NEUROPSYCHOLOGIA

Neuropsychologia 43 (2005) 754–770

www.elsevier.com/locate/neuropsychologia

Semantic relevance best predicts normal and abnormal name retrieval

G. Sartori[a,*], L. Lombardi[b], L. Mattiuzzi[a]

[a] Dipartimento di Psicologia Generale, Università di Padova, Via Venezia 8, I-35100 Padova, Italy
[b] Dipartimento di Scienze della Cognizione e della Formazione, Università di Trento, via Matteo del Ben 5, I-38068 Rovereto (TN), Italy

Received 24 January 2004; received in revised form 30 July 2004; accepted 3 August 2004

Abstract

The relevance of a semantic feature measures its contribution to the "core" meaning of a concept. In a naming-to-description task, we investigated the predictive power of relevance in comparison with frequency, familiarity, typicality, and Age-of-Acquisition. In a group of Alzheimer patients with semantic disorder, relevance turned out to be the best predictor of name retrieval accuracy in a naming-to-description task. The same pattern of results was observed in normal controls. Relations between semantic relevance and the parameters of the concepts are discussed in order to highlight the mechanism of concept activation in a naming-to-description task.

Keywords: Semantic relevance; Concept representation; Semantic disorders

1. Introduction

Concepts are believed to be organized networks of semantic features. It has been argued that concepts may differ along several dimensions, such as the age they were acquired, the frequency with which they appear in various contexts, etc. These factors are known to affect naming performance in normal subjects as well as that of patients with disorders of conceptual knowledge. Indeed, name retrieval is more accurate for concepts acquired earlier in life, prototypical concepts, and frequent ones (Jolicoeur, Gluck, & Kosslyn, 1984; Lambon Ralph, Graham, Ellis, & Hodges, 1998). At the semantic feature level, the roles of type, distinctiveness, and inter-correlation of features in the representation of concepts has been strongly emphasized as determinants of naming accuracy (Moss, Tyler, Durrant-Peatfield, & Bunn, 1998; Warrington & Shallice, 1984).

Accordingly, the dimensions of conceptual knowledge which are believed to lie at the base of semantic disorders may be classified into: (i) dimensions that encode concepts, and (ii) dimensions that encode semantic features (see Cree & McRae, 2003, for a review). Dimensions of concepts include parameters like frequency, familiarity, and typicality, and their effects are now well established (Kremin et al., 2003). The dimensions of semantic features include: (1) distinctiveness, which scores high when a semantic feature is used in defining few concepts; (2) dominance, which scores high when the semantic feature is frequently mentioned by subjects in defining a concept; (3) semantic relevance, which scores high when a semantic feature is both frequently mentioned in defining a concept, but only mentioned in defining few other concepts.[1]

Semantic relevance (Sartori & Lombardi, 2004) is a parameter indexing the importance of a semantic feature in concept identification. To illustrate, the Romans called the giraffe CAMELOPARDALIS;[2] this name probably derives from the fact that a giraffe resembles a camel in its long neck, and also resembles a leopard in its spotted coat. Other examples may be found in neuroanatomy: FALX, HIPPOCAMPUS, CAUDATE, etc. Important features used as names may not only be sensory, as in the previous examples, but also abstract. Let us consider the concept PLATYPUS, as described by early Western visitors to Australia. Its scientific name,

* Corresponding author. Tel.: +39 0464 483558; fax: +39 0464 483554.
E-mail addresses: giuseppe.sartori@unipd.it (G. Sartori), luigi.lombardi@unitn.it (L. Lombardi).

[1] Details on the parameters indicated here will be presented later in Section 2.3.

[2] Concept names are printed in uppercase (e.g., DOG) and names of semantic features in angled brackets (e.g., ⟨has a tail⟩).

0028-3932/$ – see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2004.08.001

Ornithorhyncus paradoxus, contains the most important characteristic, at least for a Westerner: paradoxus refers to the strange associations of semantic features such as ⟨can breed⟩ and ⟨has a beak⟩, never previously encountered by Westerners.

The notion of relevance of semantic features is intended to capture the importance of a given semantic feature in the distinction of one concept from other similar ones (Sartori & Lombardi, 2004). Relevance-based approaches lie implicitly at the base of a great deal of theorizing about semantic knowledge (Rosch, 1975; Warrington & Shallice, 1984). For example, according to Warrington and Shallice (1984) Living items rely more on perceptual semantics, whereas Non-living ones rely more on functional features. Sartori and Lombardi (2004) proposed a model to measure the relevance of a semantic feature for a concept, in which concepts are represented by a vector of semantic features, and relevance is a measure of the contribution of semantic features to the "core" meaning of a concept. Semantic features with high relevance are those which are useful for distinguishing the target concept from similar concepts. In fact, when we are asked to define a concept, we usually do not list all its semantic features, but only those useful to differentiate it from closely related concepts. Vocabulary definitions are organized in this way. For example, ⟨has a trunk⟩ is a semantic feature of high relevance for the concept ELEPHANT because most subjects use it to define this concept, whereas very few people use the same feature to define other concepts; on the other hand, ⟨has 4 legs⟩ is a semantic feature with lower relevance for the same concept, because few subjects use it in the definition of ELEPHANT but do apply it to many other concepts. Sartori and Lombardi (2004) have proposed a procedure to derive algorithmically relevance values from concepts descriptions and this procedure is described in Appendix A.

In determining naming accuracy, knowing the relative weight of these parameters of concepts and parameters of semantic features will contribute towards highlighting the mechanisms involved in normal and abnormal name retrieval. Although the effects of variables measuring concepts have been extensively investigated, the role of some parameters which describe semantic features are not yet clear. In the same way, studies that compare the influence, in naming, of parameters of concepts and parameters of semantic features are not available to our knowledge. Which factors are most important in this caldron of contributing ingredients will be the main object of the paper.

Specifically, this article is concerned with the relations among those dimensions which are believed to have an effect in semantic tasks in normal and semantically impaired patients. Specifically, we address the role, in a naming-to-description task, of parameters of the concepts and semantic features in predicting identification accuracy. To anticipate our results, we show that semantic relevance is the best predictor of accuracy in name retrieval, in both patients and normal controls, when compared with variables such as frequency, familiarity, and Age-of-Acquisition.

The method adopted here will allow three other related issues to be addressed in Section 4. The first concerns the best way to describe the internal structure of concepts, and also regards the mechanisms by means of which semantic features activate concepts. The second issue concerns the relational structure among those psycholinguistic parameters which are usually considered to be descriptors of concepts. Cree and McRae (2003) suggested a multifactorial theory of normal and abnormal semantic memory, in which several of the factors considered here are believed to contribute to the computation of meaning, and also to lie at the origin of semantic memory disorders. Conversely, we discuss the possibility that some of these factors manifest their effect on accuracy through the intervention of semantic relevance. The third issue regards the relation between the organization of semantic in healthy controls as compared to degraded semantic as observed in Alzheimer's patients. A DAT group is contrasted with that of healthy controls in order to verify if the same dimensions that affect performance of controls are the same that affect DAT's performance.

2. Materials and method

2.1. Participants

Data collected on two groups were used in this study: (i) a group of 15 patients with diagnoses of dementia of Alzheimer's type (DAT) (mean age = 75.6 years, S.D. = 7.72; mean education = 6.6 years, S.D. = 4.64), and (ii) a control group of 37 normal controls (mean age = 76.49 years, S.D. = 6.78; mean education = 5.03 years, S.D. = 1.25), matched for age and education to the DAT group. All participants were native Italian speakers. Some degree of semantic impairment is commonly seen in the early stages of dementia of Alzheimer's type (Chertkow & Bub, 1990; Hodges & Patterson, 1995), and this investigation was conducted on DAT patients with this characteristic. The 15 DAT patients (12 women, 3 men) met the National Institute of Neurological and Communicative Disorders and Stroke/Alzheimer's Disease and Related Disorders Association (NINCDS/ADRDA) criteria for probable Alzheimer's disease (McKhann et al., 1984). All 15 patients had Hachinski scores (Hachinski et al., 1975) below 4 and an MMSE (Folstein, Folstein, & Mc Hugh, 1975) below 24/30. All DATs were at least 2 S.D. below average scores of the normative sample on two anterograde and two semantic memory tests (see Table 1). All underwent CT or MRI scanning, together with a screening battery to exclude treatable causes of dementia. Patients with major depression, past history of known stroke or TIA, alcoholism, head injury or major medical illnesses were excluded. Patients were recruited in three hospitals and four nursing homes located in the Veneto (North-East Italy). The background neuropsychological data collected on participants are given in Table 1. Although premorbid IQ (as measured by TIB, an Italian analog of NART) did not differ between DATs and controls, the DAT group

Table 1
Background neuropsychological and semantic memory screening tests for group of DAT patients and normal controls

Description	DAT				Controls				Difference
	Mean	S.D.	Minumum	Maximum	Mean	S.D.	Minumum	Maximum	
Neuropsychology tests									
MMSE[a] correct maximum = 30	18.69	2.48	13.20	24.10	26.20	1.45	24	29.30	$p < .05$
T.I.B.[b] premorbid IQ	90.35	12.28	77.36	114.05	93.03	8.55	81.15	110.01	$p = .37$
Prose memory test[c]	1.73	0.70	0.50	3	9.51	3.47	1	16	$p < .05$
Incid. phon. memory (maximum = 20)[d]	1.21	0.70	0	2	3.62	1.47	1	8	$p < .05$
Semantic memory tests									
Picture naming[d]									
Non-living (%) ($N = 32$)	55	11.56	25	71.87	85.33	10.18	50	100	$p < .05$
Living (%) ($N = 32$)	46.25	22.91	12.50	90.62	84.11	10.57	53.12	100	$p < .05$
Naming to description[e]									
Verbal description (%) ($N = 14$)	54.28	13.71	21.43	71.43	93.45	9.41	71.43	100	$p < .05$
Visual description (%) ($N = 11$)	22.42	16.43	0	54.54	79.54	14.80	36.36	100	$p < .05$

[a] Mini Mental State Examination (Folstein et al., 1975); corrected score.
[b] Premorbid IQ, Italian analog of NART (Sartori, Colombo, Vallar, Rusconi, & Pinarello, 1995).
[c] Spinnler and Tognoni (1987).
[d] Sartori, Job, and Zago (2002).
[e] Silveri and Gainotti (1988).

had lower scores on the MMSE and anterograde memory tests (Prose memory, Phonemic incidental memory). DATs had also a poorer performance than controls on picture naming and naming-to-description tests, used here as semantic memory screening tests.

2.2. Naming-to-description task

Naming-to-description is largely used in investigations on patients with semantic disorders (e.g., Lambon Ralph et al., 1998; Silveri & Gainotti, 1988). In contrast to picture naming, it has the advantage of allowing full control over the presented semantic features[3]. The task consists of presenting participants with a sentence describing the target concept, and including a set of three semantic features. For example the sentence “has an handle, has two wheels and has two pedals” was presented orally to the partecipant who was required to retrieve the name BICYCLE. Semantic features could be of any type including, perceptual, associative, encyclopaedic and functional features. The response was scored either correct or wrong. For each concept, accuracy was calculated for DATs and controls separately, and the relation between accuracy, the parameters of the concept and of the semantic features was the object of this investigation.

2.3. Stimuli and procedure

One hundred concepts were used, randomly selected from a larger pool of 254 concepts, for which the set of parameters described in the next section was either available or computed expressly. The database of 254 concepts included 13 categories (i.e., birds, buildings, clothes, flowers, furniture, fruits, houses, wares, mammals, musical instruments, vegetables, vehicles and weapons; Dell’Acqua, Lotto, & Job, 2000). The number of elements in each category[4] varied from 11 to 32.

The 100 concepts selected for use in this investigation guaranteed: (i) a sufficient number of stimuli in order to run the necessary analyzes, (ii) a test suited to DAT patients, and (iii) the various parameters under investigation spanned all ranges. Each concept was described by a sentence consisting of three semantic features randomly selected from the set of all features that applied to the target concept. The three semantic features were presented orally to the participants who were required to retrieve the corresponding concept. The required responses were oral.

2.3.1. Parameters of concepts and parameters of semantic features

As indicated earlier, the structure of the concepts may be analyzed from several points of view. Here, for each concept and its corresponding description (consisting of three semantic features), a number of parameters were considered and used to predict naming accuracy. These parameters were classified into parameters of concepts and parameters of semantic features.

[3] No computational models are available, to our knowledge, which account for spreading activations from the presented features to not presented ones prior to name retrieval. This is a limit that should be kept in mind when generalizing the results.

[4] The database used here and other published databases (e.g., Cree & McRae, 2003; Garrard, Lambon Ralph, Hodges, & Patterson, 2001; Vigliocco, Vinson, Lewis, & Garrett, 2004) included concepts selected on the belief that the organization of semantic memory is based on categories (e.g., Warrington & Shallice, 1984). This was also the assumption that guided the selection of concepts included in our database.

2.3.1.1. Parameters of concepts. These are parameters which do not take into account semantic features, but are rather estimates of the "difficulty" of concepts measured from different facets. Some of them have well-established effects on naming performance, and those considered here are:

(1) *Frequency:* This refers to the frequency with which a word is encountered in adult language. Thus, frequency norms, such as those used here and reported in Dell'Acqua et al. (2000), reflect how often words are used. The influence of frequency in naming is well established in both normal and neurological populations. Indeed, high-frequency words are retrieved more quickly and accurately than low-frequency ones (McRae, Jared, & Seidenberg, 1990).

(2) *Familiarity:* Familiarity is a context-free measure related to the amount of experience with the concept (Mandler, 1980). It is usually rated by subjects, and here we used the norms collected by Dell'Acqua et al. (2000). Highly familiar items are named more accurately, and familiarity may influence both normal and patients' response accuracy (e.g., Funnell & Sheridan, 1992).

(3) *Age-of-Acquisition:* Age-of-Acquisition is a measure of how early in life a certain concept is acquired. One measure of when children have actually acquired words has been provided by Morrison, Chappell, & Ellis (1997) who asked children of various ages to name picture. Measures of Age-of-Acquisition have generally been collected by means of adult estimations of when they learned particular words (e.g., Gilhooly & Gilhooly, 1980). It is striking that such Age-of-Acquisition ratings correlate impressively highly with more objective measures of the age at which words are actually learned (Carroll & White, 1973; De Moor, Ghyselinck, & Brysbaert, 2001; Gilhooly & Gilhooly, 1980; Jorm, 1991; Lyons, Teer, & Rubenstein, 1978; Morrison et al., 1997; Pind, Jonsdottir, Tryggvadottir, & Jonsson, 2000), which suggests that the ratings are valid.

    The norms used here were taken from Dell'Acqua et al. (2000) and were based on subjective estimations. Words learned early in life can be recognized and produced faster than later-learned words. This effect has been observed in a variety of tasks including picture naming (Barry, Morrison, & Ellis, 1997) and reading aloud (Gerhand & Barry, 1998). The influence of this variable in the performance of semantic patients has recently been reviewed by Capitani, Laiacona, Mahon, and Caramazza (2003).

(4) *Typicality:* This refers to what extent the concept is considered a good representative of a category (Rosch, 1975) and is usually collected through subjective ratings. The norms used here were those reported in Dell'Acqua et al. (2000). Highly typical items are named better by normal and by brain-damaged patients (Jolicoeur et al., 1984; Kiran & Thompson, 2003).

2.3.1.2. Parameters of semantic features. These parameters differ from the previous ones, in that they take into account semantic features. The parameters reported below, rather than based on subjective ratings, were computed by starting from the norms of features derived from a feature-listing task (Rogers et al., 2004).

(1) *Dominance:* This is a measure of how frequently a semantic feature is used in defining a concept (Ashcraft, 1978). Garrard et al. (2001) computed dominance by counting the number of times subjects listed a given feature in defining a concept, divided by the total number of all the instances of those features listed in defining the same concept. To our knowledge, empirical evidence for the predictive role of dominance in naming is not available. Calculation of dominance for the 100 descriptions was carried out using a variant of the procedure of Garrard et al. (2001). The dominance of the concept description results from the sum of raw dominance values of the three semantic features, divided by the total number of occurrences of the features in the concept.

(2) *Distinctiveness:* Highly distinctive semantic features are those which appear in the definition of a few concepts, whereas low distinctive features appear in the definition of many concepts. The distinctiveness of a semantic feature is defined as the complement of sharedness. Sharedness is a normalized factor that is computed by dividing the number of different concepts in which the semantic feature appears by the number of concepts in the database (see Devlin, Gonnerman, Andersen, & Seidenberg, 1998; Tyler, Moss, Durrant-Peatfield, & Levy, 2000). Cree and McRae (2003) introduced the similar notion of distinguishing feature. Following their suggestion we calculated distinctiveness using the full set of 254 concepts and not, as suggested by Garrard et al. (2001), using the contrast set limited to the category which the target concept belonged to. Therefore, distinctiveness is calculated as 1—*sharedness* and ranges between 0 (when the semantic feature appears in all concepts) and approaches 1 (when it appears in one concept only). Garrard et al. (2001) reported distinctiveness calculated within category. Here, to evaluate the relative weight which distinctiveness plays in semantic relevance, it was calculated on all 254 concepts from which the 100 concepts were selected. It has been claimed that distinctiveness modulates typicality judgements (Rosch & Mervis, 1975) and category verification (Smith, Shoben, & Rips, 1974) but, to our knowledge, direct evidence related to how distinctiveness facilitates naming is not available. Also distinctiveness, like dominance, was calculated by summing the values of distinctiveness of the three semantic features of the concept description.

(3) *Semantic relevance:* When a set of semantic features is presented, the overall relevance results from the sum of the individual relevance values associated with each of the semantic features. The concept with the highest

summed relevance is the one which will be retrieved. For instance, the three features ⟨similar to a goose⟩, ⟨lives in ponds⟩ and ⟨has a beak⟩ have, in the database considered here, top relevance for DUCK, followed by SWAN and OSTRICH. Given these three features, the retrieved concept will be DUCK, because it has the highest relevance. It may happen that, in the presence of degraded features of DUCK, SWAN is erroneously retrieved, as resulting in higher relevance than DUCK. Hence, overall accuracy in name retrieval is poor when concepts have low relevance, and when they have many other semantically similar concepts with which they may be confused.

Relevance of semantic features is different from distinctiveness. Distinctiveness is a dimension which is not concept-dependent, since scores are high when the feature is found in only a few concepts. Instead, the relevance of a given semantic feature varies across different concepts and, in a way, may be considered concept-dependent. For example, the feature ⟨has a beak⟩ has higher relevance for the concept DUCK than for the concept SWAN.

As an example of the computational procedure[5] suppose that the semantic feature ⟨yields milk⟩ appears in 7 of 300 concepts, and suppose also that the same feature is listed, by subjects, 12 times in defining the concept COW. The semantic relevance of ⟨yields milk⟩ for COW will be, according to Eq. (A.5), equal to $k = 12 \times \log_2(300/7) = 65.057$.

Sartori and Lombardi (2004) have shown that: (a) concepts are better retrieved when semantic features with higher relevance values are presented; (b) Living items have semantic features with lower relevance than Non-living, thus creating an advantage for Non-living items; (c) Living have perceptual, functional and specific semantic features with lower relevance than Non-living; (d) Living have high semantic similarity between exemplars; (e) animals and vegetables show similar profiles in terms of relevance and within-category similarity; (f) musical instruments have the same relevance, like other objects but very high similarity among exemplars of the category. The heterogeneous distribution of semantic features with different relevance values has been used to explain category specificity for Living, Non-living, and many other effects reported in the neuropsychological literature. The semantic relevance model is in line with the specific impairment for Living, because of the intrinsic characteristics of Living items which have semantic features that are, on average, less relevant than Non-living ones. If matching for relevance is not carried out carefully, exemplars of Living which have lower relevance are likely to be selected, thus reducing response accuracy on semantic tasks. Instead, selecting Non-living items which have lower relevance than Living ones, used as benchmarks, will yield greater impairment for Non-living. Two different measures of semantic relevance were analyzed here: (i) the relevance of the three semantic features presented in the naming-to-description task that results from summing the relevance values of these three semantic features, (ii) total relevance resulting from summing the relevance values of all the semantic features which are listed in defining the concept. Some examples of stimuli used in the verbal-to-description task together with their parameter values are reported in Table 2.

2.4. Statistical methods

The aim of our analysis was two-fold. On one hand, we were interested in how the accuracy of DATs and controls is modulated by concept parameter structure and feature parameter structure. The model consists of two hierarchically connected main components: (1) the semantic structure component as a predictor multivariate variable, which is further divided into two distinct subcomponents (concept structure and feature structure); (2) the retrieval accuracy component as the target-dependent variable. For each subcomponent, we also tested a set of hierarchical relations among the parameters of the structure.

On the other hand, we wanted to test, in the concept structure, how the concept parameters might influence total semantic relevance. Specifically, we wanted to evaluate whether and how total relevance might be mediated through interrelationships among concept parameters.

Both analyses were conducted on a pairwise correlation matrix of the variables represented by the two models. Path analyses on the resulting correlation matrices were performed by using LISREL (Jöreskog & Sorböm, 1993). Following the recommendations of Hu and Bentler (1999), we evaluated model fit using the *non-normed fit index* (NNFI), *root-mean-square error of approximation* (RMSEA), *comparative fit index* (CFI) along with the standard chi-square statistic.[6]

[5] Similarly to distinctiveness, semantic relevance values depend on the total number (I) of concepts in the database, which ideally should correspond to the size of the mental lexicon. This may be a critical point as we ignore on what set of concepts the actual computations of our mind are based. If relevance is greatly influenced by I then its estimation may loose reliability. In our investigation I has a value of 254. To investigate the effects on relevance of varying levels of I, we compared relevance values when computed in a set of 50, 100 and 150 concepts (subsets of the original 254 concepts). Details of the analysis are reported in Section A.2.4. Here it is worth mentioning that relevance values, when computed using these subsamples, predicted very accurately the original relevance values calculated on the 254-concepts database.

[6] The NNFI and CFI offer a way to quantify the degree of fit along a continuum. They are incremental fit indices that measure the proportionate improvement in fit by comparing a target model with a more restricted nested baseline model. In contrast RMSEA is an absolute fit index that assesses how well an a priori model reproduces the sample data. Hu and Bentler (1999) recommended that values exceeding .90 for the NNFI, .06 for the RMSEA, and .08 for the CFI should be used as cutoffs, representing a good fit of the data to the model.

Table 2
Examples of stimuli used

	DIS	DOM	REL	FREQ	FAM	TYP	AA	RELT
Bicycle				2.45	6.87	5.27	2.27	671.74
⟨Has a handle⟩	.99	6	41.93	–	–	–	–	–
⟨Has two wheels⟩	.98	7	39.67	–	–	–	–	–
⟨Has two pedals⟩	.97	7	37.83	–	–	–	–	–
Sum over the three features	2.95	20	119.42	–	–	–	–	–
Cow				1.70	6.13	5.47	2.40	577.08
⟨Similar to a calf⟩	.99	5	39.94	–	–	–	–	–
⟨Has udders⟩	.99	3	23.97	–	–	–	–	–
⟨Yields milk⟩	.97	12	62.18	–	–	–	–	–
Sum over the three features	2.96	20	126.08	–	–	–	–	–

Procedures for calculating dominance, distinctiveness and semantic relevance are reported in Appendix A. REL: semantic relevance; DIS: distinctiveness; DOM: dominance; FREQ: frequency; FAM: familiarity; TYP: typicality; AA: Age-of-Acquisition; RELT: total semantic relevance.

3. Results

3.1. Semantic parameters and observed accuracies

3.1.1. DAT group

As expected, the DAT patients showed a moderate degree of semantic impairment in the naming-to-description task. Their overall accuracy on the 100 items was 29.62%; that of the control group was 68.14% ($t(198) = -9.97$, $p < .001$).

The results of the correlation analysis (Table 3) refer to empirical correlations for DAT patients. In regard to concept structure, Age-of-Acquisition (AA) and frequency (FREQ) were significantly correlated with DAT accuracy (ACC) ($r = -.321$, $p < .01$, for AA and $r = .260$, $p < .01$, for FREQ). Typicality (TYP) did not reach significance ($p = .32$), whereas familiarity (FAM) was close to significance ($p = .06$). As regards feature structure, a moderate relationship was found between dominance (DOM) and ACC ($r = .234$, $p < .05$), whereas a stronger association resulted between semantic relevance (REL) and ACC ($r = .443$, $p < .01$). Lastly, the correlation between distinctiveness (DIS) and ACC turned out to be non-significant ($p = .363$). Similar correlational patterns were also found for the controls, although their relationships were somewhat weaker than those observed for the DATs (see Table 4).

3.1.2. Path analyses

The correlational analyzes discussed above give an overview of the relationships among our variables. However, they do not provide a test of the structure of the relationships. Nor do they provide information regarding unique or incremental relationships above and beyond the variance explained by other variables in the structure. To test the complete structure of the relationships, including estimation of the unique variance explained by each hypothetical link, we evaluated the correlation matrix using structural equation modelling for observed variables (path analysis). The first step was a fully connected model (Fig. 1), in which concept parameters and feature parameters were hypothesized to affect DAT accuracy independently. The chi-square test for this model was significant ($\chi^2(8, N = 100) = 16.56$, $p < .05$), and the fit indices indicated a moderately good fit (NNFI = .84, RMSEA = .11, CFI = .96).

The relevance model makes specific hypotheses about how distinctiveness and dominance are integrated into relevance (see Appendix A). We tested a revised model reproducing the scheme of the relevance formula reported in Eq. (A.5) and discussed in Section A.2.1 by eliminating all paths not implied by that formula. We also removed all non-significant paths on the concept structure side. The final result was a

Table 3
DAT patients: Pearson's correlations among parameters of semantic features (relevance, distinctiveness, dominance), parameters of concepts (frequency, familiarity, typicality, Age-of-Acquisition) and naming accuracy

	ACC	REL	DIS	DOM	FREQ	FAM	TYP	AA
ACC	–							
REL	.443**	––						
DIS	.092(n.s.)	.515**	–					
DOM	.234*	.647**	.159(n.s.)	–				
FREQ	.260**	.054(n.s.)	.081(n.s.)	−.147(n.s.)	–			
FAM	.197(n.s.)	.125(n.s.)	−.023(n.s.)	.015(n.s.)	.260**	–		
TYP	.101(n.s.)	−.141(n.s.)	−.203*	−.154(n.s.)	.324**	.452**	–	
AA	−.321**	−.063(n.s.)	−.107(n.s.)	.162(n.s.)	−.529**	−.480**	−.310**	–

ACC: retrieval accuracy of patients; REL: semantic relevance; DIS: distinctiveness; DOM: dominance; FREQ: frequency; FAM: familiarity; TYP: typicality; AA: Age-of-Acquisition; n.s.: non-significant.

* $p < .05$.

** $p < .01$.

Table 4
Control subjects: Pearson's correlations among parameters of semantic features (relevance, distinctiveness, dominance), parameters of concepts (frequency, familiarity, typicality, Age-of-Acquisition) and naming accuracy

	ACC	REL	DIS	DOM	FREQ	FAM	TYP	AA
ACC	–							
REL	.344**	–						
DIS	.098(n.s.)	.515**	–					
DOM	.185(n.s.)	.647**	.159(n.s.)	–				
FREQ	.171(n.s.)	.054(n.s.)	.081(n.s.)	−.147(n.s.)–				
FAM	.126(n.s.)	.125(n.s.)	−.023(n.s.)	.015(n.s.)	.260**	–		
TYP	−.013(n.s.)	−.141(n.s.)	−.203*	−.154(n.s.)	.324**	.452**	–	
AA	−.243**	−.063(n.s.)	−.107(n.s.)	.162(n.s.)	−.529**	−.480**	−.310**	–

ACC: retrieval accuracy of patients; REL: semantic relevance; DIS: distinctiveness; DOM: dominance; FREQ: frequency; FAM: familiarity; TYP: typicality; AA: Age-of-Acquisition; n.s.: non-significant.

* $p < .05$.

** $p < .01$.

more parsimonious model (Fig. 2) that results in an increase in model fit and which reads as follows: REL is expected to be affected by DIS and DOM. Although the chi-square test for the revised model was significant ($\chi^2(14, N = 100) = 23.80$, $p < .05$), the model demonstrated a good fit with the data (NNFI = .90, RMSEA = .086, CFI = .95), meeting Hu and Bentler's (1999) recommended cutoff for RMSEA, NNFI and CFI.

The reconstructed correlational parameters showed a significant positive correlation between distinctiveness and relevance ($r = .42$, $p < .001$) and between dominance and relevance ($r = .58$, $p < .001$). Further details about the mutual relationships between distinctiveness, dominance, and semantic relevance is given in Section 4 and Appendix A (Section A.2.1).

On the concept structure side, frequency, typicality and familiarity were expected to affect Age-of-Acquisition. The high correlation between Age-of-Acquisition and frequency is a well-established fact (Zevin & Seidenberg, 2004) although complete understanding of it is not available. In particular, effects due to Age-of-Acquisition per se, which are not related to frequency, cannot definitely be observed. For the purpose of this paper, it is important to note that, whereas Age-of-Acquisition is an estimate of the moment in life at which a concept is acquired, frequency is an estimate of the usage of that concept in adulthood. The reconstructed correlational parameters (Fig. 2) showed a significant negative correlation between frequency and Age-of-Acquisition ($r = -.43$, $p < .001$) and between familiarity and Age-of-Acquisition ($r = -.37$, $p < .001$). A final remarkable result was that semantic relevance was clearly the best predictor, when compared with Age-of-Acquisition ($r = .42$, $p < .001$ versus $r = -.29$, $p < .001$).

It is important to note that REL derives from a non-linear combination of DIS and DOM (see Section A.2.1) and this property might actually have influenced the general result as reported above. In order to check whether similar results were obtained using path analyses that do not model the semantic feature-structure we ran new analyses by (a) including all three parameters (REL, DOM, DIS) separately in the path

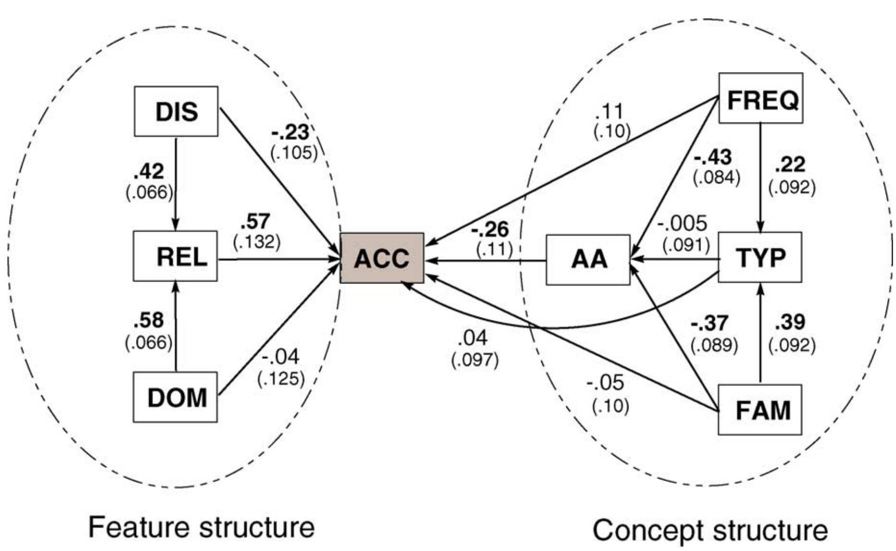

Fig. 1. Initial path model (DAT group) with standardized regression weights. $\chi^2(8, N = 100) = 16.56$, $p < .05$; RMSEA = .11; NNFI = .84; CFI = .96. Standardized regression coefficients in bold are significant at $p < .05$. Values in parentheses are standard errors for the regression coefficient.

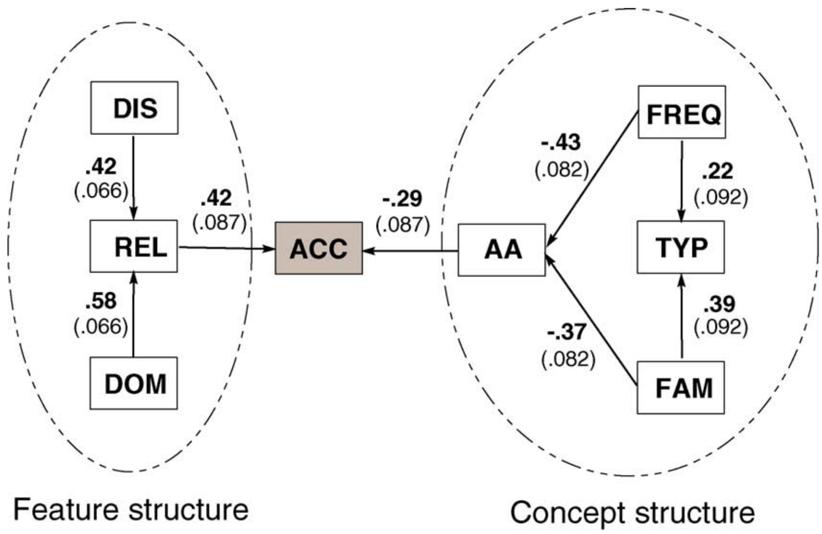

Fig. 2. Final path model (DAT group) with standardized regression weights. $\chi^2(14, N = 100) = 23.80$, $p < .05$; RMSEA = .086; NNFI = .90; CFI = .95. Standardized regression coefficients in bold are significant at $p < .05$. Note: non-significant paths of base model have been removed. Values in parentheses are standard errors for the regression coefficient.

model (b) keeping the rest of the path model-structure (i.e., the structure modelling the relation among the parameters of concepts), unchanged. The new results[7] were in line with what already observed in the main analysis. In particular, (i) REL showed up to be the best predictor among all parameters (ii) DIS (resp. DOM) taken singly did not predict naming accuracy better than AA. Summarizing the ranking as implied by accuracy prediction was as follows:

$$\text{REL}(.42) \succ \text{AA}(-.32) \succ \text{DOM}(.29) \succ \text{DIS}(.06),$$

with the AA value ($-.32$) computed as the average of the three AA values ($-.29$, $-.31$, $-.37$) over the three distinct analyses.

We also evaluated the impact of error in measuring naming accuracy on the DAT group. In order to check whether measurement error and sampling error (e.g., small group size, differing clinical conditions between patients, differing levels of severity between patients, etc.) might have affected our results, we performed an uncertainty analysis. Specifically, we used a new approach developed by Lombardi, Pastore, and Nucci (2004), named Sample Generation by Replacements (SGR), to evaluate the robustness of our results. This method can be used to analyze model acceptability-criteria assuming that the empirical data set is perturbed with predefined levels of error. The results of the uncertainty analysis showed that the qualitative pattern among variables was still observed when a 25% perturbation of naming accuracy was produced. The perturbation was created by replacing approximately 25% of the 15 dichotomous accuracy-responses y with $1 - y$, and this for each of 100 concepts in our study. This analysis indicated that the superiority of relevance in predicting naming accuracy is a robust result, not explained by calling into account measurement and sampling errors (see Fig. 3).

3.1.3. Control group

Similar results were also observed for the control group (Fig. 4), although these relationships were somewhat weaker than those observed for the DATs (Table 4), presumably due

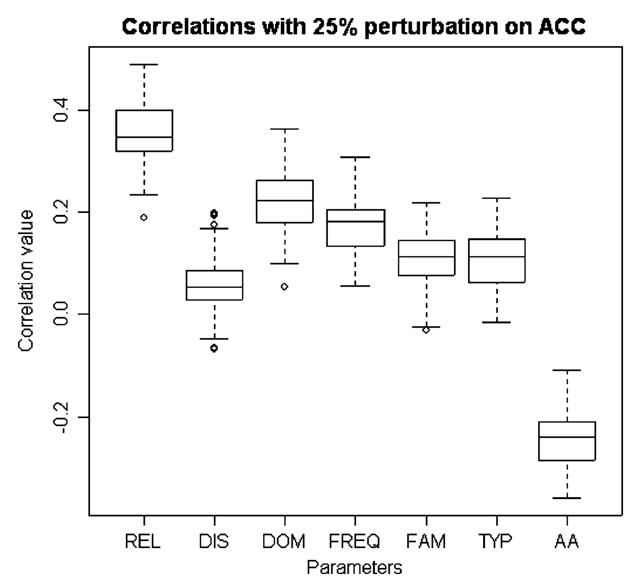

Fig. 3. Path analysis results after running 100 simulations with 25% perturbation on ACC. The figure reports the distributions of the reconstructed LISREL correlations between independent variables (REL, DIS, DOM, FREQ, FAM, TYP, AA) and perturbed ACC.

[7] Path models (DAT group) with, respectively, REL taken singly: $\chi^2(6, N = 100) = 7.24$, $p = .29$; RMSEA = .046; NNFI = .97; CFI = .99; DIS taken singly: $\chi^2(6, N = 100) = 8.67$, $p = .19$; RMSEA = .068; NNFI = .93; CFI = .97; DOM taken singly: $\chi^2(6, N = 100) = 6.09$, $p = .41$; RMSEA = .012; NNFI = 1.00; CFI = 1.00.

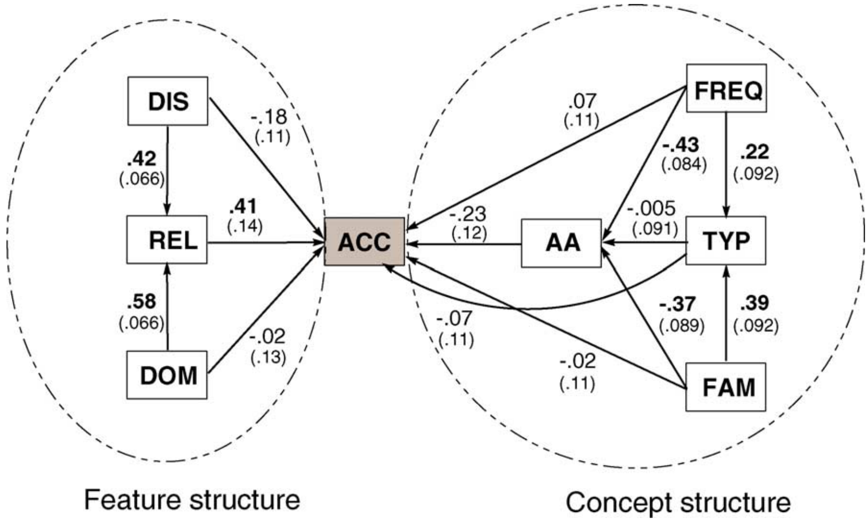

Fig. 4. Initial path model (control group) with standardized regression weights. $\chi^2(8, N = 100) = 16.56$, $p < .05$; RMSEA = .11; NNFI = .82; CFI = .95. Standardized regression coefficients in bold are significant at $p < .05$. Values in parentheses are standard errors for the regression coefficient.

to the ceiling effect, which diminishes the magnitude of all effects. Fig. 5 depicts the final model obtained after removing non-significant paths and after modelling connections according to Eq. (A.5). The result is an increase in model fit in the revised model. The chi-square test was close to significance ($p = .06$), and the fit indices indicated a good model fit (NNFI = .90, RMSEA = .080, CFI = .94). Most notably, Age-of-Acquisition no longer influenced accuracy. Therefore, relevance remained the sole reliable predictor of accuracy.

3.2. Semantic parameters and total semantic relevance

An interesting issue is the relation between Age-of-Acquisition, frequency, familiarity and total relevance. As total relevance results from adding the relevance values of all the semantic features that are listed in defining the concept, it may be suggested that an increase in total relevance will be observed for concepts acquired early in life and used frequently. To investigate this possibility, the results of a preliminary correlation analysis are reported in Table 5.

Moderate-to-large relationships were found among all three selected concept parameters (FREQ, FAM, AA) and total relevance (RELT), ranging from $r = .202$ ($p < .01$) for FAM to $r = .400$ ($p < .05$) for FREQ. The first model tested represents the pattern of interrelations postulated above, in which FREQ and FAM were hypothesized to affect RELT independently, with AA on RELT mediated by FREQ and FAM. This model (Fig. 6) was saturated and therefore yielded a perfect fit, due to overparametrization. In order to avoid overfitting, we tested a revised model in which all non-

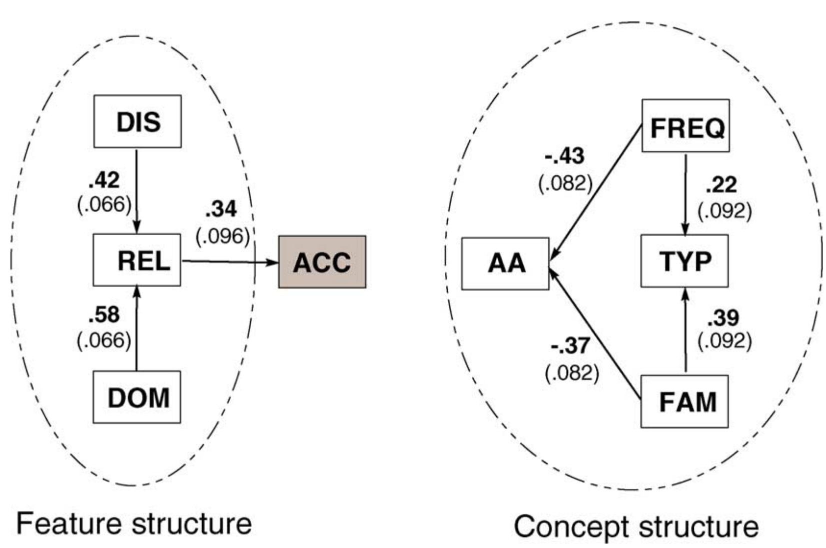

Fig. 5. Final path model (control group) with standardized regression weights. $\chi^2(15, N = 100) = 24.22$, $p = .06$ (n.s.); RMSEA = .080; NNFI = .90; CFI = .94. Standardized regression coefficients in bold are significant at $p < .05$. Note: non-significant paths of base model have been removed. Values in parentheses are standard errors for the regression coefficient.

Table 5
Pearson's correlations among parameters of concepts (frequency, familiarity, Age-of-Acquisition) and total semantic relevance

	RELT	FREQ	FAM	AA
RELT	–			
FREQ	.400**	–		
FAM	.202*	.260**	–	
AA	−.316**	−.529**	−.480**	–

RELT: total semantic relevance; FREQ: frequency; FAM: familiarity; AA: Age-of-Acquisition; n.s.: non-significant.

* $p < .05$.
** $p < .01$.

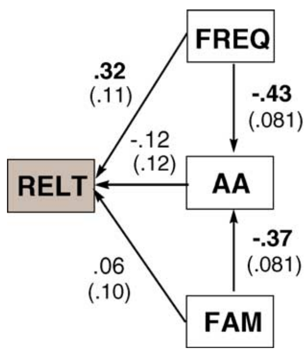

Fig. 6. Initial saturated path model with standardized regression weights. Standardized regression coefficients in bold are significant at $p < .05$. Values in parentheses are standard errors for the regression coefficient.

significant paths were removed (Fig. 7). The chi-square test for the revised model was non-significant ($p = .34$), and the fit indices indicated a very good model fit (NNFI = .99; RMSEA = .030; CFI = 1.00). The reconstructed correlations confirmed the strong effects of FREQ and FAM on AA (resp. −.43 for FREQ and −.37 for FAM). Moreover, FREQ was the only independent factor which positively affected RELT (.40), thus indicating that, the more frequently a subject is exposed to a concept, the more relevant the concept itself is.

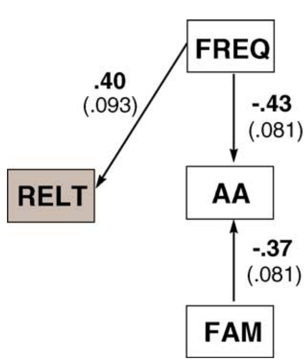

Fig. 7. Final path model with standardized regression weights. $\chi^2(2, N = 100) = 2.14$, $p = .34$ (n.s.); RMSEA = .03; NNFI = .99; CFI = 1.00. Standardized regression coefficients in bold are significant at $p < .05$. Note: non-significant paths of base model have been removed. Values in parentheses are standard errors for the regression coefficient.

4. Discussion

Concepts may be compared along a number of different dimensions. Some of the parameters convey compact information at the concept level (e.g., frequency, familiarity, Age-of-Acquisition, typicality), others are intended to measure semantic features (e.g., dominance, distinctiveness, semantic relevance).

In a naming-to-description task, administered to normal participants and to DATs with semantic impairment, we studied the predictive power of these parameters with respect to naming accuracy. First, qualitatively similar patterns were obtained for both DATs and normal controls. For this reason, unless otherwise stated, conclusions refer to both groups. Second, when normal controls and semantic memory patients were given a set of three semantic features and asked to retrieve the corresponding concept, their performance was best predicted by the relevance of the three presented semantic features. Most importantly, relevance turned out to be a better predictor than Age-of-Acquisition or frequency. Lastly, the relevance of the presented features was independent of frequency, familiarity, and Age-of-Acquisition.

4.1. Relation between DAT and control performances

Only minor differences were observed regarding the importance of the predictors between DAT and controls. The relevance of the three features and total relevance were the best predictors of accuracy in both groups. However, controls consistently had lower correlations between accuracy and the independent variables. A possible explanation calls into account the phenomenon known as range restriction. In this view, lower correlations in the control group are caused by the reduction in score range of the dependent variable; indeed, the control group showed lower variability in scores when compared with DATs.

Selection of material when constructing tests for assessing semantic memory patients is critical. Thus, similarity between the two groups has important empirical consequences. As the various parameters have similar effects in DATs and controls, norms collected on controls may reasonably be used with semantically impaired subjects (see Garrard et al., 2001; Rogers et al., 2004).

DATs with a general cognitive level similar to that of our patients, as measured by the MMSE, typically show impairment on semantic tasks (Hodges, Patterson, Graham, & Dawson, 1996). Degradation of conceptual knowledge follows the severity of the disease (Garrard et al., 2001) and it is the contention of Gainotti, Silveri, Daniele, and Giustolisi (1995) that DAT causes widespread damage to the temporal lobes and consequently impairment of semantic knowledge.

In our view, semantic degradation may be modelled assuming that damage reduces the connection strength between semantic features and concepts (this is a widely accepted assumption; e.g., see McLeod, Shallice, & Plaut, 2000). As the weight of connections between semantic features and concepts may be a way of conceptualizing relevance, the more relevant a feature for a concept, the more probable that that concept will be misnamed when the feature is damaged. Hence, the behavioral consequence of damage is expected to be proportional to the relevance of the lost/damaged feature. Nevertheless, we are not committed to any specific hypothesis about how brain damage may be mimicked by a neural network. Alternative hypotheses on this issue are possible. For example, if we consider a feed-forward neural network (e.g., Small, Hart, Nguyen, & Gordon, 1995), then features of similar relevance may be captured by the same hidden units. This suggests that focal damage affects specific hidden units and has disproportionate effects on individual categories. In sum, given random damage, the likelihood of correctly retrieving a concept will be reduced proportionally to the magnitude of the damage while the same qualitative pattern among controls and DATs will be reproduced. And this is what we observed.

4.2. Relations between parameters of the concept

Although Age-of-Acquisition and frequency reflect how often concepts are encountered, only Age-of-Acquisition predicted accuracy significantly in both groups. For frequency, the results were less coherent; frequency was significantly correlated with accuracy for DATs but not for controls.

As regards the mutual relations among frequency, familiarity, Age-of-Acquisition, and typicality, there has been some debate on whether Age-of-Acquisition influences behavior independently of frequency (e.g., Turner, Valentine, & Ellis, 1998) or merely embodies cumulative frequency (Lewis, Gerhand, & Ellis, 2001), because high-frequency words are likely to be acquired earlier than low-frequency ones. The cluster observed here, which included frequency, Age-of-Acquisition, and typicality, was observed several times. This indicates a trend common to both groups; normal controls and DAT patients tend to name concepts more accurately when they were acquired earlier in life and used extensively.

4.3. Relations between parameters of semantic features

The semantic relevance of the concept description predicts response accuracy in name retrieval better than distinctiveness and dominance of the same description. A soundly based explanation of the relation among these variables is required in order to avoid making a "correlation is causation" error in the interpretation.

According to Sartori and Lombardi (2004), the relevance of a semantic feature is determined by two separate components. The first, *local component* (Eq. (A.3)), measures the importance of the feature for the concept. The second, *global component* (Eq. (A.4)), measures how much the same feature contributes to the meaning of all the other concepts. Relevance integrates the effects of both local and global importance. It gains higher value when the semantic feature is used frequently in defining the concept, and few times in defining other concepts. As shown in Section A.2.1, the first part of the formula for computing semantic relevance (the local component) may be read as a function of dominance, whereas the second (the global component) may be read as a function of distinctiveness. Dominance is a measure of the frequency with which a given semantic feature is listed in defining the concept. Instead, distinctiveness is high when the semantic feature is used in defining few concepts. Distinctiveness is not linear, and this means that differences in features occurring in only a few concepts should be weighted more than those occurring in many concepts. For this reason, a logarithm is used to represent this non-linearity in the relevance formula. In this view, dominance and distinctiveness are integrated into semantic relevance, and this may explain the high correlation between the three variables that we observed.

The unexpected result was that, individually, dominance and distinctiveness were not significantly correlated with naming accuracy. More precisely dominance was only weakly correlated for patients: $r = .23$, $p < .05$, but not for controls. Instead, when combined into relevance, they became highly correlated with naming accuracy. Their relative contribution to relevance was highlighted: dominance at path analysis was more important than distinctiveness. As dominance and distinctiveness alone do not contribute to prediction of accuracy, their correlation with this parameter should be considered indirect. The fact that dominance is more predictive than distinctiveness indicates that the local component plays a more important role than the global component in name retrieval.

The number of intercorrelations among semantic features used in concept definition is believed to be an important aspect of name retrieval (Garrard et al., 2001; McRae, de Sa, & Seidenberg, 1993). One issue which arises concerns the possible relation between intercorrelation and semantic relevance. Interestingly, the Automatic Information Retrieval literature shows that, in order to increase accuracy in retrieval of documents, correlated terms should be added to queries (Van Rijsbergen, 1979). The semantic model we propose here is a modified version of the Vector Space Model within the information retrieval approach (Robertson & Sparck Jones, 1977; Van Rijsbergen, 1979) in which, in particular, concepts stand for documents and features stand for terms. Therefore, as regards intercorrelation, *mutatis mutandis*, adding a highly correlated feature to the concept description is expected to increase the likelihood of a correct response. A corollary is the following: if a group of semantic features yields correct name retrieval, then those semantic features will tend to be correlated to each other. We found corroborative evidence for this theoretical claim by analyzing our database post hoc. As expected, intercorrelation among the three presented features increased with relevance (average intercorrelations: first quartile of relevance = .45; second quartile = .61; third quartile = .75; fourth quartile = .81). Therefore, when semantic relevance is high, features tend to be correlated to each other, and when it is low they tend to be less correlated.

4.4. Relation between semantic relevance and parameters of the concept

Below we give reasons for the high correlation between total relevance and both Age-of-Acquisition and frequency. We also give reasons for the claim that total relevance is refined by long experience with the concept, as measured by Age-of-Acquisition and frequency.

Other things being equal, total relevance is high when there are many semantic features of high relevance. Features increase their relevance when fine-grained distinctions are required to discern among similar concepts. If several similar concepts exist in the mental lexicon, the target concept must be described more precisely in order to be correctly identified. Let us suppose, for example, that we do not know of the existence of an animal called OKAPI (a rare mammal living in central Africa, with a neck like a giraffe's, and a striped back like a zebra's), and suppose that we are asked to define the concept ZEBRA. In this case, the two features ⟨is a mammal⟩ and ⟨has black and white stripes⟩ may be sufficient to identify the concept. However, they do not allow us to distinguish ZEBRA from OKAPI if OKAPI forms part of our mental lexicon. This is because the entry of OKAPI into our lexicon determines a reorganization of the relevance values of the two semantic features which have their original relevance values decreased.[8] Whether this update is accomplished as soon as the new semantic information is encountered, or rather when this new information is actually used, is an open question. We assume here, but we have no empirical data in support, that the update of the relevance weights is carried out as soon as the new semantic information enters the lexicon.

Progression from broad to finer-grained distinctions of semantic representation have been investigated by Keil (1979), who showed that statements that children accept as true for a concept progressively change through experience. For example, at kindergarten the feature ⟨can feel sorry⟩ applies to both MAN and PIG, whereas at sixth grade it only applies to MAN.

Evidence that relevance weights can be shaped through experience comes from Macario (1991), who showed that children young as 3 years seem to know that the color of an object is more important than its shape if it is a kind of food, and slightly less important if it is a kind of toy (see also Jones, Smith, & Landau, 1991, for similar findings). According to Mervis (1984), conceptual representation in childhood includes a smaller number of semantic features relative to adults, and Steyvers and Tenenbaum (in press) showed that early acquired concepts have a more complex semantic network than late acquired concepts. In their model, new concepts are built on older ones and, consequently, the order in which concepts are learned is important. Early acquired concepts possess more connections than later acquired ones. In sum, longer experience leads to a level of expertise which may change the relevance values of semantic features. In this view, the history of concept learning, as measured by Age-of-Acquisition and frequency, may be considered a form of compact measure of the amount of experience concerning the concept.

Not only does the amount of experience increase with Age-of-Acquisition, but also the number of contexts in which the concept appears increases. Early Age-of-Acquisition and high frequency also increase the likelihood of encountering the target concept in many contexts. The longer subjects are exposed to experience with a concept (as grossly measured by frequency and Age-of-Acquisition), the more fine-grained distinctions they can make between this same concept and similar ones. Hence, high frequency and early Age-of-Acquisition may be considered representative of the high variety of contexts in which a certain concept appears. Indeed, Steyvers and Tenenbaum (in press) have shown that the number of contexts in which a concept appears is highly correlated with frequency ($r = .98$, $p < .05$). As a result, total relevance is expected to be high when frequency is high and when Age-of-Acquisition is low. Age-of-Acquisition captures the level of experience about the concept, and early acquired concepts may develop richer semantic descriptions than later acquired ones, and this, in turn, has an influence on total semantic relevance. This claim may find corroborative evidence in the correlation between total semantic relevance and Age-of-Acquisition. A figure of $r = -.417$ ($p < .05$) indicates that, when the concept is acquired earlier, its total semantic relevance increases.

In summary, semantic features are likely to be assigned adequate relevance weights in proportion to the variety of experiences that subjects have had with it which, in turn, may be measured by Age-of-Acquisition and frequency. In this view, Age-of-Acquisition effects do not imply that late learned words are encoded less effectively than early learned ones, but simply that they have lower relevance semantic features, as they have reduced occasions on which concepts are contrasted with similar items. We must remember here that the identification of distinguishing features among pairs of concepts drives the semantic relevance of a given semantic feature to higher levels. Hence, developing fine-grained relevance values guarantees fine-grained distinctions among concepts.

All in all, in this work we attempted to analyze and discuss, within a soundly based framework, the underlying structure of the relationships among parameters of concepts and parameters of semantic features. In particular, we showed that the accuracy in a naming-to-description task is best predicted by the semantic relevance of the concept description. This

[8] As the distinctiveness of a semantic feature is based on the ratio between the number of concepts in which the feature appears and the total number of concepts, then increasing the numerator and denominator by 1 decreases the distinctiveness value and consequently decreases relevance. For further details, see Section A.2.3

result was observed both in healthy controls and in DAT patients. Other parameters of semantic features (dominance and distinctiveness; the components of relevance) were less correlated with accuracy if taken singly. Among concept parameters, Age-of-Acquisition was the second best predictor. Based on these results we have also put forward an explanation that links together most of the variables that affect semantic representations and semantic impairments. As a final remark we must note that our conclusions are limited to the naming-to-description task and to the specific database used for estimating parameters of semantic features. Although we have shown, in our database, that semantic relevance is a reliable measure (see Section A.2.4), we also think that the important issue of what can actually be considered a representative sample of the mental lexicon has not been yet addressed by us or other investigators. Therefore, specific characteristics of the database of concepts might influence the estimation of parameters of semantic features such as dominance, distinctiveness, relevance and intercorrelation thus possibly biasing results.

Appendix A

We briefly describe here the general model for the semantic relevance of concepts, as described in Sartori and Lombardi (2004). We then introduce a particular instance of this model which clarifies the interrelationships between semantic relevance and dominance (resp. distinctiveness).

A.1. General model

This model is an adapted version of the Vector Space Model within the information retrieval approach (Robertson & Sparck Jones, 1977; Van Rijsbergen, 1979).

Relevance analysis transforms an I (concepts) $\times J$ (semantic features) intensity data matrix $\mathbf{X}$ into an $I \times J$ relevance model matrix $\mathbf{K}$, which represents the semantic relevance model for the domain under investigation. Entry $x_{ij} \in \Re^+ \cup \{0\}$ of $\mathbf{X}$ denotes a degree of positive association between Feature j and Concept i, whereas entry $k_{ij} \in \Re^+ \cup \{0\}$ of $\mathbf{K}$ denotes the relevance of Feature j for Concept i.

The fundamental assumption of our model is that relevance matrix $\mathbf{K}$ may be decomposed into an $I \times J$ matrix $\mathbf{L}$ and a $J \times J$ diagonal matrix[9] $\mathbf{G}$, by means of the matrix product:

$$\mathbf{K} = \mathbf{LG} \tag{A.1}$$

In the above equation, $\mathbf{L}$ represents an $I \times J$ matrix of weights with entry $l_{ij} \in \Re^+ \cup \{0\}$ of $\mathbf{L}$, denoting the *local importance* of Feature j for Concept i; hence, $\mathbf{L}$ is called the *local importance matrix*. Main diagonal $diag(\mathbf{G})$ of $\mathbf{G}$ represents a vector of J weights with entry $g_j \in \Re^+ \cup \{0\}$ of $diag(\mathbf{G})$, denoting the *overall importance* of Feature j for all I Concepts; hence, $\mathbf{G}$ is called the *global importance matrix*.

$\mathbf{L}$ and $diag(\mathbf{G})$ may be derived by means of two weighting mappings $\langle \phi, \psi \rangle$:

$$\mathbf{L} = \phi(\mathbf{X}), \qquad diag(\mathbf{G}) = \psi(\mathbf{X}) \tag{A.2}$$

which act as a linking structure between intensity matrix $\mathbf{X}$ and relevance matrix $\mathbf{K}$. Several weighting schemes may be derived from information retrieval models (Dumais, 1991) and adopted, after appropriate modifications, within a relevance analysis approach. In this paper, we refer to a simple weighting scheme called FF × ICF (*Feature Frequency × Inverse Concept Frequency*), adapted from Salton's well-known TF × IDF (*Term Frequency × Inverse Document Frequency*) measure (Salton, 1989).

A.2. Semantic relevance as FF × ICF instance

The whole procedure may be split into three consecutive steps. First, cued verbal descriptions of 254 concepts belonging to the corpus of Dell'Acqua et al. (2000) were collected for five Italian-speaking subjects. 2619 features were extracted from verbal descriptions. Second, 254 (concepts) × 2619 (semantic features) intensity matrix $\mathbf{X}$ was computed by setting entry x_{ij} of $\mathbf{X}$ as equal to the frequency of occurrence of Feature j in Concept i over all subjects' descriptions (for details, see Sartori & Lombardi, 2004). Lastly under the FF × ICF assumption, we set:

$$l_{ij} = \phi(x_{ij}) = x_{ij} \tag{A.3}$$

$$g_j = \psi(\mathbf{x}_{.j}) = \log\left(\frac{I}{I_j}\right) \tag{A.4}$$

($\forall i = 1, \ldots, I = 254; \forall j = 1, \ldots, J = 2619$) with $\mathbf{x}_{.j}$ and I_j, respectively, denoting the jth-column of $\mathbf{X}$ and the number of concepts in which Feature j occurs. Note that, under the FF × ICF assumption, mapping ϕ reduces to the identity function.

By applying (1), entry k_{ij} ($\forall i = 1, \ldots, I; \forall j = 1, \ldots, J$) of relevance matrix $\mathbf{K}$ takes the form:[10]

$$k_{ij} = l_{ij} g_{jj} = x_{ij} \log\left(\frac{I}{I_j}\right) \tag{A.5}$$

In words, Eq. (A.5) states that a feature which captures the core meaning of a concept will have both high local importance and high global importance.

A.2.1. Relationship of semantic relevance with dominance and distinctiveness

Dominance is a measure of how frequently a semantic feature is used in defining a concept (Ashcraft, 1978; Garrard et al., 2001). It is defined as the number of times subjects

[9] Matrix $\mathbf{A}$ is a diagonal matrix if (i) $\mathbf{A}$ is a square matrix, and (ii) $a_{ij} = 0$ whenever $i \neq j$.

[10] As $\mathbf{G}$ is a diagonal matrix, Eq. (A.5) may be considered in place of the standard product $k_{ij} = \sum_{h=1}^{J} l_{ih} g_{hj}$.

mention a given feature in defining a concept, divided by the sum of the counts of all features occurring in the definition of that concept. The distinctiveness of a given feature is defined as the complement of *sharedness* which, in turn, is defined as the number of concepts in which the semantic feature appears, divided by the total number of concepts in the lexicon (or database) (Tyler et al., 2000). Using our notation, we can rewrite these parameters as:

$$\text{Dominance}_{ij} = \frac{x_{ij}}{\sum_{j'=1}^{J} x_{ij'}} \tag{A.6}$$

$$\text{Distinctiveness}_j = \left(1 - \frac{I_j}{I}\right) = \left(\frac{I - I_j}{I}\right), \tag{A.7}$$

with $I_j > 0$. Therefore, by simple algebra, we can rewrite the relevance parameter as:

$$\begin{aligned} k_{ij} &= \left(\sum_{j'=1}^{J} x_{ij'} \times \text{Dominance}_{ij}\right) \\ &\quad \times \log\left(1 + \frac{I}{I_j} \times \text{Distinctiveness}_j\right) \\ &= x_{ij} \log\left(1 + \frac{I}{I_j}\left(\frac{I - I_j}{I}\right)\right) = x_{ij} \log\left(\frac{I}{I_j}\right). \end{aligned}$$

Stated in words, global weight g_{jj} as defined in Eq. (A.4) is simply distinctiveness after its transformation by an appropriate function *f*(Distinctiveness) (see equation above), then damping by the base 2 logarithm; l_{ij} is the non-normalized dominance of Feature *j* over Concept *i*.

A.2.2. Relevance ranking of concepts

In the general case, a ranking of concepts is done by ordering their summed relevance values (*total relevance*). More precisely, a Concept *i* is more relevant than a Concept i^* if and only if the sum of the relevance values of the former is greater than the sum of the relevance values of the latter. In symbols:

$$C_i \succ_{Rel} C_{i^*} \iff \sum_{j=1}^{J} k_{ij} > \sum_{j=1}^{J} k_{i^*j}$$

In the special case of a naming-to-three descriptions task, the dominance relation ($\succ_{Rel}$) is reduced to:

$$C_i \succ_{Rel} C_{i^*} \iff \sum_{t=1}^{3} k_{ij(t)} > \sum_{t=1}^{3} k_{i^*j^*(t)}$$

where $\{j(1), j(2), j(3)\}$ (resp. $\{j^*(1), j^*(2), j^*(3)\}$) denotes the three features describing Concept *i* (resp. Concept i^*).

A.2.3. Effect on relevance of a larger ($I + n_j$) lexicon

We now prove how increasing the number of concepts in the mental lexicon for which feature *j* occurs, diminishes the relevance of the concepts in a new upgraded lexicon. Let us suppose that n_j new concepts associated with Feature *j* enter the lexicon. The recalculated distinctiveness of Feature j ($\forall j = 1, \ldots, J$) then becomes:

$$1 - \frac{I_j + n_j}{I + n_j} = \frac{I - I_j}{I + n_j} \leq \frac{I - I_j}{I} \tag{A.8}$$

which is not greater than the original distinctiveness. Therefore, by simple algebra and by Eqs. (A.5) and (A.8), the relevance values corresponding to the original *I* concepts, as recoded in upgraded vector $\mathbf{k}_j^\star$ turn out to be not greater than the corresponding values as recorded in original vector $\mathbf{k}_j$:

$$\begin{aligned} k_{ij}^\star &= \left[x_{ij} \log\left(1 + \frac{I + n_j}{I_j + n_j}\left(\frac{I - I_j}{I + n_j}\right)\right)\right] \\ &\leq \left[x_{ij} \log\left(1 + \frac{I}{I_j}\left(\frac{I - I_j}{I}\right)\right)\right] = k_{ij} \end{aligned}$$

for all $i = 1, \ldots, I$.

A.2.4. Reliability of relevance in relation to the dimension of the mental lexicon

Given that relevance depends on *I*, there is the possibility that it might change significantly with varying levels of *I*. In order to evaluate this possibility, we recalculated relevance values on different subsets of our original 254 concepts database using a bootstrapping procedure as follows:

(1) For each of the 100 concepts used in our naming-to-description task, the sum of relevance values of the three presented features was computed using the entire database of the 254 concepts. These values were therefore collected into a vector **k** of 100 relevances (one for each target concept).
(2) (a) A sample without replacements of the 254 concepts was randomly drawn (we used three different sampling sizes: 50, 100 and 150).
 (b) For each of the 100 target concepts the sum of the relevance of the three associated semantic features was recalculated, this time using only 50 (+1 = target concept) (resp. $100 + 1$, and $150 + 1$) concepts.
 (c) Steps (a) and (b) were iterated 100 times. This yielded a distribution of relevance values for each of the 100 target concept. Therefore the mean of the distribution of the recalculated relevance values was computed for each of the 100 concepts. Finally we collected these 100 means into a new vector $\mathbf{k}^\star$.
(3) In order to check whether $\mathbf{k}^\star$ might predict the original relevance values as codified in **k** (see Step 1), we ran three regression analyses (one for each sample size) using $\mathbf{k}^\star$ as independent variable and **k** as dependent variable.

The results of these analyses are reported in Figs. A.1–A.3.

These results indicate that even if, in theory, the number of concepts in the mental lexicon influences the absolute values

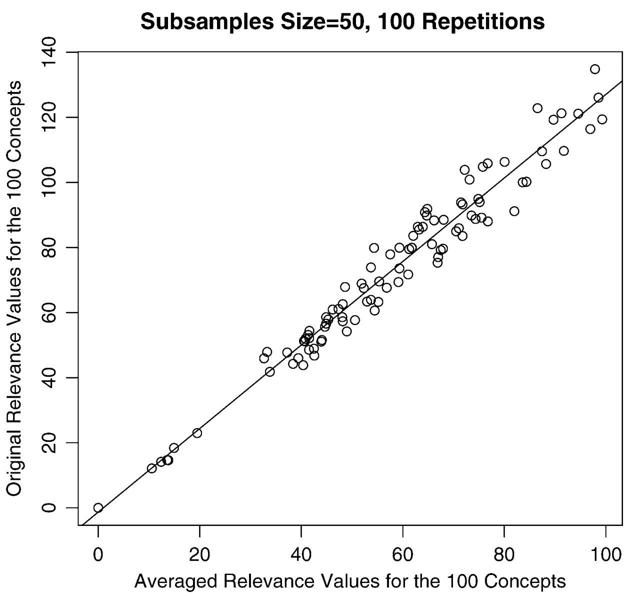

Fig. A.1. Relevance values of the 100 concepts computed on the whole 254 concepts database plotted as a function of the averaged relevance values of the same 100 concepts computed on 100 subsamples of size = 50(+1). Regressing these data ($R^2 = .962$, $p < .001$) we found a slope of 1.284 (with standard error of 0.025) and an intercept value of −1.307 (with standard error of 1.566). The slope (1.284) was significantly different from 0 ($t = 50.421$, $p < .001$); the intercept (−1.307) was not significantly different from 0 ($t = -.835$, $p = .406$).

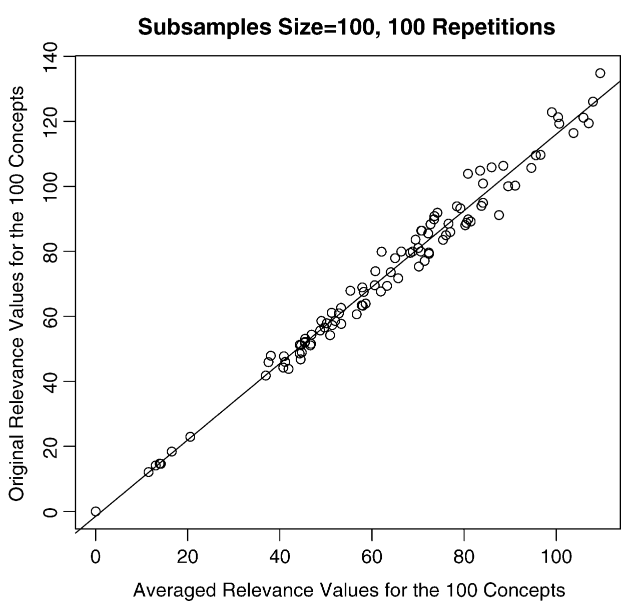

Fig. A.2. Relevance values of the 100 concepts computed on the whole 254 concepts database plotted as a function of the averaged relevance values of the same 100 concepts computed on 100 subsamples of size = 100(+1). Regressing these data ($R^2 = .982$, $p < .001$) we found a slope of 1.175 (with standard error of .015) and an intercept value of −1.501 (with standard error of 1.064). The slope (1.175) was significantly different from 0 ($t = 74.443$, $p < .001$); the intercept (−1.501) was not significantly different from 0 ($t = -1.418$, $p = .159$).

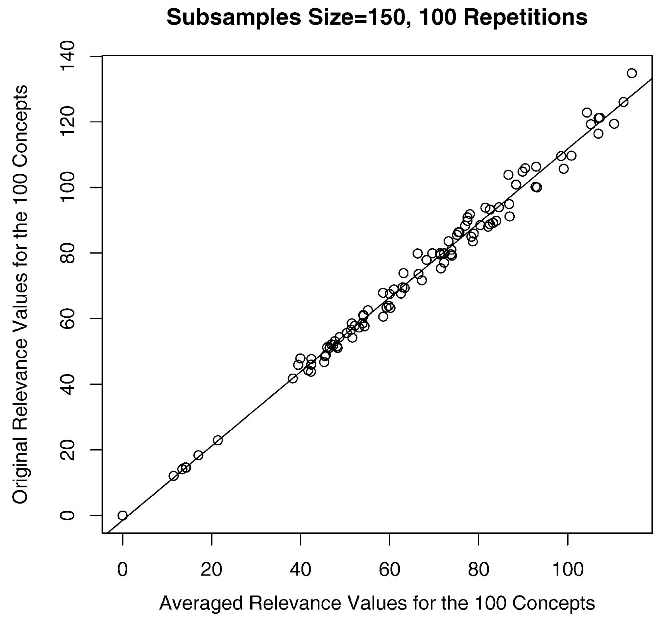

Fig. A.3. Relevance values of the 100 concepts computed on the whole 254 concepts database plotted as a function of the averaged relevance values of the same 100 concepts computed on 100 subsamples of size = 150(+1). Regressing these data ($R^2 = .990$, $p < .001$) we found a slope of 1.132 (with standard error of .011) and an intercept value of −1.437 (with standard error of .783). The slope (1.132) was significantly different from 0 ($t = 101.133$, $p < .001$); the intercept (−1.437) was not significantly different from 0 ($t = -1.834$, $p = .069$).

of relevance (see Section A.2.3), in practice ranking of concepts, in terms of relevance, remains substantially unchanged. And this shows the robustness of relevance estimations.

References

Ashcraft, M. H. (1978). Property norms for typical and atypical items from 17 categories: A description and discussion. *Memory and Cognition*, *6*, 227–232.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of Age-of-Acquisition, frequency, and name agreement. *Quarterly Journal of Experimental Psychology*, *50A*, 560–585.

Capitani, E., Laiacona, M., Mahon, B., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*, 213–261.

Carroll, J. B., & White, M. N. (1973). Word frequency and age of acquisition as determiners of picture naming latency. *Quarterly Journal of Experimental Psychology*, *25A*, 85–95.

Chertkow, B., & Bub, D. (1990). Semantic memory loss in dementia of Alzheimer's type. What do various measures measure? *Brain*, *113*, 397–417.

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.

De Moor, W., Ghyselinck, M., & Brysbaert, M. (2001). The effects of frequency of occurrence and age-of-acquisition in word processing. In F. Columbus (Ed.), *Advances in psychology research* (Vol. 5, pp. 71–84). Huntington, NY: Nova Science.

Dell'Acqua, R., Lotto, L., & Job, R. (2000). Naming time and standardized norms for the Italian PD/DPSS set of 266 pictures: Direct comparisons with American, English, French, and Spanish published databases. *Behaviour Research Methods, Instruments and Computers*, *32*, 588–612.

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.

Dumais, S. T. (1991). Improving the retrieval of information from external sources. *Behaviour Research Methods, Instruments and Computers*, *23*, 229–236.

Folstein, M. F., Folstein, S. E., & Mc Hugh, P. R. (1975). Mini Mental State: A pratical method for grading the cognitive state of patients for clinician. *Journal of Psychiatric Research*, *12*, 189–198.

Funnell, E., & Sheridan, J. (1992). Categories of knowledge? Unfamiliar aspects of living and non-living things. *Cognitive Neuropsychology*, *9*, 135–153.

Gainotti, G., Silveri, M. C., Daniele, A., & Giustolisi, L. (1995). Neuroanatomical correlates of category-specific semantic disorders: A critical survey. *Memory*, *3*, 247–264.

Garrard, P., Lambon Ralph, M., Hodges, J. R., & Patterson, K. (2001). Prototipicality, distinctiveness, and intercorrelations: Analyses of the semantic attributes of Living and nonLiving concepts. *Cognitive Neuropsychology*, *18*, 125–174.

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely Age-of-Acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*, 267–283.

Gilhooly, K. J., & Gilhooly, M. L. (1980). The validity of age of-acquisition ratings. *British Journal of Psychology*, *71*, 105–110.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1944 words. *Behavior Research Methods and Instrumentation*, *12*, 395–427.

Hachinski, V. G., Iliff, L., Du Boulay, G. H., McAllister, V. L., Marshall, J., Ross Russell, R. W., et al. (1975). Cerebral blood flow in dementia. *Archive of Neurology*, *32*, 633–637.

Hodges, J. R., & Patterson, K. (1995). Is semantic memory consistently early in the course of Alzheimer´s disease? Neuroanatomical and diagnostics implications. *Neuropsychologia*, *33*, 441–459.

Hodges, J. R., Patterson, K., Graham, N., & Dawson, K. (1996). Naming and knowing in dementia of Alzheimer's type. *Brain and Language*, *54*, 302–325.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.

Jolicoeur, P., Gluck, M. A., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, *16*, 243–275.

Jones, S. S., Smith, L. B., & Landau, B. (1991). Object properties and knowledge in early lexical learning. *Child Development*, *62*, 499–516.

Jorm, A. F. (1991). The validity of word age-of-acquisition ratings: A longitudinal study of a child word knowledge. *British Journal of Developmental Psychology*, *9*, 437–444.

Jöreskog, K. J., & Sorböm, D. (1993). *PRELIS2 user's reference guide*. Chicago: Scientific Software International, Inc.

Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.

Kiran, S., & Thompson, C. K. (2003). Effect of typicality on online category verification of animate category exemplars in aphasia. *Brain and Language*, *85*, 441–450.

Kremin, H., Akhutina, T., Basso, A., Davidoff, J., De Wilde, M., Kitzing, P., et al. (2003). A cross-linguistic data bank for oral picture naming in Dutch, English, German, French, Italian, Russian, Spanish, and Swedish (PEDOI). *Brain and Cognition*, *53*, 243–246.

Lambon Ralph, M. A., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia—What matters? *Neuropsychologia*, *36*, 775–784.

Lewis, M. B., Gerhand, S., & Ellis, H. D. (2001). Re-evaluating Age-of-Acquisition effects: Are they simply cumulative-frequency effects? *Cognition*, *78*, 189–205.

Lombardi, L., Pastore, M., & Nucci, M. (2004). Evaluating uncertainty of model acceptability in empirical applications: A replacement approach. In K. Van Monfort, J. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models: Theory, applications* (pp. 69–82). Amsterdam: Kluwer Academic Publishers.

Lyons, A. W., Teer, P., & Rubenstein, H. (1978). Age-at-acquisition and word recognition. *Journal of Psycholinguistic Research*, *7*, 179–187.

Macario, J. F. (1991). Young children's use of color and classification: Foods and canonically colored objects. *Cognitive Development*, *6*, 17–46.

Mandler, G. (1980). Recognizing: The judgement of previous occurrence. *Psychological Review*, *87*, 252–271.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadan, E. M. (1984). Clinical diagnosis of Alzheimer disease: Report of the NINCDS-ADRDA work group under the auspices of the Department of Health and Human Services Task Force on Alzheiemer disease. *Neurology*, *34*, 939–944.

McLeod, P., Shallice, T., & Plaut, D. C. (2000). Attractor dynamics in word recognition: Converging evidence from errors by normal subjects, dyslexic patients and a connectionist model. *Cognition*, *74*, 91–114.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1993). Modeling property intercorrelations in conceptual memory. In *Proceedings of the 15th Annual Meeting of the Cognitive Science Society* (pp. 729–734). Hillsdale, New York: Erlbaum.

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.

Mervis, C. (1984). Early lexical development: The combination of mother and child—Origins of cognitive skills. In C. Sophian (Ed.), *The origin of cognitive skills*. Hillsdale, New York: Erlbaum.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology*, *50A*, 528–559.

Moss, H. E., Tyler, L. K., Durrant-Peatfield, M., & Bunn, E. M. (1998). "Two eyes of a see -through": Impaired and intact semantic knowledge in a case of selective deficit for Living things. *Neurocase*, *4*, 291–310.

Pind, J., Jonsdottir, H., Tryggvadottir, H. B., & Jonsson, F. (1980). Icelandic norms for the Snodgrass and Vanderwart pictures: Name and image agreement, familiarity, and age of acquisition. *Scandinavian Journal of Psychology*, *41*, 41–88.

Robertson, S., & Sparck Jones, S. (1977). Relevance weighting of search terms. *Journal of the American Society for Information Science*, *27*, 129–146.

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., et al. (2004). The structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review*, *111*, 205–235.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, *104*, 192–233.

Rosch, E., & Mervis, C. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, *8*, 382–439.

Salton, G. (1989). *Automatic text processing: The transformation, analysis and retrieval of information by computer*. Reading, MA: Addison-Wesley.

Sartori, G., Colombo, L., Vallar, G., Rusconi, M. L., & Pinarello, A. (1995). T.I.B. Test di intelligenza breve per la valutazione del quoziente intellettivo attuale. *Giornale dell'Ordine degli Psicologi*, *4*, 1–24.

Sartori, G., Job, R., & Zago, S. (2002). A case of domain-specific semantic deficit. In E. M. E. Forde & G. W. Humphreys (Eds.), *Category specificity in brain and mind* (pp. 25–49). London: Psychology Press.

Sartori, G., & Lombardi, L. (2004). Semantic relevance and semantic disorders. *Journal of Cognitive Neuroscience*, *16*, 439–452.

Silveri, M. C., & Gainotti, G. (1988). Interaction between vision and language in category-specific semantic impairment. *Cognitive Neuropsychology*, *5*, 677–709.

Small, S. L., Hart, J., Nguyen, T., & Gordon, B. (1995). Distributed representations of semantic knowledge in the brain. *Brain*, *118*, 441–453.

Smith, E., Shoben, E., & Rips, L. (1974). Structure and process in semantic memory: A featural model for semantic decision. *Psychological Review*, *81*, 214–241.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments and Computers*, *28*, 516–536.

Spinnler, H., & Tognoni, G. (1987). Standardizzazione e taratura italiana di test neuropsicologici. *The Italian Journal of Neurological Sciences*, *6*, 5–120.

Steyvers, M., & Tenenbaum, J. (in press). The large scale structure of concept networks: Statistical analyses of a model of semantic growth. *Cognitive Science*.

Turner, J. E., Valentine, T., & Ellis, A. W. (1998). Contrasting effects of Age-of-Acquisition and word frequency on auditory and visual lexical decision. *Memory and Cognition*, *26*, 1282–1291.

Tyler, L. K., Moss, H. E., Durrant-Peatfield, M., & Levy, J. (2000). Conceptual structure and the structure of the concepts: A distributed account of category specific deficits. *Brain and Language*, *75*, 195–231.

Van Rijsbergen, C. J. (1979). *Information retrieval*. London: Butterworth.

Vigliocco, G., Vinson, D. P., Lewis, W., & Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology*, *48*, 422–488.

Walley, A. C., & Metsala, J. L. (1992). Young children's Age-of-Acquisition estimates for spoken words. *Memory and Cognition*, *20*, 171–182.

Warrington, E., & Shallice, T. (1984). Category-specific semantic impairments. *Brain*, *107*, 829–854.

Zevin, J. D., & Seidenberg, M. S. (2004). Age-of-Acquisition effects in reading aloud: Tests of cumulative frequency and frequency trajectory. *Memory and Cognition*, *32*, 31–38.