, D., Dye, M., Denny, K., & Thorpe, K.** (2010). The effects of feature-label-order and their implications for symbolic learning. *Cognitive science*, *34*(6), 909–957.
- Rosch, E.** (1978). Principles of Categorization. In E Rosch and BB Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum.
- Rosch, E., & Mervis, C. B.** (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology*, *7*(4), 573–605.
- Sartori, G., & Lombardi, L.** (2004). Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience*, *16*(3), 439–452.
- Sartori, G., Lombardi, L., & Mattiuzzi, L.** (2005). Semantic relevance best predicts normal and abnormal name retrieval. *Neuropsychologia*, *43*(5), 754–770.
- Siew, C.S.Q.** (2020). Feature distinctiveness effects in language acquisition and lexical processing: insights from mega studies. *Cognitive Processing*, *21*, 669–685.
- Siew, C.S.Q.** (2021). Global and local feature distinctiveness effects in language acquisition. *Cognitive Science*, *45* (7), e13008.
- Soshi, T., Fujimaki, N., Matsumoto, A., & Ihara, A. S.** (2017). Memory-based specification of verbal features for classifying animals into super-ordinate and sub-ordinate categories. *Frontiers in Communication*, *2*(September), 12.
- Taylor, K. I., Moss, H. E., & Tyler, L. K.** (2007). The conceptual structure account: A cognitive model of semantic memory and its neural instantiation. In *Neural Basis of Semantic Memory*. <https://doi.org/10.1017/CBO9780511544965.012>
- Tiwari, P., Mishra, T., Singh, T., Singh, A., Tiwari, T. & Singh, I.** (2020). Cue validity effects on word and non-word processing during reading. *Journal of the Indian Academy of Applied Psychology*, *46*, 240–246.
- Ursino, M., Cuppini, C., & Magosso, E.** (2013). The formation of categories and the representation of feature saliency: analysis with a computational model trained with an Hebbian paradigm. *Journal of Integrative Neuroscience*, *12*(04), 401–425.
- Vivas, J., Kogan, B., Romanelli, S., Lizarralde, F., & Corda, L.** (2020). A cross-linguistic comparison of Spanish and English semantic norms: looking at core features. *Applied Psycholinguistics*, *41*(2), 285–297.
- Vivas, J., Kogan, B., Yerro, M., Romanelli, S. & Vivas, L.** (2021). Describing the structure of concepts through different feature levels. *Journal of Cognitive Psychology*, *33*, 1.
- Vivas, J., Lizarralde, F., Huapaya, C., Vivas, L., & Comesaña, A.** (2014). Organización reticular de la memoria semántica. Natural Finder y Definition Finder, dos métodos informatizados para recuperar conocimiento. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, *19*(40), 235–252.
- Vivas, J., Vivas, L., Comesaña, A., Coni, A. G., & Vorano, A.** (2017). Spanish semantic feature production norms for 400 concrete concepts. *Behavior Research Methods*, *49*(3), 1095–1106.
- Vivas, L., Montefinese, M., Bolognesi, M. & Vivas, J.** (2020). Core features: measures and characterization for different languages. *Cognitive Processing*, *21*, 651–667.
- Vivas, L., Yerro M, Romanelli S, García Coni A, Comesaña A, Lizarralde F, Passoni I, Vivas J.** (2022). New Spanish semantic feature production norms for older adults. *Behaviour Research Methods*, *54*(2), 970–986.
- White, A., Storms, G., Malt, B. C., & Verheyen, S.** (2018). Mind the generation gap: differences between young and old in everyday lexical categories. *Journal of Memory and Language*, *98*, 12–25.
- World Medical Association.** (2013). World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *Jama*, *310*(20), 2191–2194.
- Yoon, C., Feinberg, F., Hu, P., Gutchess, A. H., Hedden, T., Chen, H. Y. M., . . . & Park, D. C.** (2004). Category norms as a function of culture and age: comparisons of item responses to 105 categories by American and Chinese adults. *Psychology and Aging*, *19*(3), 379.

---

**Cite this article:** Vivas, L., Yerro Avincetto, M., Romanelli, S., Lizarralde, F., & Vivas, J. R. (2024). Semantic feature salience matters: new values of significance and relative weight for the Spanish feature norms for young and older adults. *Applied Psycholinguistics*. <https://doi.org/10.1017/S014271642400033X>