

Semantic memory: A feature-based analysis and new norms for Italian

Maria Montefinese · Ettore Ambrosini · Beth Fairfield · Nicola Mammarella

Published online: 9 October 2012
© Psychonomic Society, Inc. 2012

Abstract Semantic norms for properties produced by native speakers are valuable tools for researchers interested in the structure of semantic memory and in category-specific semantic deficits in individuals following brain damage. The aims of this study were threefold. First, we sought to extend existing semantic norms by adopting an empirical approach to category (Exp. 1) and concept (Exp. 2) selection, in order to obtain a more representative set of semantic memory features. Second, we extensively outlined a new set of semantic production norms collected from Italian native speakers for 120 artifactual and natural basic-level concepts, using numerous measures and statistics following a feature-listing task (Exp. 3b). Finally, we aimed to create a new publicly accessible database, since only a few existing databases are publicly available online.

Keywords Semantic norms · Feature listing · Conceptual representations · Accessibility · Significance

Semantic memory is our knowledge about the world. It is the system that stores and organizes the representation and processing of semantic knowledge (Tulving, 1972)—that is, internal representations about words, things, and their properties that allow us to understand and interact with the world around

Electronic supplementary material The online version of this article (doi:10.3758/s13428-012-0263-4) contains supplementary material, which is available to authorized users.

M. Montefinese (✉) · B. Fairfield · N. Mammarella
Department of Psychological Sciences, University of Chieti,
Via dei Vestini 31,
66100 Chieti, Italy
e-mail: m.montefinese@unich.it

E. Ambrosini
Department of Neuroscience and Imaging and Institute for
Advanced Biomedical Technologies (ITAB), University of Chieti,
Chieti, Italy

us and with the objects present in it. This knowledge largely derives from daily observation of living and nonliving things, from interacting with them, and from witnessing others' interaction with them from birth on, although it can also be acquired culturally or by talking and reading about concepts. Importantly, semantic knowledge plays a central role in both online and offline processing of the environment: Online, it guides perception, categorization, and inferences, and offline, it reconstructs memories, underlies the meanings of linguistic expressions, and provides the representations manipulated by thought (Barsalou, Simmons, Barbey, & Wilson, 2003).

Semantic features are assumed to be the building blocks of semantic knowledge by a variety of theories (e.g., Collins & Quillian, 1969; Jackendoff, 1990, 2002; Minsky, 1975; Norman & Rumelhart, 1975; Saffran & Sholl, 1999; Smith & Medin, 1981) and by computational models of conceptual representation (e.g., Caramazza & Shelton, 1998; Farah & McClelland, 1991; Humphreys & Forde, 2001). Furthermore, some of these computational models of semantic representations, in which concepts are described as distributed patterns of activation across units representing semantic features, have proposed that at least a subset of these features are modality-specific (e.g., Cree & McRae, 2003; Plaut, 2002; Rogers et al., 2004; Vigliocco, Vinson, Lewis, & Garrett, 2004)—that is, grounded in sensory–motor processing—as has been advocated by the embodied-cognition view on mental representation (see, e.g., Barsalou, 1999, 2008; Grush, 2004) and confirmed by neurophysiological studies (Martin & Chao, 2001; Vigliocco et al., 2006).

Some models have used semantic feature norms obtained from naive participants to construct empirically derived conceptual representations that can be used to test theories about semantic representation and computation (e.g., Cree, McNorgan, & McRae, 2006; Cree & McRae, 2003; McRae, de Sa, & Seidenberg, 1997; Garrard, Lambon Ralph, Hodges, & Patterson, 2001; Moss, Tyler, & Devlin, 2002; Rogers & McClelland, 2004). In this empirical approach to

the study of semantic knowledge, participants are presented with a set of concept names and are asked to produce the features that they feel best describe each of the concepts. From the data obtained, it is then possible to calculate numerous measures and distributional statistics at both the feature and concept levels, such as estimates of semantic similarity/distance between concepts, or measures of how features of various knowledge types are distributed across concepts. Several other measures can also be used to increase the information obtained from the norms themselves, such as collecting them either by means of participants' explicit ratings, as in the case of familiarity or typicality ratings, or using language corpora, as in the case of word frequency or the linguistic variables associated with a concept's name.

These measures and statistics allow researchers to construct stimuli for further experiments while controlling for nuisance variables; to model human behavior in computational simulation models of semantic knowledge; and to test theories about semantic memory and its deficits. Many studies have, in fact, used feature representations deriving from feature production norms to account for empirical phenomena, such as semantic priming (Cree, McRae, & McNorgan, 1999; McRae et al., 1997; Vigliocco et al., 2004), feature verification (Ashcraft, 1978; McRae, Cree, Westmacott, & de Sa, 1999; Solomon & Barsalou, 2001), categorization (Hampton, 1979; Smith, Shoben, & Rips, 1974), and category learning (Kruschke, 1992), among others (for a more detailed discussion of the aims and limits of feature norms, see McRae, Cree, Seidenberg, & McNorgan, 2005; for a discussion of theories of semantic memory organization that use feature representations to account for category-specific semantic deficits, see Cree & McRae, 2003; Zannino, Perri, Pasqualetti, Caltagirone, & Carlesimo, 2006).

The present study

The principal aims of the present study were the following. First, we aimed to extend existing sets of semantic feature production norms. Second, we aimed to make the categories and concepts included in the database more representative of our participants' semantic memory by adopting empirical methods to choose categories (Exp. 1) and concepts (Exp. 2), while in Experiment 3, a large sample of Italian native speakers were asked to list the features of each of the resultant concepts (a property generation task). In most normative studies, researchers have typically designated categories and their exemplars a priori on the basis of intuition or by selecting them from previously published databases. In this way, participants are made to list characteristics for concepts that may not be most representative of the structure of their conceptual knowledge. Indeed, some

authors (e.g., Bueno & Megherbi, 2009; Van Overschelde, Rawson, & Dunlosky, 2004) have questioned the appropriateness of earlier databases (especially the norms of Battig & Montague, 1969) and have highlighted the fact that semantic norms rapidly become obsolete, due to the speed of language (and cultural) changes. More importantly, to the best of our knowledge, no previous study has actually controlled, or even measured, the representativeness of the categories or concepts used in creating norms.

Third, we described the measures and statistics included in our norms and briefly discussed their application in cognitive neuroscience and neuropsychology, in line with the idea of creating a freely accessible public database for researchers interested in testing theories about semantic memory. Semantic norms are useful and powerful tools for the research community, and although several normative databases currently exist, only a few research groups have made their norms freely available (De Deyne et al., 2008; Garrard et al., 2001; McRae et al., 2005; Vinson & Vigliocco, 2008); to the best of our knowledge, only one set of semantic norms in Italian is freely available (Kremer & Baroni, 2011).

Before proceeding to the description of the norms, we will describe the terminology and notation that will be used throughout the article. The notion of an *exemplar* denotes a concept (i.e., member of a category), such as *cane/dog*, *coltello/knife*, and so forth, and will be printed in italics, while a <category> refers to a group of exemplars, such as <*animali/animals*>, <*utensili da cucina/kitchenware*>, and so forth, and will be printed in italics between triangular brackets. Features that participants listed in response to an exemplar, such as “*verde/green*,” “*alto/high*,” and “*ha ago/has needle*” in response to *abete/fir*, will be indicated in the text with an italic typeface between quotation marks. Note that throughout we present the Italian word first, followed by a slash and the English translation.

We will also refer to Excel files that can be found in the [Supplementary Materials](#). Four files are provided: C-C_Pairs.xls contains concept production measures; Concepts.xls contains information about individual concepts derived from the feature-listing task; C-F_Pairs.xls contains information about concept–feature pairs derived from the feature-listing task; and Distance_Matrix.xls contains the semantic distance matrices. We will refer to the variables in the files by indicating their names between square brackets (e.g., [Prod_Fr]).

Experiment 1

Experiment 1 was preliminary to the rest of the normative data collection and was designed to empirically determine the representative categories of semantic memory. Existing norms in the literature are typically based on concepts or categories that either are selected from previously published

databases (De Deyne et al., 2008; Garrard et al., 2001; Izura, Hernández-Muñoz, & Ellis, 2005; Kremer & Baroni, 2011; McRae et al., 2005; Ruts et al., 2004; van Dantzig, Cowell, Zeelenberg, & Pecher, 2011; Zannino et al., 2006) or are chosen a priori by researchers (Lynott & Connell, 2009; McRae et al., 2005; Vinson & Vigliocco, 2008). In this manner, participants describe categories or concepts that are only supposed to be representative of semantic memory, without controlling for their actual representativeness, or importance, in the semantic organization. In our study, we chose to use an empirical approach to construct our norms. Specifically, in order to determine which semantic categories were more strongly represented in our participants' semantic memory, we asked participants to generate a number of concrete concepts, without constraints.

Method

Participants A group of 435 native Italian speakers (105 male, 330 female) participated in our study. Their ages ranged from 18 to 35 years ($M = 22.7$, $SD = 3.3$), and their mean years of education were 13.7 ($SD = 2.1$). All participants were naïve as to the purpose of the study, and most of them were undergraduate psychology students from the University of Chieti.

Materials and procedure Context is essential for basic-level concept representations, since this level of abstraction is suitable for using, thinking about, or naming an object in most situations in which the object occurs. In situations in which context is not specified (e.g., in laboratory conditions), participants generally assume that the object is in a typical context (Rosch, 1978). Thus context, defined as the place in which participants perform a task, could influence the frequency and order of production of concepts in a free concept-listing task, and participants may likely take advantage of environmental clues when producing concepts. We therefore asked participants to freely produce words that referred to concrete concepts in one of three different environmental conditions—namely, a classroom, an online context, and a laboratory—in order to balance the context influences.

In all environmental conditions, we asked participants to list a minimum of 20 words that referred to concrete entities. In the classroom context, data collection took place in two large classrooms. All participants ($N = 310$ undergraduate psychology students from the University of Chieti; 51 male, 259 female; mean age $\pm SD = 21.7 \pm 2.6$ years) were present at the same time. The participants received a piece of paper with written instructions at the top, followed by 60 empty lines. They were instructed to write one word for each line, starting from the top of the sheet. In the online context, we created a form (Word document) similar to the one used in

the classroom context. These forms were then e-mailed to 150 native Italian speakers. Sixty-one participants (23 male, 38 female; mean age $\pm SD = 25.8 \pm 4.0$ years) returned the completed forms. Finally, in the laboratory context, we collected the forms in our laboratory. Sixty-four participants (31 male, 33 female; mean age $\pm SD = 24.8 \pm 3.1$ years) were asked to list concepts out loud while standing in a dark, unfurnished room. The experimenter took note of the concepts produced.

It is important to specify that while we chose to use the pencil-and-paper (classroom) condition because it has been the most used procedure in previous works, the other two environmental conditions were chosen to try to cope with the influence of context in prompting participants' responses. In particular, the controlled laboratory context was chosen because it should reduce the influence of environmental cues. In addition, since data collection is not feasible in every possible context, we chose to use the online context. Indeed, this condition, even if uncontrolled, should represent several contexts in which we daily live. In fact, the Internet is accessible from a vast variety of locations by means of smartphones, tablets, laptops, and similar devices. However, as we will discuss below, the analysis that we employed allowed us to cope with possible contextual influences.

Transcription and labeling The data consisted of a number of concepts listed by each participant, separated for context. First, we removed the abstract concepts that were occasionally listed (e.g., *amore/love* or *giustizia/justice*; less than 1 % of the total recorded data), and plurals were converted into their respective singular forms. Moreover, synonyms referring to the same concept were recorded with the same label (e.g., *gatto/cat* and *micio/kitty* were both recorded as *gatto/cat*). Next, we added information about the order of production of each word for each participant. Participants listed a different number of concepts, so we calculated a normalized order of production by converting each order-of-production value into a centile. We computed the centile order of production $C_{i,m}$ for the concept i and participant m as $C_{i,m} = [1 - (O_{i,m} - 1) / N_m]$, where $O_{i,m}$ represents the order of production of the concept i for participant m , and N_m represents the total number of concepts produced by that participant. For example, a concept listed as third by a participant who produced 16 concepts would take a centile order value of $[1 - (3 - 1) / 16] = .875$. Therefore, the centile order value assumed a value of 1 for a concept listed first by a participant and approached 0 for a concept listed last.

Finally, we classified concepts into their respective semantic categories, irrespective of their level of abstraction. In other words, words referring to superordinate, subordinate, and basic-level concepts, such as *pianta/plant*, *rosa/rose*, and *fiore/flower*, respectively, were all classified as

belonging to the <*piante*/plants> category. Concepts were classified into 33 different categories with the same abstraction level. Two native Italian speakers (authors M.M. and E.A.) independently classified all of the concepts, and a third native speaker judged disagreements. We also assessed the intercoder reliability in concept classification by asking a fourth native Italian speaker to classify the entire set of included concepts using the same set of categories that we used. Agreement between our labeling and that of the secondary rater was high, with a Cohen's κ value of .71 ($N = 1,007$, $p < .0001$). (We will use the term *intercoder reliability* or *agreement* when using the procedure outlined above).

Results and discussion

From a total of 11,357 concepts, we collected 1,007 distinct concepts belonging to 33 semantic categories. The average number of words produced per participant was 26.08 ($SD = 7.28$, range = 10 to 57). The reliability of the collected data, in terms of the categories to which concepts belonged, was evaluated by computing split-half correlations corrected with the Spearman–Brown formula after randomly dividing participants into two subgroups of equal size. This reliability index was calculated on 10,000 different randomizations of the participants, and the resultant median reliability index was very high (.994; range = .977 to .999) (unless explicitly specified, this holds true for all of the reliability estimations reported throughout this article, and we will refer to *reliability* when using the procedure outlined above).

We then calculated the mean dominance value of each category for each context as the proportion of participants who produced at least one concept belonging to that category. For example, 33 out of 64 participants listed one or more <*articoli di cancelleria*/stationery> exemplars in the laboratory context, so this category took a mean dominance value of $\approx 52\%$ (i.e., 33/64). We also calculated the mean centile order value of each category for each context. Specifically, we averaged the centile order values across all of the concepts produced in that context that belonged to that category. For example, if three <*veicoli*/vehicles> concepts were listed in the classroom context, with corresponding centile order values of .80, .63, and .37, the <*veicoli*/vehicles> category would assume a mean centile order value of .60. Note that by computing the average across participants of the mean centile values across the concepts listed by each participant, the main result of the subsequent cluster analysis would not change (i.e., with both of the averaging methods, the same categories were included in the “more representative” cluster in all three environmental conditions; see below). Note that both of these measures range from 0 to 1, with 1 corresponding to a category that was always listed as first (centile order) or

that was produced by all of the participants (dominance measure).

Subsequently, in order to choose the most representative categories of semantic memory, we conducted a K -means cluster analysis for each environmental condition. K -means clustering is a data analysis technique that seeks to segregate category sets into a predefined number of K subsets of the greatest possible distinction. This analysis classified each semantic category into one of two a priori clusters (“more representative” or “less representative”) on the basis of the mean dominance and the mean centile order values of that category. We assumed that both dominance and mean centile order values reflect the representativeness of a given concept in semantic memory.

We repeated this procedure 100 times using different random seeds to calculate cluster centroids, since K -means cluster analysis is sensitive to starting conditions (Berkhin, 2006). From the 100 solutions, the best clustering solution was selected by calculating the ratio between the between- and within-cluster variances. On the basis of these results, we then assigned each category a value of 1 if it was included in the cluster with both higher dominance and centile order values—that is, the more representative (MR) cluster—and a value of 0 if it entered in the other (less representative, LR) cluster. This procedure was repeated for each of the three context conditions. Consequently, by summing this binary measure for the three context conditions, each category assumed a comprehensive value that ranged from 3 (if that category was included in the MR cluster for all three of the contexts) to 0 (if that category was included in the LR cluster for all three of the contexts) (see Table 1). Figure 1 presents the results of the cluster analysis.

Twelve categories were not included in the MR cluster in any of the three K -means cluster analyses, and thus assumed a comprehensive inclusion value of 0. Interestingly, several of the most commonly used categories in the literature (e.g., in Battig & Montague, 1969; Bueno & Megherbi, 2009; Garrard et al., 2001; McRae et al., 2005; Van Overschelde et al., 2004; Zannino et al., 2006) were among these LR categories—namely <*frutta*/fruit>, <*strumenti musicali*/musical instruments>, <*armi*/weapons>, and <*vegetali*/vegetables>—since their concepts were listed on average by only a minority of our participants (20.1 %, 11.1 %, 8.5 %, and 6.2 %, respectively).

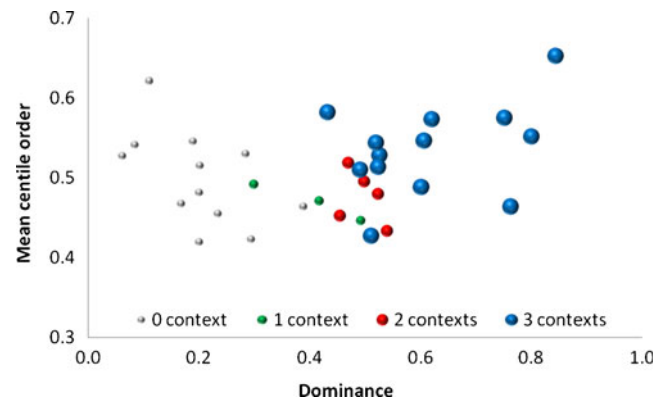
Thirteen categories, instead, were constantly produced by a high percentage of participants, irrespective of context condition, and were included in the MR cluster in all three K -means cluster analyses. Accordingly, when a category emerged in the MR cluster in only one environmental condition (thus assuming a comprehensive value of 1), we did not include it among the selected categories for subsequent experiments, because in this case it was not possible to exclude possible influences of environmental cues in

Table 1 Dominance and mean centile order values of categories, ranked by number of inclusions in the “more representative” cluster (out of three contexts)

Domain	Category	Dominance	Mean Centile	Inclusions
Artifact	Furniture	.846	.653	3
Artifact	Housing buildings	.801	.551	3
Artifact	Clothing	.764	.463	3
Artifact	Stationery	.752	.575	3
Natural	Plants	.621	.573	3
Artifact	Hobbies	.607	.546	3
Artifact	Vehicles	.602	.488	3
Natural	Body parts	.527	.528	3
Artifact	Kitchenware	.524	.513	3
Artifact	Informatics	.521	.544	3
Artifact	Furnishings/fitings	.512	.427	3
Artifact	Food	.492	.510	3
Natural	Animals	.433	.582	3
Artifact	Fashion accessories	.541	.433	2
Artifact	Leatherwork	.524	.480	2
Artifact	Electronics devices	.499	.495	2
Natural	Nature	.470	.519	2
Artifact	Tools	.455	.453	2
Artifact	Costume jewelry	.493	.446	1
Artifact	Urban buildings	.418	.470	1
Artifact	Objects	.300	.492	1
Artifact	Household appliances	.389	.464	0
Artifact	Linens	.295	.423	0
Artifact	Beverages	.285	.530	0
Artifact	Containers	.235	.455	0
Artifact	Materials	.202	.515	0
Natural	Fruit	.201	.481	0
Artifact	Cosmetics	.201	.419	0
Natural	People/person	.190	.546	0
Artifact	Sport Items	.168	.467	0
Artifact	Musical instruments	.111	.621	0
Artifact	Weapons	.085	.541	0
Natural	Vegetables	.062	.528	0

The category in bold was excluded (see the text for details).

prompting the participants’ responses. We thus assumed that the categories with a comprehensive value of 3 primarily represented semantic memory, at least in our sample. Again, it is noteworthy that participants produced a number of concepts belonging to semantic categories that are neglected in existing norms, such as <articoli di cancelleria/stationery>, <costruzioni abitative/housing buildings>, and <passatempi/hobbies>, as well as more “classical” categories such as <vestiti/clothes> and <veicoli/vehicles>. Finally, we chose to exclude one of the categories, <apparecchiatura informatica/informatics>, from Experiment 2 because it was composed of

**Fig. 1** Mean dominance and centile order values for the 33 categories of Experiment 1, as a function of the number of context conditions in which that category was more representative (MR). Each concept is plotted in the Dominance × Centile Order space (with both of these measures averaged over the three contexts), and the size and color of the corresponding points reflect the number of context conditions in which that category entered the MR cluster

a low number of exemplars (five), even though it was one of the more representative categories (see Table 1).

One of the most interesting results is that the majority of categories (26 out of 33, or 78.8 %) belong to the artifact domain, and this imbalance holds true even when we take into consideration the concepts produced instead of the categories. In fact, 84 % (9,539 out of 11,357) of the total number of concepts collected belong to the artifact domain, as well as 70.7 % (712 out of 1,007) of the distinct concepts (i.e., without considering concepts listed by more than one participant). Thus, our results suggest an asymmetrical distribution of natural and artifactual concepts in semantic memory, which is reflected in the 12 categories that we chose to use for our normative study (in fact, 75 % were artifact categories).

Experiment 2

The results from Experiment 1 indicated nine artifact (<articoli di cancelleria/stationery>, <costruzioni abitative/housing buildings>, <veicoli/vehicles>, <mobili/furniture>, <cibi/food>, <suppellettili/furnishings and fittings>, <utensili da cucina/kitchenware>, <vestiti/clothes>, and <passatempi/hobbies>) and three natural (<animali/animals>, <parti del corpo/body parts>, and <piante/plants>) categories as representative of our participants’ semantic memory. Experiment 2 was designed to investigate the distribution of concepts within each of these categories in order to select the categories’ most representative exemplars to be used in the feature-listing task (Exp. 3b). According to the prototype theory (Rosch, 1973), knowledge is clustered into semantic categories, and some members are better exemplars of a category than others. This

process, called *categorization*, is a basic mental process in cognition, and without it the world would appear chaotic (Bueno & Megherbi, 2009).

Several tasks have been used to investigate the organization of categories, such as the category verification task, in which participants have to decide whether or not a stimulus belongs to a given category (Casey, 1992; McFarland, Kellas, Klueger, & Juola, 1974), or the exemplar generation task, in which participants are asked to list as many members of a given category as possible (Bueno & Megherbi, 2009; Ruts et al., 2004; Storms, 2001; Storms, De Boeck, & Ruts, 2000; Storms, De Boeck, Van Mechelen, & Ruts, 1996). In Experiment 2, we adopted an exemplar generation task and identified four alternative measures to describe the strength of the relationship between a given category and its exemplars: exemplar generation dominance, mean rank order, first occurrence, and exemplar availability.

Method

Participants A group of 62 undergraduate psychology students from the University of Chieti (23 male, 39 female) participated in Experiment 2. Their ages ranged from 18 to 34 years ($M = 25.6$, $SD = 3.4$), and the mean years of education were 15.5 ($SD = 2.4$). All participants were naïve as to the purpose of the experiment.

Materials and collection procedure We used the 12 semantic categories selected from Experiment 1 as stimuli (see Table 1). Participants were given a 13-page, A5-format booklet with written instructions on the first page and a different category, followed by eight empty lines, at the top of each of the next 12 pages. The instructions on the first page told participants to list eight different exemplars for each of the 12 given semantic categories. They were also asked to write the exemplars in the order that they thought of them. The order of category presentation within the booklet was determined randomly for each participant so as to reduce order effects. Data collection took place in a classroom where all of the participants were present at the same time.

Recording procedure Responses for the 12 categories from the 62 booklets were manually entered into an Excel spreadsheet that included their ranking order in each category for every participant. We collected 5,761 responses, because some participants failed to give eight responses per category. About 1 % of the responses were excluded due to illegible handwriting. We also eliminated legible responses that were clearly category nonmembers, as well as synonyms and superordinate concepts (i.e., we discarded responses such as *foresta/wood* in response to the *<piante/plants>* category). Concepts at the subordinate level were

replaced by corresponding basic-level concepts (i.e., if a participant listed *pitbull* for the *<animali/animals>* category, we recorded the entry as *cane/dog*) only if three independent native Italian speakers (M.M., E.A., and a naïve speaker) reached unanimous consensus. We then added information about the abstraction level of each recorded concept. The intercoder reliability was very high, with a Cohen's κ value of .89 ($N = 824$; $p < .0001$).

Results and discussion

The data set was composed of 5,444 words or, without taking into account those repeated across participants, 824 distinct concepts belonging to the 12 semantic categories. The average number of concepts per category was 7.32 ($SD = 1.01$), ranging from 6.68 ($SD = 1.62$), for the *<suppellettili/furnishing and fittings>* category, to 7.73 ($SD = 1.07$), for the *<vestiti/clothes>* category. The estimated reliability of the data was very high (.966, range = .954–.975).

We calculated the dominance value of each concept as the proportion of the participants (out of 62) who produced a determinate concept in response to a given category label and the mean output position of each concept for the corresponding category (mean rank order). Together, these measures reflect the accessibility of a concept for a given category. In previous research, the order of production has been neglected by studies of semantic memory organization. In fact, only Ruts et al. (2004) found a significant correlation between generation frequency and the mean rank position of the exemplars.

Successively, we computed the first-occurrence value—that is, the proportion of participants who produced a given concept as their first response (Bueno & Megherbi, 2009; Van Overschelde et al., 2004). This was computed by dividing the number of participants who produced a given concept as the first response for a given category by the total number of participants. Finally, we calculated the availability of each concept in response to the category label (Izura et al., 2005), a measure that represents the readiness with which a concept is produced as a member of a given category, by taking into account its position in a given category for each participant, its production frequency within a category, the lowest position in which it was produced across participants, and the total number of participants who responded with it. The formula (adapted from Lopez-Chavez & Strassburger-Frias, 1987) is

$$z_i = \sum_{p=1}^n \frac{f_{pi}}{N} \cdot e^{[-2.3 \cdot (\frac{p-1}{n-1})]}, \quad (1)$$

where z_i is the accessibility of the concept i for a given category, p is the position of that concept in that category, n

is the lowest position occupied by that concept for that category, f_{pi} is the number of participants who produced the concept i at the position p in that category, and N is the total number of participants. We included all of the aforementioned variables for all concepts with a dominance value higher than .1 (i.e., concepts produced by at least 10 % of our participants) in the C-C_Pairs.xls file. As expected, all four of the measures described above were significantly correlated, reflecting the strength of the relationship between an exemplar and its corresponding category. In particular, exemplar dominance, availability, and first occurrence were all strongly and positively correlated ($r_s = .95, .76, \text{ and } .90$ for dominance vs. availability, dominance vs. first occurrence, and availability vs. first occurrence, respectively; $N = 824$; $p_s < .0001$), whereas the mean rank order of production correlated weakly and negatively with the other three measures ($r_s = -.22, -.29, \text{ and } -.28$ for dominance, availability, and first occurrence, respectively; $N = 824$; $p_s < .0001$).

In order to select the best exemplars for the construction of our norms, we chose to retain only those concepts produced by at least seven participants (i.e., those with a dominance value higher than .1). This cutoff criterion resulted in the inclusion of at least 12 exemplars per category (range 12–19; $M = 15.75, SD = 2.05$). In order to select the same number of exemplars for each category, we then selected the 12 concepts per category with the highest availability values (these concepts have a value of 1 in the [Final_Dataset] variable in the C-C_Pairs.xls file) for Experiments 3a and 3b. Moreover, we chose to exclude the <cibi/food> and <passatempi/hobbies> categories from Experiments 3a and 3b because these categories included a high number of exemplars that could be included in other semantic categories or that belonged to both the natural and artifact domains.

Experiment 3a

Concept *familiarity* represents the degree to which people have heard, read, or otherwise been exposed to a given concept in their lives, and it has been proposed that semantic knowledge about any given concept is reflected, at least in part, by the degree of familiarity with that concept (Marques, 2007; see also Funnell, 1995; Glass & Meany, 1978). For example, Glass and Meany showed that knowledge of some unfamiliar concepts (e.g., mite) is restricted to the information about category membership (e.g., “is an insect”). Although familiarity is most often considered a nuisance variable, several studies have highlighted its influence as a statistical variable in tasks requiring word processing (e.g., Gernsbacher, 1984) as, for example, in feature verification tasks (McRae et al., 1999; McRae et al., 1997).

Note that the concept of familiarity is not independent from typicality, and therefore it is not easy to determine the exact contributions of these two variables in experimental results (see McCloskey, 1980, for a detailed discussion). An exemplar’s *typicality*, also known as *goodness of example*, *representativeness*, and *prototypicality*, reflects how representative an exemplar is of a category (Hampton, 2007) and has been explained by both prototype models (Hampton, 1979; Rosch & Mervis, 1975) and exemplar models (Heit & Barsalou, 1996; Storms et al., 2000). Typicality is usually measured by asking participants to rate, on a numerical scale, the degree to which a category item is a representative member of a category. The impact of typicality is evident in categorization tasks: The more typical an exemplar of a category is, the more quickly it is verified as being a member of that category (e.g., Larochelle & Pineau, 1994; Rosch, 1975). Several studies have shown that typicality is an influential variable that explains a wide range of behavioral findings, including priming effects (Rosch, 1975), categorization probability (e.g., Hampton, Dubois, & Yeh, 2006), inductive inference (e.g., Osherson, Smith, Wilkie, López, & Shafir, 1990), memory interference effects (Keller & Kellas, 1978), and semantic substitutability (Rosch, 1977), as well as performance in production (Hampton & Gardiner, 1983) and naming (Dell’Acqua, Lotto, & Job, 2000) tasks.

Method

Participants A group of 120 native Italian speakers who had not participated in the previous experiments participated in this experiment. All of the participants gave informed consent before beginning the experimental session. They were equally divided into two groups: the first (mean age 26.4 years, range 21–56) performed a familiarity rating task, and the second (mean age 26.4 years, range 18–51) a typicality rating task.

Materials and collection procedure Familiarity with a concept (i.e., “How familiar are you with the thing the word refers to?”) was rated using a Likert-type 10-point rating scale, with 1 corresponding to *not familiar at all* and 10 to *extremely familiar*. The typicality of a concept for the corresponding category was also rated using a Likert-type rating scale, ranging from 1 (*extremely atypical exemplar*) to 10 (*extremely typical exemplar*). In both tasks, every participant provided ratings for five of the ten categories (60 concepts), so that each concept was rated by 30 different participants. The order of concepts was randomized for each category, and the assignment of categories to participants was pseudorandomized, with the restriction that each participant received either one or two categories among the three that belonged to the natural domain: <animali/animals>, <piante/plants>, and <parti del corpo/body parts> .

Results and discussion

In the Concepts.xls file, we provide the mean familiarity and typicality ratings (the variables [Familiarity] and [Typicality]) for each concept. As expected, a feature's rated familiarity and typicality were highly correlated ($r = .77$; $N = 120$; $p < .0001$). The estimated reliability of the ratings of both familiarity and typicality were quite high (.93 and .88, respectively). The reliability of the familiarity ratings was also assessed by correlating the values collected from our participants with those reported by Della Rosa, Catricalà, Vigliocco, and Cappa (2010), for the 28 words found in both sets of norms, and with those of McRae et al. (2005), with which we held 61 concepts in common. The significant correlations ($r_s = .86$ and $.69$, respectively; $p_s < .001$) further indicated that concept familiarity values were highly reliable.

Experiment 3b

Although it is widely accepted that concepts cannot be defined by a list of singly necessary and jointly sufficient features, semantic features provide a vast amount of information about the representation of semantic concepts (Cree & McRae, 2003; Garrard et al., 2001; McRae et al., 2005), especially regarding their coherence (Sloman, Love, & Ahn, 1998), similarity (Markman & Gentner, 1993; Markman & Wisniewski, 1997), and cognitive economy in hierarchical taxonomies (Collins & Quillian, 1969; Smith et al., 1974).

Experiment 3b was designed to provide normative data for 120 concrete concepts from the ten different concept classes resulting from Experiments 1 and 2. We collected normative data by asking native Italian speakers to perform a feature generation task similar to the one that has been used for other norms (De Deyne et al., 2008; Kremer & Baroni, 2011; McRae et al., 2005; Vinson & Vigliocco, 2008; Zannino et al., 2006).

Method

Participants A group of 417 volunteers (126 male, 291 female), principally undergraduate psychology students at the University of Chieti, participated in Experiment 3b. Their ages ranged from 18 to 35 years ($M = 24.6$, $SD = 3.6$), and the mean years of education were 13.6 ($SD = 1.6$). All of the participants were native Italian speakers and were naïve as to the purpose of the study.

Materials and collection procedure The stimuli came from the ten artifact and natural categories, with the 12 exemplars in each retained from Experiments 1 and 2, for a total of 120 words. The words were randomized and subdivided into

norming forms, each of which contained five concepts belonging to five different categories. This subdivision was done for a number of reasons: to keep participants from producing features for multiple semantically similar concepts or exploiting explicit comparisons among concepts belonging to the same category; to avoid order effects; and to leave as much time as needed to complete the feature-listing task. After completing the first feature-listing task, participants were free to complete a second feature-listing task with the remaining five categories. Thus, each participant listed the features for either five or ten concepts (215 and 202 participants, respectively), and each concept were listed by a minimum of 25 participants (and at most 28; the mean number of participants per concept was 25.78).

Forms were distributed during psychology classes at the University of Chieti and collected in our laboratory. The participants received a form with written instructions at the top and five concepts, followed by ten blank lines on which to list features. The instructions avoided excessive constraints, and participants were asked to define and describe each word on the list by using features. The participants were asked to enter one descriptive word or phrase per line and to try to list at least six features for each target word. An example was provided.

Transcription and labeling The data consisted of a large quantity of features for each concept listed by participants. We interpreted, organized, and labeled the feature set as follows: Participants often produced nonelementary pieces of information, such as adjective–noun or adjective–adjective composite attributes (e.g., “*ha quattro gambe*/has four legs” and “*è gialla e nera*/is yellow and black” for *tigre*/tiger), disjunctive features (e.g., “*singolo o matrimoniale*/single or double” for *letto*/bed), or compound phrases (e.g., “*usato per trasportare persone*/used to transport people” for *autobus*/bus). In these cases, we broke nonelementary attributes down into elementary pieces of information, assuming that the participant had provided two pieces of information. In less obvious cases (e.g., when talking about a tiger, “*è giallo e nero*/is yellow and black” has a different meaning than “*è giallo*/is yellow” and “*è nero*/is black”), three native Italian speakers (M.M., E.A., and a naïve speaker) judged whether the multiple terms taken together had a different meaning than the terms considered separately. Terms were divided only if the three speakers reached unanimous consensus. Quantifiers relative to the occurrence frequency of a given feature in a given concept were always dropped from the participants' responses—assuming, in line with McRae et al. (1997), that the information provided by these quantifiers would be indexed by production frequency. Finally, to ensure coherent recording of synonymous features, both within and across concepts, synonymous features were recorded under the same label (e.g., “*spazioso*/spacious”

and “*ampio/roomy*” were both recorded as “*spazioso/spacious*”) only when the three aforementioned raters reached unanimous agreement. Instead, to distinctly label features that differed in meaning but were defined by the same word, we followed the criterion used by McRae et al. (2005). We interpreted responses conservatively, without modifying feature names, to avoid potential ambiguity (see McRae et al., 2005). We also used a number of key words and phrases to organize and code the features in order to indicate their feature type; some examples are “*un carnivoro/a carnivore*” (taxonomic feature); “*richiede un biglietto/requires a ticket*,” “*si trova in casa/found in houses*,” and “*produce latte/produces milk*” (encyclopedic features); “*fatto di plastica/made of plastic*” (sensorial feature); or “*usato per scrivere/used to write*” and “*usato dai fumatori/used by smokers*” (functional features).

After modifying the participants’ responses, we added information regarding the order of production and entered the data in an exemplar-by-feature intensity matrix, with each cell corresponding to the production frequency of a feature for a given concept. This value represented the number of participants who listed that feature for that concept, ranging from 0 to 25–28, with the upper limit depending on the concept. Idiosyncratic features—those whose production frequency did not reach a value of 4 in at least one concept in the set—were discarded. We chose an inclusion cutoff that was slightly different in magnitude from that used with McRae et al.’s (2005) norms, but the cutoffs were similar in terms of relative dominance. In fact, McRae et al. (2005) chose an inclusion cutoff of five of 30 participants who listed each concept (see also Zannino et al., 2006), corresponding to a relative dominance of 16.7 %. We, on the other hand, used a cutoff corresponding to an average relative dominance of 15.5 %.

The included features were then classified into feature types according to the kinds of information that they represented. In the main classification, all features were classified into basic knowledge types in line with McRae et al.’s (2005) criteria: A feature was defined as sensorial if it represented information regarding visual (*visual-color*, *visual-form and surface*, or *visual-motion*) or other (*sound*, *tactile*, *smell*, or *taste*) primary sensory systems, and as functional if it referred to functional/motor information regarding the ways in which people interact with the object. In particular, the *functional* label was applied to features that represented information regarding what an entity is used for by people, the agent who typically interacts with an exemplar, or the action or motor movement associated with the interaction with an entity (see Cree & McRae, 2003, for a detailed discussion; see also Zannino et al., 2006). Thus, in line with McRae et al.’s (2005) procedure, we did not consider a feature as functional if it referred to “what an item does”, but only if it referred to “what it is used for”. In

this manner, we teased apart two types of information that Farah and McClelland’s (1991) original definition considered comparable. Moreover, a feature was classified as encyclopedic if it described some other type of knowledge or associative relationship. Finally, we added the taxonomic feature type to McRae et al.’s (2005) classifications, because participants often listed features regarding the superordinate category of a concept or subordinate and coordinate concepts as well.

All features were independently labeled by two native Italian speakers (M.M. and E.A.), and disagreements were mediated by a third colleague. The agreement between our labeling and that of the secondary rater was high, with a Cohen’s κ value of .94 ($N = 730$; $p < .001$).

Measures and statistics

The following sections describe the measures that we collected for the norms, their past application, and the statistics calculated from them. Note that, where possible, we also calculated (and included in our files) the same distributional measures excluding all taxonomic features, according to the assumption that this type of knowledge provides a somewhat different type of information (mainly referring to the superordinate category of a concept) as compared to other knowledge types (see McRae et al., 2005). However, for ease of reading, we will not discuss the measures calculated here without also considering taxonomic features.

Overall database characteristics Overall, we collected data for 23,696 features or, excluding those repeated across participants, 4,481 distinct features. Each participant listed features for 7.66 concepts, on average ($SD = 2.62$), while the average total number of features obtained for a concept was 197.47 ($SD = 22.51$). By applying our production frequency cutoff for feature inclusion, the total set was reduced to 730 unique features, yielding 1,715 concept–feature pairs. Finally, the average number of distinct features per concept ([n_Feat] in the Concepts.xls file) was 14.29 ($SD = 3.24$). In particular, the latter index provides information about the semantic richness of the conceptual representation—that is, the amount of information that a concept can evoke—and its importance in the semantic system has been proven by various studies (for a review of this topic, see Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008)

Concept lexical measures Concept-level measures for the 120 concepts are presented in the Concepts.xls file. Column names and a short explanation are presented in Appendix C. The first four columns contain the names of each concept in Italian and English, as well as the corresponding category and domain. Next, we included the number of letters for

each Italian and English word and the number of syllables for the Italian word denoting each concept. The last lexical measures were word frequency—taken from the COLFIS corpus (*Corpus e Lessico di Frequenza dell’Italiano Scritto*, the “Corpus and Frequency Lexicon of Written Italian”; Bertinetto et al., 2005), a three-million word corpus of contemporary written Italian—summed over a word’s singular and plural uses, and the natural logarithm of the frequency. The more frequently a word is read, heard, written, or spoken, the more often the conceptual representation of the corresponding object is computed. Thus, the frequency with which a word is encountered likely influences how easily the meaning of a word can be computed (Cree & McRae, 2003), as well as how quickly and easily that word can be recognized and produced (Duncan, 1977; Forster & Chambers, 1973; McRae, Jared, & Seidenberg, 1990). Moreover, this index correlates with both concept familiarity and word length: Lower-frequency words tend to be less familiar and longer (Graves, Grabowski, Mehta, & Gordon, 2007).

Exemplar generation measures We included the measures reflecting the relationship between concepts and categories, computed as described above (see the **Results and Discussion** section of Exp. 2), in the Concepts.xls file. These four variables—[Exempl_M_Rank], [Exempl_Dominance], [Exempl_1°Occur], and [Exempl_Availability]—refer to the internal structure of semantic categories and measure different aspects of categorization processes. Since the most widely studied semantic property in categorization research is typicality (Koivisto & Laine, 1999; McFarland, Duncan, & Kellas, 1978; Mervis & Rosch, 1981; Smith et al., 1974), we assessed the relationship between the four exemplar generation measures and typicality ratings. Correlational analyses showed that typicality ratings were significantly correlated with the exemplar generation measures ($r_s = .56, -.55$, and $.52$ for the exemplar dominance, mean rank, and first-occurrence measures, respectively; $N = 120$; $p_s < .0001$), and especially with the exemplar availability measure ($r = .62$; see Fig. 2). Moreover, since other variables, such as concept familiarity or word frequency, have also proven to influence the ease of exemplar production for a given category (e.g., Hernández-Muñoz, Izura, & Ellis, 2006), we also assessed the correlation between the four exemplar generation measures and familiarity ratings. We found significant correlation indexes ($r_s = .52, -.47$, and $.45$ for the exemplar dominance, mean rank, and first-occurrence measures, respectively; $N = 120$; $p_s < .0001$), and, again, especially with the exemplar availability measure ($r = .54$; see Fig. 2).

Feature dominance Feature production frequency, also called *dominance* (especially when calculated as the proportion of the total sample of participants), represents the

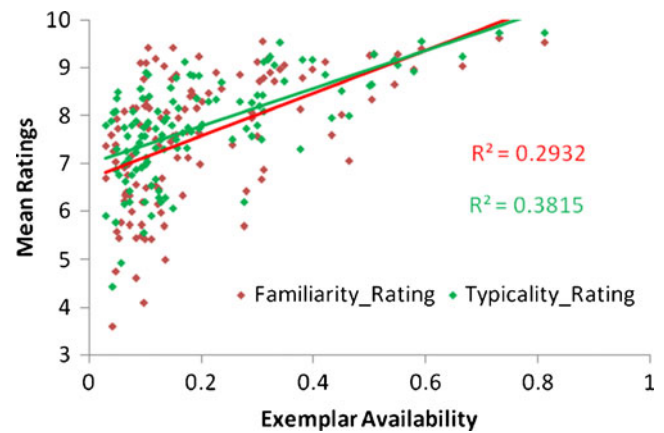


Fig. 2 Scatterplot of participants’ mean familiarity (red diamonds) and typicality (green diamonds) ratings, as a function of exemplar availability values. The corresponding best-fitting linear regression lines are shown in the same colors

number of participants who listed a given feature for a given entity in a feature-listing task. This index refers to the salience or importance of a feature for a given concept (Smith & Medin, 1981) and reflects the accessibility of that feature when a concept is computed from its name. In the C-F_Pairs.xls file, we have included the raw production frequency of each concept–feature pair [Prod_Fr], indicating the number of participants who listed that feature for that concept. On the basis of the number of participants presented with each concept ([n_Subj]), we calculated dominance ([Domin]) as the ratio between [Prod_Fr] and [n_Subj]. The dominance distributions for concept–feature pairs are shown in Fig. 3, which shows that a plurality of features (420, corresponding to nearly 25 %) were listed for a given concept by four participants (i.e., the production frequency cutoff that we used to include concept–feature pairs in our norms, corresponding to an average relative dominance value of 15.5 %).

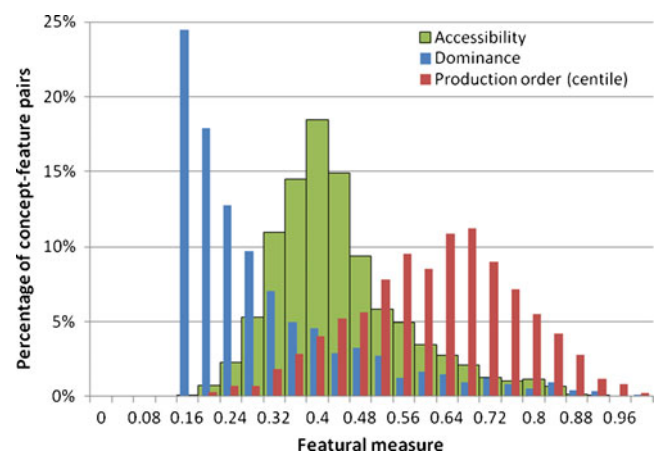


Fig. 3 Distributions of feature accessibility (green bars), dominance (blue bars), and normalized order of production (red bars) for the total set of 1,715 concept–feature pairs in Experiment 3b

We included the corresponding measures calculated for each concept, with [sum_Prod_Fr] indicating the summed frequency of production of all the features listed for a given concept, and [M_Prod_Fr] indicating the mean production frequency of the features listed for that concept (calculated by dividing [sum_Prod_Fr] by [n_Feat], the number of features listed for that concept) in the Concepts.xls file. Moreover, we calculated a [Domin] variable for every concept that indicates the average of the dominance values of all of the features listed for a given concept and that was calculated as the ratio between [M_Prod_Fr] and [n_Subj].

Feature accessibility The order of production can also be used in the computation of the accessibility of a feature for a given concept, as well as as an index of the generation of an exemplar in a given category. Thus, feature accessibility can also be reflected in the order of generation of that feature. For example, the features “*fatto di legno/made of wood*,” for *tavolo/table*, and “*lunga/long*,” for *sciarpa/scarf*, both have the same dominance, but “*fatto di legno/made of wood*” was listed sooner than “*lunga/long*,” indicating that it was more readily accessible in semantic memory when the corresponding concept cue was presented. Therefore, the accessibility of a feature in response to a concept’s name might be independent of that feature’s absolute production frequency and might be related to a feature’s salience, defined in relation to other features.

In C-F_Pairs.xls, we included the variable [M_Prod_Ord], representing the mean order of production of each feature for a given concept. Because participants listed a different number of features, we also calculated a normalized order of production ([M_Prod_Cent]) by converting each order of production value in centiles, as explained above (see Exp. 1, Transcription and Labeling section). Thus, the [M_Prod_Cent] index assumes a maximum value of 1 for a feature listed first for a concept by a participant, and it approaches 0 for a feature listed last (see Fig. 3 for the distribution of this index across concept–feature pairs). To confirm that the order in which a feature was listed reflected the accessibility of that feature for a given concept, we performed a correlation analysis between the normalized order of production and the dominance of each feature. The Pearson’s correlation coefficient ($r = .27$; $N = 1,715$; $p < .0001$) was significant, confirming that features with high dominance were listed sooner, and features with low dominance were generated later.

We also calculated a new measure to reflect the accessibility of a feature in response to a concept’s name, by taking into account both the order and the frequency of production of a given feature. Assuming that a feature with an ideal accessibility would be generated first by all participants (i.e.,

it would have both a normalized order of production and a dominance of 1), we measured the accessibility value a_{ij} of a feature by calculating its Euclidean distance from the point (1, 1) in the Dominance \times Order of Production space. We calculated this measure ([Access] in the C-F_Pairs.xls file) for each feature j in a concept i as follows:

$$a_{ij} = 1 - \frac{\sqrt{(1 - x_{ij})^2 + (1 - y_{ij})^2}}{\sqrt{2}}, \quad (2)$$

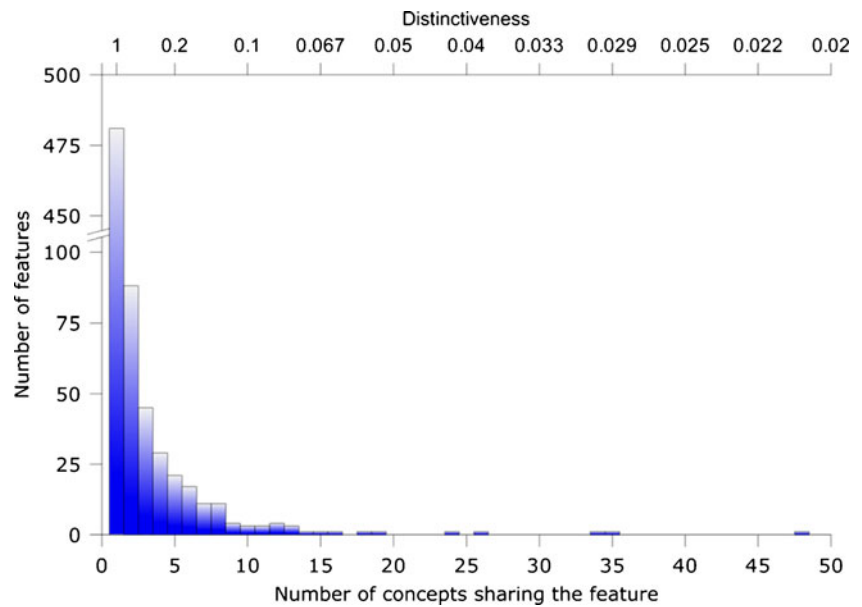
where x_{ij} and y_{ij} indicate, respectively, the dominance and the normalized order of production of a given feature j in a given concept i .

Accessibility values range from 0 to 1, with 1 indicating ideal accessibility (see Fig. 3 for the distribution of the accessibility value across concept–feature pairs). With this measure, it is possible to find two features with the same accessibility but different dominance values: For example, both the feature “*annusa/sniffs*” for *naso/nose* and the feature “*usato per riscaldare/used to warm up*” for *cappotto/coat* show the same accessibility value of .64, with the latter listed sooner but by fewer participants, as compared to the former.

Feature distinctiveness An important property of feature-based semantic representations is that features differ in terms of how informative they are about the entity they describe—that is, some features provide more information than others in identifying a concept, by distinguishing it in general or among other similar concepts in the same category. For example, “*miagola/meows*” is a highly distinctive feature for *gatto/cat*, whereas “*ha quattro gambe/has four legs*” is very nondistinctive, because many things (not only cats) share this attribute. Thus, distinguishing features provide necessary or sufficient information for identifying an entity and are critical when performing tasks that require discriminating a specific concept among similar concepts, such as word–picture matching and naming from definitions or pictures.

We defined a concept’s feature as distinguishing if it was listed for only one concept in the norms (variable [Disting] in the C-F_Pairs.xls file). On the basis of this information, we included the number of distinguishing features per concept ([n_Disting_Feat]) and, by dividing this measure by the number of features listed for that concept, we also calculated the corresponding percentage of distinguishing features ([%_Disting_Feat]) in the Concepts.xls file. As can be seen in Fig. 4, the majority of features (481 out of 730, corresponding to 65.9 %) were listed for only one of the concepts, and were accordingly classified as distinguishing. Furthermore, the number of features dramatically decreased as a power function of the number of concepts for which they were listed.

Fig. 4 Number of features (out of 730) as a function of the number of concepts in which they occurred (and the corresponding distinctiveness measures)



Although the measures above are useful for understanding which attributes best discriminate a given concept among others, the amount of information provided by a feature about a concept's identity falls along a continuum, and thus is not perfectly represented by a dichotomic variable. For example, “*ha gambe/has legs*” and “*ha una coda/has a tail*” are not very distinguishing features at all, whereas “*ha una criniera/has a mane*” is quite distinguishing, since it is shared by only two concepts in our norms, *leone/lion* and *cavallo/horse*, and only the highly distinguishing feature “*usato per trainare carrozze/used to pull coaches*” permits one to correctly identify the corresponding entity as a horse. Therefore, we calculated the distinctiveness, also known as *informativeness*, a measure that allows for a continuum representation of the amount of distinguishing information provided by a given feature for a concept. Distinctiveness values range from highly shared to highly distinguishing (Cree & McRae, 2003; Devlin, Gonnerman, Andersen, & Seidenberg, 1998; Garrard et al., 2001; McRae et al., 2005; Zannino et al., 2006). In line with Devlin et al. and McRae et al. (2005), but contrary to Garrard et al. and Zannino et al., we computed the distinctiveness of a feature (the variable [Distinctiveness] in C-F_Pairs.xls) as the inverse of the number of concepts in which the feature appears in the entire norms (i.e., as $1/[n_Concepts]$), without limiting the contrast set of concepts to those within a category. In this manner, distinctiveness was defined as a continuous but nonlinear measure, since the difference in distinctiveness for a feature occurring in one versus two concepts is larger than that for a feature occurring in two versus three concepts, so that the distinguishing nature of a feature is represented by a binary measure, whereas feature distinctiveness is a continuous index. Importantly, features that appear in only one

concept have distinguishing and distinctiveness values that are equal (i.e., 1). Next, for every concept, we calculated the mean distinctiveness of all of the features that composed it (the variable [M_Distinctiveness] in the Concepts.xls file).

Finally, we computed another continuous measure regarding a feature's informativeness, the so-called *cue validity* (Bourne & Restle, 1959; Rosch & Mervis, 1975). This is the conditional probability of a concept, given a feature, and represents the importance of a property, or how easily it “pops up” when one thinks of a concept (Smith & Medin, 1981). We included in the C-F_Pairs.xls file the corresponding variable [Cue_Val], calculated as the production frequency of a given feature for a given concept ([Prod_Fr]) divided by the sum of the production frequencies of that feature for all concepts in which it appears in the norms ([Cum_Prod_Fr]). Again, we calculated the mean cue validity of each feature for every concept (the variable [M_Cue_Val] in the Concepts.xls file). As expected, the cue validity and distinctiveness were highly correlated ($r = .99$; $N = 1,715$; $p < .0001$), since the former is a measure analogous to distinctiveness, except that it additionally takes into account an estimate of feature salience. Therefore, if a feature is truly distinguishing, and thus is listed for only one concept in the norms (as “*miagola/meows*” for *gatto/cat*), it will have a maximum value (i.e., 1) for both cue validity and distinctiveness. On the contrary, if a feature is not very distinguishing, being shared by many concepts, it will have a very low score (i.e., approaching to 0) for both cue validity and distinctiveness.

Feature relevance Previous authors have proposed a feature-based model of semantic memory containing a new measure, semantic relevance, considered the factor that

“best captures the importance of a feature for the meaning of a concept” (Sartori & Lombardi, 2004, p. 440). In other words, high-relevance features are those that mainly contribute to the “core” meaning of a concept and are crucial both in identifying that concept and in discriminating it from other, similar concepts (Sartori, Gnoato, Mariani, Prioni, & Lombardi, 2007; Sartori & Lombardi, 2004).

Semantic relevance can be construed as a nonlinear combination of a feature’s dominance and distinctiveness (Sartori et al., 2007; Sartori, Lombardi, & Mattiuzzi, 2005), since it is computed as the result of two components: a local component, comparable to dominance, reflecting the importance of a semantic feature for a given concept, and a global component, comparable to distinctiveness, reflecting the importance of the same semantic feature for the whole set of concepts. For example, the semantic feature “*sente/hears*” is of very high relevance for the concept *orecchio/ear*, since all participants listed it when describing that concept, and at the same time, none of them used the same feature to define other concepts. In other words, the feature “*sente/hears*” for the concept *orecchio/ear* has a maximum score in both dominance and distinctiveness measures—that is, it has both high local importance and high global importance. In contrast, the feature “*grande/big*” is of low relevance for the concept *orecchio/ear*, because very few participants used it to describe that concept, and it was listed at the same time for many concepts. In this case, the feature “*grande/big*” has both low dominance and distinctiveness values. It is worth noting that relevance is different from distinctiveness. Distinctiveness is a concept-independent measure, and therefore the distinctiveness score of a feature is the same for all concepts (i.e., regardless of the specific concept in which that feature appears). Instead, relevance may be considered a concept-dependent measure, because a given feature can take different relevance scores across different concepts (i.e., depending on the concept in which that feature appears). For example, the feature “*fatto di vetro/made of glass*” has a low relevance score for the concepts *centrotavola/centerpiece* and *teglia/baking pan*, but it has a high relevance score for the concept *bicchiere/glass*.

We calculated the semantic relevance of each concept–feature pair as described in Sartori et al. (2007) (the variable [Relevance] in the C-F_Pairs.xls file). Briefly, we computed the relevance value k_{ij} of a feature j in a concept i as follows:

$$k_{ij} = 10x_{ij} \times \ln\left(\frac{I}{I_j}\right), \quad (3)$$

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 our norms (i.e., 120); I_j indicates the number of concepts in which the feature j appears; and $\ln(I/I_j)$

represents the global component of k_{ij} . Note that we used dominance instead of raw production frequency (used by Sartori, Lombardi, & Mattiuzzi, 2005) in computing relevance because the concepts in our norms were presented to a variable number of participants. We also included the measure [M_Relevance], computed as the mean relevance of each concept’s features in the Concepts.xls file.

Moreover, for each concept–feature pair, we computed a new measure that shares the same underlying principle as semantic relevance—that is, semantic significance (the variable [Significance] in the C-F_Pairs.xls file). As for relevance, this measure is composed of a local and a global component, but in this case, the local component of semantic significance corresponds to the accessibility of a feature, as described above. In this way, the importance of a given semantic feature for its corresponding concept is reflected by the accessibility. This measure accounts for both the dominance and the priority with which that feature emerged in the participants’ semantic conceptual representations. We calculated the significance score s_{ij} of a feature j in a concept i , as in the case of relevance, but modified the local component of s_{ij} by replacing the feature’s dominance— x_{ij} in the Eq. 3—with the accessibility of that feature, a_{ij} . We have included the corresponding variable [M_Significance], computed as the mean significance of each concept’s features, in the Concepts.xls file. Relevance and significance values are strongly correlated ($r = .90$; $N = 1,715$; $p < .001$)

Feature correlation Features are considered correlated if they tend to appear together in the semantic representation of a concept. For example, the features “*ha ruote/has wheels*” and “*usato per muoversi/used to move*” are highly correlated, because typically, something that is used to move probably also has wheels. Thus, these two features tend to co-occur in the representation of that concept. Correlations among features represent a central factor in various models of semantic organization (Malt & Smith, 1984; Rosch, 1978; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Rosch (1978) claimed that clumps of correlated features form the basis of basic-level conceptual representations (e.g., *mucca/cow* or *cavallo/horse*), a level at which feature correlations best fit together within categories and can be used to distinguish among concepts (Rosch et al., 1976).

Feature correlations are also central in connectionist models of featural representations of word meaning (McRae et al., 1997) and are believed to play an important role in the organization of semantic memory. Feature correlations influence the early computation of word meaning in online tasks, in contrast to slow, offline tasks, which require high-level reasoning processes (i.e., similarity and typicality ratings). To test these hypotheses, McRae et al. (1997)

compared performance on offline concept similarity rating and typicality judgment tasks with performance for online speeded priming semantic and feature verification tasks. They found that feature correlations predicted performance only in online tasks. Other studies replicated this finding in a feature–concept priming paradigm: Strongly intercorrelated features primed lexical decisions to target concepts significantly more than weakly intercorrelated features did (Taylor, Moss, Randall, & Tyler, 2004).

We calculated several variables based on the correlational structure of feature representations. First, we constructed a Feature \times Concept intensity matrix, with 730 rows (one for each feature) and 120 columns (one for each concept) in which each cell corresponded to the production frequency of a given feature for the corresponding concept (see the Fr_Production sheet in the C-F_Matrix.xls file). In line with McRae et al. (2005; McRae et al., 1997), we chose to exclude from the correlational analyses features that occurred in only one or two concepts in the norms, in order to avoid spurious correlations. In fact, including very distinguishing features would have caused an artificial increase in the number of perfectly correlated feature pairs. We then performed correlational analyses on the resulting 161 \times 120 Feature \times Concept matrix by computing the Pearson product–moment correlation coefficient for each resulting pairwise combination of feature vectors. A feature pair was considered correlated if the features were associated by $|r| \geq .234$, indicative of a significant correlation at the $p \leq .01$ level, and thus shared at least 5.5 % of their variance (i.e., with $r^2 \geq .055$). Finally, in line with McRae et al. (2005), we calculated a number of concept-dependent measures indicating the degree to which a concept's features were intercorrelated (see Zannino et al., 2006, for a comparison of concept-dependent and concept-independent measures of feature correlation).

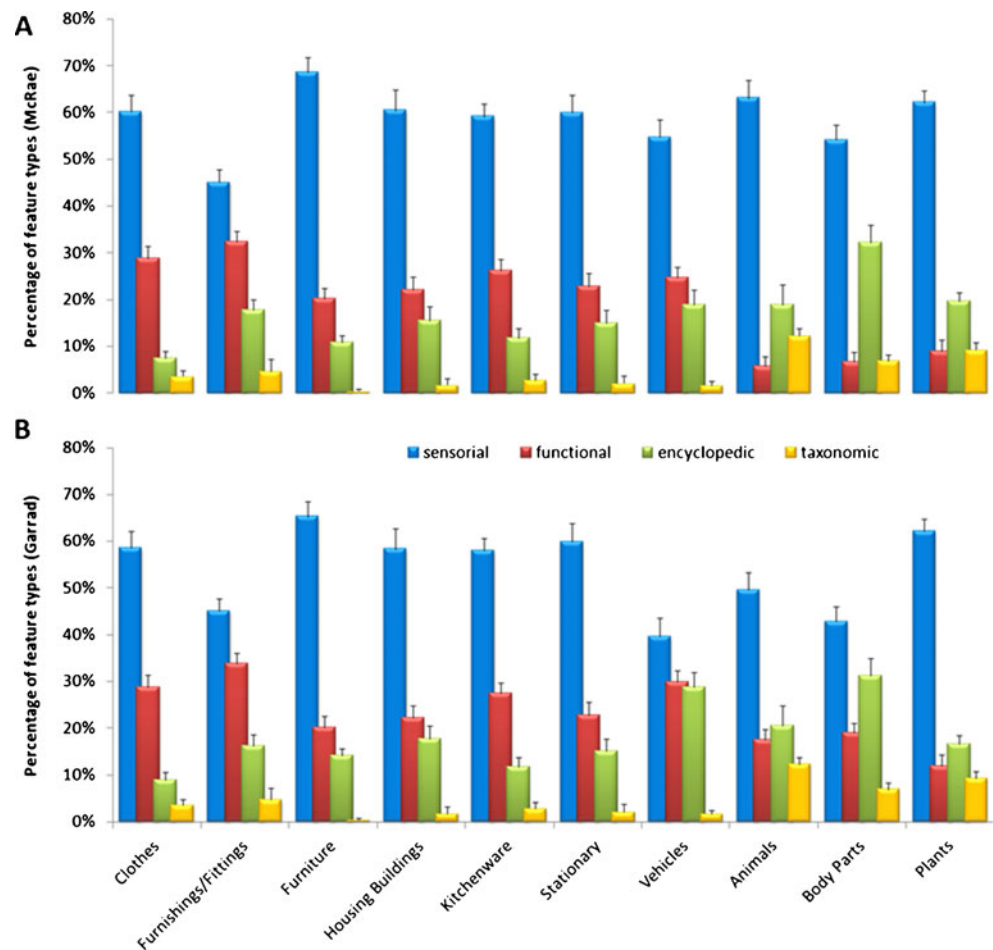
We included the variable [n_Correl], which indicates, for a given feature, the number of features occurring in the same concept (and in at least two further concepts) with which that feature is significantly correlated, in the C-F_Pairs.xls file. Then we calculated the percentage of significant correlation [%_Correl] based on [n_poss_Correl] for each feature, which in turn indicates the number of possible pairwise combinations for a feature (i.e., for each feature, the number of other features occurring in the same concept, excluding those listed for only one or two concepts). In the same manner, we included [n_Correl_Pairs], indicating the number of significantly correlated feature pairs per concept, and [%_Correl_Pairs], indicating the percentage of correlated feature pairs per concept, and calculated by dividing [n_Correl_Pairs] by [n_poss_Correl_Pairs], which in turn indicates the number of possible pairwise combinations of features that appeared in that concept and were listed for three or more concepts, in the Concepts.xls file.

Second, we included in the C-F_Pairs.xls file a measure of the intercorrelational strength of a given feature for its corresponding concept, [Intercorr_Str], which indicates the degree to which a feature is correlated with the other features that occur in the same concept and is computed by summing the proportions of shared variance (r^2) between that feature and each of the other features occurring in the same concept with which it is significantly correlated (McRae et al., 2005). Because this measure is no longer a percentage (being actually a sum of percentages), we also calculated [M_Intercorr_Str] by computing the ratio between [Intercorr_Str] and [n_Correl]. Similarly, in the Concepts.xls file we included [Intercorr_Density], a measure of the amount of intercorrelation between features occurring in a given concept, calculated by summing the proportions of shared variance across all significantly correlated feature pairs of that concept. Again, to have a somewhat standardized measure, we also included the measure [M_Intercorr_Density], which is calculated by dividing [Intercorr_Density] by [n_Correl_Pairs].

Feature type distribution As is described above (see the Recording Procedure section of Exp. 2), we first classified each feature into four principal types of knowledge, according to the kind of information that it represented (i.e., sensorial, functional, encyclopedic, or taxonomic). This gave us a measure of the importance of the type of knowledge for each concept (see Cree & McRae, 2003, for a detailed discussion). This information about each feature's main knowledge type was included in the C-F_Pairs.xls file as [Main_Feat_Type]. We also included [Det_Feat_Type], representing a more detailed classification in which each feature was classified into one of the nine knowledge types defined by Cree and McRae. On the basis of the main classification in knowledge types, we also computed four variables representing the number of features of each of the four principal knowledge types that were listed for each concept. These measures were included in the Concepts.xls file and were called [n_Sens_Feat], [n_Funct_Feat], [n_Encyc_Feat], and [n_Taxon_Feat]. We also computed the corresponding percentages ([%_Sens_Feat], [%_Funct_Feat], [%_Encyc_Feat], and [%_Taxon_Feat]) by dividing each of the four aforementioned measures by [n_Feat]—that is, the total number of features listed for each concept (see Fig. 5a).

The supposed difference between living and nonliving things in the distribution of features representing different types of knowledge depends on the precise definition of sensorial and functional information. In fact, different definitions have revealed discrepancies in the literature (Caramazza & Shelton, 1998; Farah & McClelland, 1991; Garrard et al., 2001; McRae & Cree, 2002; Vinson & Vigliocco, 2008; Zannino et al., 2006). As far as functional features are concerned, in contrast with Farah and

Fig. 5 Mean percentages of feature types in exemplars from the ten semantic categories of Experiment 3b. Knowledge types are labeled according to the definitions of McRae et al. (2005) (a) or Garrard et al. (2001) (b). Error bars represent standard errors of the means



McClelland and in line with Garrard et al., we chose to use a quite restrictive classification (see also McRae et al., 2005, and Zannino et al., 2006). In this manner, however, it was unlikely that a living concept could assume a functional feature, and this may have artificially caused the difference in the distributions of sensorial and functional features across semantic domains. For this reason, we performed an additional classification of all features according to the sensorial and functional definitions proposed by Garrard et al. These authors defined features that can be appreciated in a sensorial modality as sensorial, and those that describe an action, activity, or use of a given item as functional. We included this feature-based information in the C-F_Pairs.xls file in the variable [Feat_Type(Garrard)]. We also included the number and the percentage of features of each of the four principal types of knowledge listed for each concept (according to Garrard et al.'s classification), in the Concepts.xls file (see Fig. 5b).

Concept distance Semantic representations can be characterized as points in a multidimensional space in which dimensions are defined by features. In these spatial representational models, the similarity between a pair

of concepts is defined as the proximity between the pair of corresponding points: The closer that two concepts are, the more similar their semantic representations. Different measures of semantic similarity have been used to predict graded semantic effects (i.e., effects that reflect the degree of similarity between concepts, rather than simply a related/unrelated dichotomy), such as performance in a category-learning task (Kruschke, 1992), category verification latencies (Smith et al., 1974), semantic priming effects (Cree et al., 1999; Vigliocco et al., 2004), and semantic interference effects in object naming (Vigliocco, Vinson, Damian, & Levelt, 2002).

In psychological research, semantic similarity has typically been measured subjectively by asking participants to rate the pairwise similarity of a set of concepts. However, when the stimulus set is quite large, rated pairwise similarities are not feasible, due to the extremely large number of stimulus pairs. In this case, semantic distance measures can be computed from an Exemplar \times Feature matrix either by correlating the exemplars' feature vectors or by computing Euclidean or cosine distances between the exemplars' feature vectors, as well as in sorting tasks, direct and triad ratings, conditional probability ratings, and proximity ranks in

generation tasks (see Borg & Groenen, 2005, for a review).

We chose to compute the different concept distance measures directly from our feature norms, starting from the raw Feature \times Concept matrix. The first semantic distance measure that we defined was the cosine distance between two concepts—that is, the complement of the cosine of the angle between two feature vector representations. The Cos_Domin sheet in the Distance_Matrix.xls file is a 120×120 matrix of cosine distances between each pairwise combination of concepts in our norms. This matrix was calculated from the 730×120 Feature \times Concept dominance matrix, in which each cell corresponds to the dominance of a given feature for its corresponding concept. We then calculated the cosine distances $d_{\text{Cos}}(m, n)$ as the complement of the dot product between two concept vectors, divided by the product of their length, as follows:

$$d_{\text{Cos}}(m, n) = 1 - \frac{\sum_J x_{mj} \cdot x_{nj}}{\sqrt{\sum_J (x_{mj})^2} \cdot \sqrt{\sum_J (x_{nj})^2}}, \quad (4)$$

where x_{mj} and x_{nj} represent dominance values of the feature j for the concepts m and n , respectively, and J indicates the number of features. Cosine distance values range from 0 (identical concepts: minimum distance, total similarity) to 1 (concept vectors are orthogonal: maximum distance, no similarity) and provide a symmetric and non-linear measure of proximity that reflects the similarity between active features in a given pair of concepts, emphasizing shared features.

We also calculated two further familiar and commonly used geometric measures—that is, Euclidean and city block distances—that are well-known Minkowski metrics (see the Euc_Domin and City_Domin sheets in the Distance_Matrix.xls file). The Euclidean distance $d_{\text{Euc}}(m, n)$ was calculated as

$$d_{\text{Euc}}(m, n) = \sqrt{\sum_J |x_{mj} - x_{nj}|^2}, \quad (5)$$

whereas the city block distance $d_{\text{City}}(m, n)$, which emphasizes distinctive features (e.g., Dry & Storms, 2009), was calculated as

$$d_{\text{City}}(m, n) = \sum_J |x_{mj} - x_{nj}|. \quad (6)$$

For both Euclidean and city block distances, highly similar concept pairs assume values approaching zero. We also included three other distance matrixes (the sheets Cos_Binary, Euc_Binary, and City_Binary) in the Distance_Matrix.xls file that were similar to the three just described, except that they were computed from the 730×120 Feature \times Concept

applicability matrix, in which each cell assumes a value of 1 if a given feature is listed for the corresponding concept, or otherwise the cell assumes a value of 0.

Next, for each distance measure, we computed the mean distance between a given concept's feature vector and the feature vectors for every other concept in the norms ([M_Cos_Domin], [M_Euc_Domin], [M_City_Domin], [M_Cos_Binary], [M_Euc_Binary], and [M_City_Binary] in the Concepts.xls file). In particular, the [M_Cos_Binary] measure represents the density of a concept's distant neighborhood (see Mirman & Magnuson, 2008, for the opposite effects of near and distant neighbors on word processing), since it is strongly correlated with the total number of a concept's neighbor (the variable [n_Neighb] in the Concepts.xls file)—that is, the number of concepts that share at least one semantic feature with that concept ($r = .90$; $N = 120$; $p < .0001$). We then calculated the variable [M_Cos_Near_Neighb] for each concept as the mean cosine proximity of its four closest semantic neighbors. This measure represents the degree of a concept's potential confusability (Cree & McRae, 2003) and reflects the influence of its near semantic neighborhood.

Conclusion

In our study, we have described norms for a large set of concrete concepts in Italian, together with descriptive statistics and participants' ratings. Importantly, our norms differ from existing norms in that the concepts and categories were empirically selected, while those in past literature have been based on published databases or chosen a priori by the experimenters. Our method of selecting categories and concepts identified two properties missing from existing norms: Some categories neglected in the literature were present in the semantic memory of our participants, and a larger presence of artifacts than of natural concepts emerged.

Semantic norms provide valuable information about the representations of categories and concepts and have served as the basis to account for several empirical phenomena. In line with this, we have added new measures to better clarify feature representations in semantic memory—that is, featural accessibility and significance. These indexes are measures of the informativeness of the features that best capture the importance of a property for a given concept. In summary, we believe that these norms may be a useful tool in several areas of research. In particular, we propose that they can be used to verify models and theories about the organization of semantic knowledge and about conceptual representation.

Appendix A

Table 2 Explanations of variables included in the C-C_Pairs.xls file

Variable	Explanation
Category	English category name
Concept (IT)	Italian concept name
Concept	English concept name
Exempl_Dominance	dominance: proportion of participants that produced a determinate concept
Exempl_M_Rank	mean rank order: mean output position of each concept
Exempl_1°Occur	first occurrence: proportion of participants who produced a concept as first response
Exempl_Availability	availability: readiness with which a concept is produced as a member of a given category
Final_Dataset	concept set: concepts produced by at least 10 % of the participants and took higher availability values (1 = included in the final data set)

Appendix B

Table 3 Explanations of variables included in the C-F_Pairs.xls file

Variable	Explanation
Domain	English domain name
Category	English category name
Concept (IT)	Italian concept name
Concept	English concept name
Feature (IT)	Italian feature name
Feature	English feature name
Main_Feat_Type	feature knowledge type classification according to McRae et al. (2005)
Det_Feat_Type	feature knowledge type detailed classification according to McRae et al. (2005)
Feat_Type(Garrard)	feature knowledge type classification according to Garrard et al. (2001)
n_Subj	number of participants that were presented with each concept
Prod_Fr	number of participants who listed a feature for a given concept
Cum_Prod_Fr	sum of the production frequencies of a feature for all concepts in which it appears
Domin	dominance value of a feature
M_Prod_Ord	mean order of production of a feature for a given concept
M_Prod_Cent	normalized order of production of a feature for a given concept
Access	accessibility value: Euclidean distance from the point (1, 1) in the (dominance by order of production) space
Disting	D = distinguishing feature, ND = nondistinguishing feature, where distinguishing is defined as occurring in one concept in the norms
n_Concepts	number of concepts in which a feature was listed
Distinctiveness	inverse of the number of concepts in which the feature appears
Cue_Val	cue validity of the concept's features, including taxonomic features
n_Correl	number of features occurring in the same concept with which a given feature is correlated, including taxonomic features
n_poss_Correl	number of possible pairwise combinations for a feature, including taxonomic features
%_Correl	percentage of features occurring in the same concept with which a given feature is correlated, including taxonomic features
n_Correl_NoTax	number of features occurring in the same concept with which a given feature is correlated, excluding taxonomic features
n_poss_Correl_NoTax	number of possible pairwise combinations for a feature, excluding taxonomic features
%_Correl_NoTax	percentage of features occurring in the same concept with which a given feature is correlated, excluding taxonomic features
Intercorr_Str	intercorrelational strength of a feature for a given concept, including taxonomic features
M_Intercorr_Str	mean intercorrelational strength of a feature for a given concept, including taxonomic features
Intercorr_Str_NoTax	intercorrelational strength of a feature for a given concept, excluding taxonomic features
M_Intercorr_Str_NoTax	mean intercorrelational strength of a feature for a given concept, excluding taxonomic features
Relevance	relevance of each concept–feature pair
Significance	significance of each concept–feature pair

Appendix C

Table 4 Explanations of variables included in the Concepts.xls file

Variable	Explanation
Domain	English domain name
Category	English category name
Concept (IT)	Italian concept name
Concept	English concept name
Length (IT)	number of letters in Italian concept name
Length	number of letters in English concept name
n_Syllabes (IT)	number of syllables in Italian concept name
Word_Fr (IT_COLFIS)	COLFIS (2005) frequency of concept name: sum of singular and plural
LN_Word_Fr	natural logarithm of COLFIS frequency: sum of singular and plural
Familiarity_Rating	mean concept familiarity rated by 30 participants: 1–10 scale, 10 = <i>extremely familiar</i>
Typicality_Rating	mean concept typicality rated by 30 participants: 1–10 scale, 10 = <i>extremely typical exemplar</i>
Exempl_Dominance	concept dominance: proportion of participants that produced a concept
Exempl_M_Rank	mean rank order: mean output position of a concept
Exempl_1°Occur	first occurrence: proportion of participants who produced a concept as first response
Exempl_Availability	availability: readiness with which a concept is produced as a member of a given category
n_Subj	number of participants that were presented with a concept
n_Feat	number of features listed for a concept, including taxonomic features
sum_Prod_Fr	summed frequency of production of all the features listed for a concept, including taxonomic features
M_Prod_Fr	mean production frequency of the features listed for a concept, including taxonomic features
Domin	average of the dominance values of all the features listed for a concept, including taxonomic features
n_Feat_NoTax	number of features listed for a given concept, excluding taxonomic features
Prod_Fr_NoTax	summed frequency of production of all the features listed for a concept, excluding taxonomic features
M_Prod_Fr_NoTax	mean production frequency of the features listed for that concept, excluding taxonomic features
n_Disting_Feat	number of distinguishing features per concept (those in only 1 concept in the norms), including taxonomic features
%_Disting_Feat	percentage of distinguishing features (those in only 1 concept in the norms), including taxonomic features
n_Disting_Feat_NoTax	number of distinguishing features per concept (those in only 1 concept in the norms), excluding taxonomic features
%_Disting_Feat_NoTax	percentage of distinguishing features (those in only 1 concept in the norms), excluding taxonomic features
M_Distinctiveness	mean distinctiveness of the concept's features (1/number of concepts in which features occur), including taxonomic features
M_Distinctiveness_NoTax	mean distinctiveness of the concept's features (1/number of concepts in which features occur), excluding taxonomic features
M_Cue_Val	mean cue validity of the concept's features, including taxonomic features
M_Cue_Val_NoTax	mean cue validity of the concept's features, excluding taxonomic features
n_Corr_Pairs	number of correlated feature pairs per concept, including taxonomic features
n_poss_Corr_Pairs	number of possible pairwise combinations for a feature (excluding those listed for only one or two concept), including taxonomic features
%_Corr_Pairs	percentage of correlated feature pairs per concept, including taxonomic features
n_Corr_Pairs_NoTax	number of correlated feature pairs per concept, excluding taxonomic features
n_poss_Corr_Pairs_NoTax	number of possible pairwise combinations for a feature (excluding those listed for only one or two concept), excluding taxonomic features
%_Corr_Pairs_NoTax	percentage of correlated feature pairs per concept, excluding taxonomic features
Intercorr_Density	intercorrelational density: sum of the proportion of shared variance across all correlated feature pairs for a given concept, including taxonomic features
M_Intercorr_Density	mean intercorrelational density: ratio between the intercorrelational density and the number of correlated feature pairs per concept, including taxonomic features
Intercorr_Density_NoTax	intercorrelational density: sum of the proportion of shared variance across all correlated feature pairs for a given concept, excluding taxonomic features
M_Intercorr_Density_NoTax	mean intercorrelational density: ratio between the intercorrelational density and the number of correlated feature pairs per concept, excluding taxonomic features
M_Relevance	mean relevance of each concept's feature, including taxonomic features

Table 4 (continued)

Variable	Explanation
M_Relevance_NoTax	mean relevance of each concept's feature, excluding taxonomic features
M_Significance	mean significance of each concept's feature, including taxonomic features
M_Significance_NoTax	mean significance of each concept's feature, excluding taxonomic features
n_Sens_Feat	number of sensorial features, including taxonomic features
n_Funct_Feat	number of functional features, including taxonomic features
n_Encyc_Feat	number of encyclopedic features, including taxonomic features
n_Taxon_Feat	number of taxonomic features
%_Sens_Feat	percentage of sensorial features, including taxonomic features
%_Funct_Feat	percentage of functional features, including taxonomic features
%_Encyc_Feat	percentage of encyclopedic features, including taxonomic features
%_Taxon_Feat	percentage of taxonomic features
%_Sens_Feat_NoTax	percentage of sensorial features, excluding taxonomic features
%_Funct_Feat_NoTax	percentage of functional features, excluding taxonomic features
%_Encyc_Feat_NoTax	percentage of encyclopedic features, excluding taxonomic features
n_Sens_Feat (Garrard)	number of sensorial features according to Garrard et al.'s (2001) taxonomy, including taxonomic features
n_Funct_Feat (Garrard)	number of functional features according to Garrard's taxonomy, including taxonomic features
n_Encyc_Feat (Garrard)	number of encyclopedic features according to Garrard's taxonomy, including taxonomic features
n_Taxon_Feat (Garrard)	number of taxonomic features according to Garrard's taxonomy, including taxonomic features
%_Sens_Feat (Garrard)	percentage of sensorial features according to Garrard's taxonomy, including taxonomic features
%_Funct_Feat (Garrard)	percentage of functional features according to Garrard's taxonomy, including taxonomic features
%_Encyc_Feat (Garrard)	percentage of encyclopedic features according to Garrard's taxonomy, including taxonomic features
%_Taxon_Feat (Garrard)	percentage of taxonomic features according to Garrard's taxonomy, including taxonomic features
%_Sens_Feat_NoTax (Garrard)	percentage of sensorial features according to Garrard's taxonomy, excluding taxonomic features
%_Funct_Feat_NoTax (Garrard)	percentage of functional features according to Garrard's taxonomy, excluding taxonomic features
%_Encyc_Feat_NoTax (Garrard)	percentage of encyclopedic features according to Garrard's taxonomy, excluding taxonomic features
M_Cos_Domin	mean cosine distance between a concept and all other concepts, calculated from the dominance matrix
M_Euc_Domin	mean Euclidean distance between a concept and all other concepts, calculated from the dominance matrix
M_City_Domin	mean city block distance between a concept and all other concepts, calculated from the dominance matrix
M_Cos_Binary	mean cosine distance between a concept and all other concepts, calculated from the binary matrix
M_Euc_Binary	mean Euclidean distance between a concept and all other concepts, calculated from the binary matrix
M_City_Binary	mean city block distance between a concept and all other concepts, calculated from the binary matrix
n_Neighb	number of concepts that share at least one feature with a concept
M_Cos_Near_Neighb	mean cosine proximity of the four closest semantic neighbors of a concept

Appendix D

Table 5 Explanations of the sheets included in the Distance_Matrix.xls file

Sheet	Explanation
Cos_Domin	cosine distances between each pairwise combination of concepts (calculated from the feature by concept dominance matrix)
Euc_Domin	Euclidean distances between each pairwise combination of concepts (calculated from the feature by concept dominance matrix)
City_Domin	city block distances between each pairwise combination of concepts (calculated from the feature by concept dominance matrix)
Cos_Binary	cosine distances between each pairwise combination of concepts (calculated from the feature by concept applicability matrix)
Euc_Binary	Euclidean distances between each pairwise combination of concepts (calculated from the feature by concept applicability matrix)
City_Binary	city block distances between each pairwise combination of concepts (calculated from the feature by concept applicability matrix)

References

- Ashcraft, M. H. (1978). Property dominance and typicality effects in property statement verification. *Journal of Verbal Learning and Verbal Behavior*, *17*, 155–164. doi:10.1016/S0022-5371(78)90119-6
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660. doi:10.1017/S0140525X99002149
- Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. doi:10.1146/annurev.psych.59.103006.093639
- Barsalou, L. W., Simmons, W. K., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences*, *7*, 84–91. doi:10.1016/S1364-6613(02)00029-3
- Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*(3, Pt. 2), 1–46. doi:10.1037/h0027577
- Berkhin, P. (2006). Survey of clustering data mining techniques. In J. Kogan, C. K. Nicholas, & M. Teboulle (Eds.), *Grouping multidimensional data: Recent advances in clustering* (pp. 25–71). New York: Springer. doi:10.1007/3-540-28349-8_2
- Bertinetto, P. M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., & Thornton, A. M. (2005). *Corpus e lessico di frequenza dell'italiano scritto (CoLFIS)* [Corpus and frequency lexicon for written Italian]. Retrieved from www.istc.cnr.it/material/database/colfis
- Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications*. New York: Springer.
- Bourne, L. E., Jr., & Restle, F. (1959). Mathematical theory of concept identification. *Psychological Review*, *66*, 278–296. doi:10.1037/h0041365
- Bueno, S., & Megherbi, H. (2009). French categorization norms for 70 semantic categories and comparison with Van Overschelde et al.'s (2004) English norms. *Behavior Research Methods*, *41*, 1018–1028. doi:10.3758/BRM.41.4.1018
- Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*, 1–34. doi:10.1162/089892998563752
- Casey, P. J. (1992). A reexamination of the roles of typicality and category dominance in verifying category membership. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 823–834. doi:10.1037/0278-7393.18.4.823
- Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *8*, 240–247. doi:10.1016/S0022-5371(69)80069-1
- Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 643–658. doi:10.1037/0278-7393.32.4.643
- Cree, G. S., & 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). *Journal of Experimental Psychology: General*, *132*, 163–201. doi:10.1037/0096-3445.132.2.163
- Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. *Cognitive Science*, *23*, 371–414. doi:10.1016/S0364-0213(99)00005-1
- De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M. J., Voorspoels, W., & Storms, G. (2008). Exemplar by feature applicability matrices and other Dutch normative data for semantic concepts. *Behavior Research Methods*, *40*, 1030–1048. doi:10.3758/BRM.40.4.1030
- Dell'Acqua, R., Lotto, L., & Job, R. (2000). Naming times and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behavior Research Methods, Instruments, & Computers*, *32*, 588–615. doi:10.3758/BF03200832
- Della Rosa, P. A., Catricalà, E., Vigliocco, G., & Cappa, S. F. (2010). Beyond the abstract–concrete dichotomy: Mode of acquisition, concreteness, imageability, familiarity, age of acquisition, context availability, and abstractness norms for a set of 417 Italian words. *Behavior Research Methods*, *42*, 1042–1048. doi:10.3758/BRM.42.4.1042
- Devlin, J. T., Gonnerman, L. M., Andersen, E. S., & Seidenberg, M. S. (1998). Category-specific semantic deficits in focal and widespread brain damage: A computational account. *Journal of Cognitive Neuroscience*, *10*, 77–94. doi:10.1162/089892998563798
- Dry, M. J., & Storms, G. (2009). Similar but not the same: A comparison of the utility of directly rated and feature-based similarity measures for generating spatial models of conceptual data. *Behavior Research Methods*, *41*, 889–900. doi:10.3758/BRM.41.3.889
- Duncan, C. P. (1977). Effects of frequency on retrieval from a semantic category. *Bulletin of the Psychonomic Society*, *10*, 57–59.
- Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*, 339–357. doi:10.1037/0096-3445.120.4.339
- Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*, 627–635. doi:10.1016/S0022-5371(73)80042-8
- Funnell, E. (1995). Objects and properties: A study of the breakdown of semantic memory. *Memory*, *3*, 497–518. doi:10.1080/09658219508253162
- 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*, 125–174. doi:10.1080/02643290042000053
- Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*, 256–281. doi:10.1037/0096-3445.113.2.256
- Glass, A. L., & Meany, P. J. (1978). Evidence for two kinds of low-typical instances in a categorization task. *Memory & Cognition*, *6*, 622–628. doi:10.3758/BF03198252
- Graves, W. W., Grabowski, T. J., Mehta, S., & Gordon, J. K. (2007). A neural signature of phonological access: Distinguishing the effects of word frequency from familiarity and length in overt picture naming. *Journal of Cognitive Neuroscience*, *19*, 617–631. doi:10.1162/jocn.2007.19.4.617
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavior and Brain Sciences*, *27*, 377–396. doi:10.1017/S0140525X04000093
- Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441–461. doi:10.1016/S0022-5371(79)90246-9
- Hampton, J. A. (2007). Typicality, graded membership, and vagueness. *Cognitive Science*, *31*, 355–384.
- Hampton, J. A., Dubois, D., & Yeh, W. (2006). Effects of classification context on categorization in natural categories. *Memory & Cognition*, *34*, 1431–1443. doi:10.3758/BF03195908
- Hampton, J. A., & Gardiner, M. M. (1983). Measures of internal category structure: A correlational analysis of normative data. *British Journal of Psychology*, *74*, 491–516. doi:10.1111/j.2044-8295.1983.tb01882.x
- Heit, E., & Barsalou, L. W. (1996). The instantiation principle in natural language categories. *Memory*, *4*, 413–451.
- Hernández-Muñoz, N., Izura, C., & Ellis, A. W. (2006). Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology*, *18*, 730–755. doi:10.1080/09541440500339119

- Humphreys, G. W., & Forde, E. M. (2001). Hierarchies, similarity, and interactivity in object-recognition: On the multiplicity of “category-specific” deficits in neuropsychological populations. *The Behavioral and Brain Sciences*, 24, 453–509.
- Izura, C., Hernández-Muñoz, N., & Ellis, A. W. (2005). Category norms for 500 Spanish words in five semantic categories. *Behavior Research Methods*, 37, 385–397. doi:10.3758/BF03192708
- Jackendoff, R. (1990). *Semantic structures*. Cambridge: MIT Press.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York: Oxford University Press.
- Keller, D., & Kellas, G. (1978). Typicality as a dimension of encoding. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 78–85. doi:10.1037/0278-7393.4.1.78
- Koivisto, M., & Laine, M. (1999). Strategies of semantic categorization in the cerebral hemispheres. *Brain and Language*, 66, 341–357. doi:10.1006/brln.1998.2033
- Kremer, G., & Baroni, M. (2011). A set of semantic norms for German and Italian. *Behavior Research Methods*, 43, 97–109. doi:10.3758/s13428-010-0028-x
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22–44. doi:10.1037/0033-295X.99.1.22
- Larochelle, S., & Pineau, H. (1994). Determinants of response times in the semantic verification task. *Journal of Memory and Language*, 33, 796–823. doi:10.1006/jmla.1994.1038
- López-Chávez, J., & Strassburger Frías, C. (1987). Otro cálculo del índice de disponibilidad léxica [Another index for calculating lexical availability]. In *Presente, y perspectiva de la investigación computacional en México: Actas del IV Simposio de la Asociación Mexicana de Lingüística Aplicada*. Mexico City: UNAM.
- Lynott, D., & Connell, L. (2009). Modality exclusivity norms for 423 object properties. *Behavior Research Methods*, 41, 558–564. doi:10.3758/BRM.41.2.558
- Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23, 250–269. doi:10.1016/S0022-5371(84)90170-1
- Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language*, 32, 517–535. doi:10.1006/jmla.1993.1027
- Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic level categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 54–70. doi:10.1037/0278-7393.23.1.54
- Marques, J. F. (2007). The general/specific breakdown of semantic memory and the nature of superordinate knowledge: Insights from superordinate and basic-level feature norms. *Cognitive Neuropsychology*, 24, 879–903. doi:10.1080/02643290701789436
- Martin, A., & Chao, L. L. (2001). Semantic memory and the brain: Structure and processes. *Current Opinion in Neurobiology*, 11, 194–201. doi:10.1016/S0959-4388(00)00196-3
- McCloskey, M. (1980). The stimulus familiarity problem in semantic memory research. *Journal of Verbal Learning and Verbal Behavior*, 19, 485–502. doi:10.1016/S0022-5371(80)90330-8
- McFarland, C. E., Jr., Duncan, E. M., & Kellas, G. (1978). Isolating the typicality effect in semantic memory. *Quarterly Journal of Experimental Psychology*, 30, 251–262. doi:10.1080/14640747808400673
- McFarland, C. E., Jr., Kellas, G., Klueger, K., & Juola, J. (1974). Category similarity, instance dominance, and categorization time. *Journal of Verbal Learning and Verbal Behavior*, 13, 698–708. doi:10.1016/S0022-5371(74)80056-3
- McRae, K., & Cree, G. S. (2002). Factors underlying category-specific semantic deficits. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category specificity in brain and mind* (pp. 211–249). Hove: Psychology Press.
- 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, Instruments, & Computers*, 37, 547–559. doi:10.3758/BF03192726
- McRae, K., Cree, G. S., Westmacott, R., & de Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology*, 53, 360–373. doi:10.1037/h0087323
- McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99–130. doi:10.1037/0096-3445.126.2.99
- McRae, K., Jared, D., & Seidenberg, M. S. (1990). On the roles of frequency and lexical access in word naming. *Journal of Memory and Language*, 29, 43–65. doi:10.1016/0749-596X(90)90009-O
- Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, 32, 89–115. doi:10.1146/annurev.ps.32.020181.000513
- Minsky, M. (1975). A framework for representing knowledge. In P. H. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw-Hill.
- Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 65–79. doi:10.1037/0278-7393.34.1.65
- Moss, H. E., Tyler, L. K., & Devlin, J. T. (2002). The emergence of category-specific deficits in a distributed semantic system. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category-specificity in brain and mind* (pp. 115–145). Hove: Psychology Press.
- Norman, D. A., & Rumelhart, D. E. (1975). *Explorations in cognition*. San Francisco: Freeman.
- Osherson, D. N., Smith, E. E., Wilkie, O., López, A., & Shafir, E. (1990). Category-based induction. *Psychological Review*, 97, 185–200. doi:10.1037/0033-295X.97.2.185
- Pexman, P., Hargreaves, I., Siakaluk, P., Bodner, G., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, 15, 161–167. doi:10.3758/PBR.15.1.161
- Plaut, D. C. (2002). Graded modality-specific specialization in semantics: A computational account of optic aphasia. *Cognitive Neuropsychology*, 19, 603–639. doi:10.1080/02643290244000112
- Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, 111, 205–235. doi:10.1037/0033-295X.111.1.205
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge: MIT Press.
- Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology*, 104, 192–233. doi:10.1037/0096-3445.104.3.192
- Rosch, E. (1977). Human categorization. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 1, pp. 1–49). London: Academic Press.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Hillsdale: Erlbaum.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605. doi:10.1016/0010-0285(75)90024-9
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439. doi:10.1016/0010-0285(76)90013-X

- Ruts, W., De Deyne, S., Ameel, E., Vanpaemel, W., Verbeemen, T., & Storms, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behavior Research Methods, Instruments, & Computers*, *36*, 506–515. doi:10.3758/BF03195597
- Saffran, E. M., & Sholl, A. (1999). Clues to the functional and neural architecture of word meaning. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 241–271). Oxford: Oxford University Press.
- Sartori, G., Gnoato, F., Mariani, I., Prioni, S., & Lombardi, L. (2007). Semantic relevance, domain specificity and the sensory/functional theory of category specificity. *Neuropsychologia*, *45*, 966–976. doi:10.1016/j.neuropsychologia.2006.08.028
- Sartori, G., & Lombardi, L. (2004). Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience*, *16*, 439–452. doi:10.1162/089892904322926773
- Sartori, G., Lombardi, L., & Mattiuzzi, L. (2005). Semantic relevance best predicts normal and abnormal name retrieval. *Neuropsychologia*, *43*, 754–770. doi:10.1016/j.neuropsychologia.2004.08.001
- Sloman, S. A., Love, B. C., & Ahn, W. (1998). Feature centrality and conceptual coherence. *Cognitive Science*, *22*, 189–228. doi:10.1016/S0364-0213(99)80039-1
- Smith, E. E., & Medin, D. L. (1981). *Categories and concepts*. Cambridge: Harvard University Press.
- Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, *81*, 214–241. doi:10.1037/h0036351
- Solomon, K. O., & Barsalou, L. W. (2001). Representing properties locally. *Cognitive Psychology*, *43*, 129–169. doi:10.1006/cogp.2001.0754
- Storms, G. (2001). Flemish category norms for exemplars of 39 categories: A replication of the Battig and Montague (1969) category norms. *Psychologica Belgica*, *41*, 145–168.
- Storms, G., De Boeck, P., & Ruts, W. (2000). Prototype and exemplar-based information in natural language categories. *Journal of Memory and Language*, *42*, 51–73. doi:10.1006/jmla.1999.2669
- Storms, G., De Boeck, P., Van Mechelen, I., & Ruts, W. (1996). The dominance effect in concept conjunctions: Generality and interaction aspects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1266–1280. doi:10.1037/0278-7393.22.5.1266
- Taylor, K. I., Moss, H. E., Randall, B., & Tyler, L. K. (2004). The interplay between distinctiveness and intercorrelation in the automatic activation of word meaning. In *Abstracts of the Psychonomic Society* (Vol. 9, p. 109). Austin: Psychonomic Society
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York: Academic Press.
- van Dantzig, S., Cowell, R. A., Zeelenberg, R., & Pecher, D. (2011). A sharp image or a sharp knife: Norms for the modality-exclusivity of 774 concept-property items. *Behavior Research Methods*, *43*, 145–154. doi:10.3758/s13428-010-0038-8
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*, 289–335. doi:10.1016/j.jml.2003.10.003
- Vigliocco, G., Vinson, D. P., Damian, M. F., & Levelt, W. J. M. (2002). Semantic distance effects on object and action naming. *Cognition*, *85*, B61–B69. doi:10.1016/S0010-0277(02)00107-5
- Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic system hypothesis. *Cognitive Psychology*, *48*, 422–488. doi:10.1016/j.cogpsych.2003.09.001
- Vigliocco, G., Warren, J., Siri, S., Arciuli, J., Scott, S., & Wise, R. (2006). The role of semantics and grammatical class in the neural representation of words. *Cerebral Cortex*, *16*, 1790–1796. doi:10.1093/cercor/bhj115
- Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods*, *40*, 183–190. doi:10.3758/BRM.40.1.183
- Zannino, G. D., Perri, R., Pasqualetti, P., Caltagirone, C., & Carlesimo, A. G. (2006). Analysis of the semantic representations of living and nonliving concepts: A normative study. *Cognitive Neuropsychology*, *23*, 515–540. doi:10.1080/02643290542000067