RESEARCH REPORT

The semantic organization of the animal category: evidence from semantic verbal fluency and network theory

Joaquín Goñi · Gonzalo Arrondo · Jorge Sepulcre · Iñigo Martincorena ·
Nieves Vélez de Mendizábal · Bernat Corominas-Murtra · Bartolomé Bejarano ·
Sergio Ardanza-Trevijano · Herminia Peraita · Dennis P. Wall · Pablo Villoslada

Abstract Semantic memory is the subsystem of human memory that stores knowledge of concepts or meanings, as opposed to life-specific experiences. How humans organize semantic information remains poorly understood. In an effort to better understand this issue, we conducted a verbal fluency experiment on 200 participants with the aim of inferring and representing the conceptual storage structure of the natural category of animals as a network. This was done by formulating a statistical framework for co-occurring concepts that aims to infer significant concept–concept associations and represent them as a graph. The resulting network was analyzed and enriched by means of a missing links recovery criterion based on modularity. Both network models were compared to a thresholded co-occurrence approach. They were evaluated using a random subset of verbal fluency tests and comparing the network outcomes (linked pairs are clustering transitions and disconnected pairs are switching transitions) to the outcomes of two expert human raters. Results show that the network models proposed in this study overcome a thresholded co-occurrence approach, and their outcomes are in high agreement with human evaluations. Finally, the interplay between conceptual structure and retrieval mechanisms is discussed.

Keywords Verbal fluency · Switching-clustering · Semantic memory · Network theory

J. Goñi · G. Arrondo · J. Sepulcre · I. Martincorena ·
N. Vélez de Mendizábal · B. Bejarano
Department of Neurosciences. Center for Applied Medical Research, University of Navarra, Pamplona, Spain

J. Goñi · S. Ardanza-Trevijano
Department of Physics and Applied Mathematics, University of Navarra, Pamplona, Spain

B. Corominas-Murtra
ICREA-Complex Systems Lab, Universitat Pompeu Fabra-Parc de Recerca Biomèdica de Barcelona, Barcelona, Spain

H. Peraita
Department of Psychology, National University of Distance Education (UNED), Madrid, Spain

D. P. Wall
The Center for Biomedical Informatics, Harvard Medical School, Boston, MA, USA

P. Villoslada (✉)
Department of Neurosciences, Institut d'investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Hospital Clínic, Barcelona, Spain
e-mail: pvilloslada@clinic.ub.es

Introduction

Semantic memory is the subsystem of human memory that stores conceptual and factual knowledge. Contrary to episodic memory, which stores life experiences, semantic memory is not linked to any particular time or place. In a more restricted definition, it is responsible for the storage of semantic categories and naming of natural and artificial concepts (Budson and Price 2005; Patterson et al. 2007). How these categories are organized, and more specifically, which words or concepts are close to which others, has kept the attention of a number of studies, most of them based on verbal fluency data.

Verbal fluency tasks with either semantic or phonetic cues are widely used in neuropsychological studies (Galeote and Peraita 1999; Ardila and Ostrosky-Solís

2006). In semantic fluency tasks, participants have to produce words from a category such as animals or fruits in a given time (usually 60 or 90 s). Although other semantic categories have been used in this kind of tests, the animal category has the advantage of universality: it is a clear enough test across languages and cultures with only minor differences across different countries, educational systems and generation belonging (Ardila and Ostrosky-Solís 2006). Being the number of different words named the most common clinical measure (Lezak 1995), it has also been observed that words tend to appear in semantically grouped clusters (Bousfield and Sedgewick 1944; Gruenewald and Lockhead 1980; Raskin et al. 1992; Wixted and Rohrer 1994). This behavioral observation led Troyer et al. (1997) to propose a two component model of the semantic fluency task. The first component, *clustering*, implies the production of related words until a particular category is exhausted. The second component, *switching*, implies moving to a different semantic cluster. It has been argued that switching implies the flexibility to initiate a new category search and is related to frontal executive functioning while clustering depends on the brain's temporal lobe and is characterized by local explorations of semantic memory (Troyer et al. 1997, 1998a, b; Tröster et al. 1998).

This paper addresses the problem of semantic organization from the viewpoint of modern network theory. Network theory has arisen as an influential field of research (Albert and Barabási 2002) in the context of complex networks, i.e., those networks or graphs containing non-trivial topological features. This framework has broadened the understanding of a wide variety of systems, including social (Wasserman and Faust 1994; Rosvall and Bergstrom 2008), biological (Jeong et al. 2001; Voy et al. 2006) and neural networks (Sporns et al. 2004; Eguíluz et al. 2005). The case of language (Ferrer i Cancho and Solé 2001; Solé et al. 2010) and in particular of semantics (Sigman and Cecchi 2002; Motter et al. 2002; Steyvers and Tenenbaum 2005) has not been an exception—see Borge-Holthoefer and Arenas (2010b) for a detailed review. Regarding verbal fluency, a recent study has applied a network approach based on co-occurrences to verbal fluency data in order to assess behavioral differences between healthy subjects, patients with mild cognitive impairment and patients with Alzheimer's disease (Lerner et al. 2009).

Beyond the general statistical analysis provided by Sigman and Cecchi (2002), Motter et al. (2002), Steyvers and Tenenbaum (2005), a variety of cognitive models have proposed that semantic knowledge can be represented as a complex network, where nodes represent words or concepts and links connecting them correspond to conceptual (semantic) relationships. In earlier studies to explain semantic memory, a tree-like hierarchical structure was proposed (Collins and Quillian 1969, 1970), in which specific concepts are embedded in more general ones and at the same time nest-specific items, storing at each level of the hierarchy the shared features of its concepts. Nevertheless, this classification seems to be too strict, since cognitive categories are not clearly bounded (Rosch et al. 1976) and occasionally elements do not inherit the characteristics of their supra-ordinates (Sloman 1998). These theoretical limitations brought about unstructured network models where hierarchy is lost and nodes are linked as many times as relations found between their underlying concepts. Hence, any single concept can be defined in terms of its links to other concepts. These models are known as *spreading activation models* since information is processed through activation, beginning at a given point of the network and spreading to adjacent nodes following a decreasing energy gradient (Quillian 1967; Collins and Loftus 1975; Anderson 1976; Hayes-Roth 1977; Anderson and Pirolli 1984).

The models described above aim to represent the deep conceptual structure of semantic memory through a system of abstract propositions that characterize each concept by relating it to others. The high level of abstraction of these models forced authors to either code their representations manually (Quillian 1967; Collins and Quillian 1969) or leave them at a theoretical level (Anderson and Pirolli 1984; Hayes-Roth 1977). *Semantic association models*, focused on natural language use, emerged as an alternative to these theoretically driven representations. They consist of identifying clusters of concepts in a multidimensional space and yield less-specific relationships than preceding approaches—for a review see Griffiths et al. (Griffiths et al. 2007). This permits the creation of models based on data from semantic decision tasks (Rips et al. 1973; Henley 1969), verbal fluency tests (Henley 1969; Crowe and Prescott 2003; Schwartz et al. 2003), association norms (Henley 1969), or large linguistic corpora (Lund and Burgess 1996), in a non-supervised manner. In particular, semantic distance algorithms, which assume that nearer words within the tests are conceptually closer, have been applied to fluency tasks of both healthy controls (Henley 1969; Crowe and Prescott 2003; Schwartz et al. 2003) and neurological patients (Chan et al. 1993; Aloia et al. 1996; Schwartz and Baldo 2001; Prescott et al. 2006) in order to study the semantic structure of memory.

The aim of this work is to obtain a reliable conceptual network (CN) that represents the semantic organization of the animal category. This has been done by recruiting a large dataset of verbal fluency as the input source and by introducing a novel statistical framework for co-occurring concepts that aims to infer significant concept–concept associations. The resulting network is analyzed and enriched (ECN) by means of a missing links recovery criterion based on modularity. Finally, the accuracy of both CN and

ECN is evaluated. This is done using a subset of verbal fluency tests and comparing the network outcomes (linked pairs are clustering transitions and disconnected pairs are switching transitions) to the evaluations of two expert human raters. Results show that CN and ECN models used as classifiers are remarkably close to human evaluation, overcoming a thresholded co-occurrence strategy.

Our approach shares with the *spreading activation models* the representation of semantic memory as a network and with the *semantic association models* its unsupervised inference (no taxonomic or any other a priori knowledge is applied). In order to infer the semantic organization of concepts from verbal fluency, retrieval strategies must be taken into account. In particular, switching transitions might be altering expected co-occurrence and distance between concepts. In an effort to overcome this issue, we developed an statistical methodology that permitted us to create a network of reliable related concepts. It is noteworthy that finding a network model of semantic memory easily derives to a classifier of switching and clustering, since links between nodes would represent clustering transitions and the absence of links between two nodes would represent switching transitions. The network is later analyzed in terms of topological features with the aim of giving some insight into the characteristics of semantic memory.

Methods

Network theory and its descriptors

In this section, we outline the concepts related to network theory that will be used in this work. For detailed network theory reviews, see (Albert and Barabási 2002; Newman 2003; Boccaletti et al. 2006; Borge-Holthoefer and Arenas 2010b).

First, let us define a *conceptual network* as a graph $\mathcal{G} = (W, \Gamma)$ formed by a set of words $W \equiv \{w_1, \ldots, w_n\}$ that represent concepts (animals in this case) and a set of links $\Gamma \equiv \{\{w_i, w_j\}, \ldots, \{w_k, w_l\}\}$ that represent semantic associations between them. The graph is undirected, which ensures that if a concept w_i is associated with another concept w_j, it is also true that w_j is associated with w_i. For the sake of simplicity, we avoid the possibility that a node contains auto-loops (self-associations) or that two links are connecting the same two nodes. We define N as the *size* of the graph, i.e. the number of nodes (concepts) composing the graph. The structure of a graph is completely described by a $N \times N$ matrix, $A_{\mathcal{G}} = [a_{ij}]$, the so-called *adjacency matrix*. An entry a_{ij} is 1 when the concepts w_i and w_j are linked, and 0 otherwise. In our case, such matrix is symmetrical (i.e., every entry a_{ij} equals to its symmetric a_{ji})

since our graphs are undirected.[1] An undirected graph is said to be *connected* if there exists a possible finite path between all pairs of nodes. Not connected graphs may contain a *giant component* (GC), rawly speaking, a connected sub-graph that contains a majority of the nodes of the graph.

The *degree* of a node w_i, denoted by $k(w_i)$, indicates its number of links and can be easily obtained from the adjacency matrix as

$$k_{w_i} = \sum_{j=1}^{N} a_{ij}. \tag{1}$$

The set of nodes connected by a link to a node is usually referred as the *neighborhood* of this node. The *average degree* of a graph represents the average number of neighbors (concepts linked to a concept) and is defined as

$$\langle k \rangle \equiv \frac{2|\Gamma|}{N}, \tag{2}$$

where $|\Gamma|$ denotes the number of links contained in the set Γ.

The *clustering coefficient* C_i of a node w_i is defined as the proportion of links between the nodes that exist within its neighborhood divided by the number of links that could possibly exist between them (Watts and Strogatz 1998). Its formal expression is given by

$$C_{w_i} = \frac{2E_{w_i}}{k_{w_i}(k_{w_i} - 1)}, \tag{3}$$

where E_{w_i} are the number of actual edges that exist within the neighborhood of node w_i. The *average clustering coefficient* of the nodes is denoted by

$$\langle C \rangle = \frac{\sum_{i=1}^{N} C_{w_i}}{N}. \tag{4}$$

$\langle C \rangle$ is therefore a descriptor of the local connectivity correlations of the network.

In the current work, we will use also the concept of *diameter* (D) of the network referring to the longest among the shortest paths between any two nodes. Finally, $\langle L \rangle$ refers to the *mean path length* of pairwise shortest paths between every two nodes.

Network partitioning in modules provides fruitful information about the organization of a system and the basis of its structure and is one of the major current topics of interest in the field of network theory (Yip and Horvath 2007; Wagner et al. 2007; Danon et al. 2007; Arenas et al. 2008). The generalized topological overlap measure (GTOM) (Yip and Horvath 2007) is a generalization or extension of the topological overlap measure (TOM) (Ravasz et al. 2002) based on the selection of higher-order

---

[1] The absence of auto-loops ensures that all entries of the main diagonal (a_{ii} entries) are 0.

neighborhoods.[2] It provides a robust and sensitive measure of interconnectedness that eases the selection of a cutoff in dendrograms. Hence, the evaluation of different high-order neighborhoods with GTOM is an accurate option to find modules in networks based on empirical evidence, where missing links might be notorious. The basis of GTOM is to take into account the number of m-step neighbors that every pair of nodes share in a normalized fashion. For instance, selecting $m = 1$ is exactly TOM algorithm that measures the overlap coefficient O_{TOM} for every pair of nodes i and j,

$$O_{\text{TOM}}(i,j) = \frac{J(i,j)}{\min(k_i, k_j)}, \tag{5}$$

where $J(i, j)$ is the number of neighbors shared and $min(k_i, k_j)$ is the minimum of the degree of both nodes. However, setting $m = 2$ (GTOM2) considers not only the neighbors shared by every two nodes but also the neighbors of those neighbors. Therefore, the generalization to GTOM can be carried out by growing node neighborhoods,[3] i.e. adding links between those nodes distanced no more than m links in the original adjacency matrix before computing the overlap measure (see Eq. 5). The resulting overlap matrix is transformed to a dissimilarity matrix by converting each entry to $1 - O_{\text{TOM}}(i,j)$. A hierarchical clustering (with averaged linkage criterion) is then performed on the dissimilarity matrix and a cutoff that better separates the matrix in dark blocks (i.e., in sets of nodes with high GTOM) is used to generate a partition of the graph in modules. A Matlab (The Mathworks Inc., Natick, MA, USA) implementation is available as electronic supplementary material (see "Appendix").

Finally, in order to compare the network descriptors defined above with respect to a null model, we used the Erdös Rényi graph (Erdös and Rényi 1960) as a random network model. It consists of spreading links on nodes at random, preserving both the number of nodes and links with respect to the network under study.

Verbal fluency data

Two hundred Spanish speakers were recruited (83 men, 117 women). Participants ranged from 18 to 61 years (mean = 31.8, SD = 11.75), and their education ranged from 5 to 30 years (mean = 15.2, SD = 3.85). Subjects were asked to name all the animals they could in 90 s and responses were transcribed to a text file.[4] Verbal fluency data are included as electronic supplementary material (see "Appendix").

Inference of conceptual associations

Our first aim was to extract relations between concepts based on test evidence in order to obtain a conceptual network (CN). For this, we assumed that a relationship between two words existed when their rate of co-occurrence was significantly higher than what could be expected by chance. The known high rate of switching in fluency tests, averaged as 0.48 by Troyer et al. (1997), indicates that two consecutive words are not necessarily semantically related. Therefore, the use of a statistical assessment in addition to a basal approach based on co-occurrences is critical to discern which concepts are associated when the data comes from verbal fluency tests.[5]

Given the complete set of distinct words $W \equiv \{w_1, w_2, \ldots, w_n\}$ and assuming that words happen within tests at random, the probability of a word w_i to occur in a test is independent of the rest of the test. It corresponds to a Bernoulli variable that can be expressed as

$$\hat{P}_{w_i} = \frac{f_{w_i}}{M}, \tag{6}$$

where f_{wi} is the frequency of w_i within the tests and M is the number of tests (200 in our case). Therefore, the probability of two words being in the same test by chance, $P^{test}_{w_i, w_j}$, is also determined by the product of two Bernoulli variables that occur independently. Their rates of success are obtained independently from the number of occurrences divided by the number of tests evaluated. Hence, $P^{\text{test}}_{w_i, w_j}$ is defined by

$$P^{\text{test}}_{w_i, w_j} = \hat{P}_{w_i} \hat{P}_{w_j} = \frac{f_{w_i} f_{w_j}}{M M}, \tag{7}$$

where f_{wi} and f_{wi} are the frequencies of w_i and w_j, respectively.

Let us define l as the distance between two words in a test.[6] See Fig. 1 for an example of $l = 2$. Given two words occurring in the same test, the probability of being at a distance l, i.e., separated by exactly $l - 1$ words, is

---

[2] Performing a hierarchical clustering directly on the adjacency matrix and setting a threshold in the dendrogram is among the most basic and common approaches used to find modules. Nevertheless, it must be acknowledged that inferred adjacency matrices from empirical data are often noisy or incomplete. This severely affects hierarchical clustering evaluation and misleads the selection of an accurate cutoff value for module detection.

[3] For any m value, GTOM output is a normalized overlap matrix with values between 0 and 1 containing interconnectedness shared information for every pair of nodes.

[4] Every word was converted to its singular and three pure synonyms were unified. Finally, one word that was not an animal was removed.

[5] While methodologies based on co-occurrences have been successfully used to study language networks (Solé et al. 2010), it is important to remark that syntactic constraints severely reduce the possible orderings of items with respect to verbal fluency outputs, where position of concepts is unrestricted.

[6] For instance, $l = 1$ indicate that they are consecutive words. In general, $l = n$ indicate that there are $n - 1$ words between the two words under study.
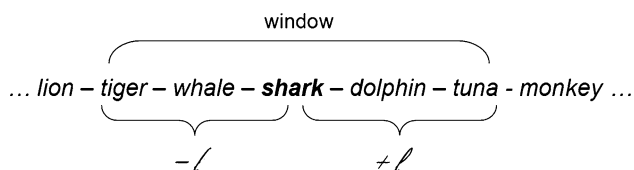
Fig. 1 Example of window length when $l = 2$, as done in the present work. The word sequence represents part of an individual test. When analyzing *shark* relationships, neighbors distanced no more than two words on both sides are taken into account. Hence, in this toy example, *tiger* and *whale* on the left and *dolphin* and *tuna* on the right shark-related candidates

$$P_{w_i,w_j}^{(l)} = 2\frac{N-l}{\left[\begin{smallmatrix}N\\2\end{smallmatrix}\right]} = 2\frac{N-l}{N(N-1)}, \quad 1 \leq l < N. \tag{8}$$

where N is the mean length of tests (a *mean field* approach).[7] The term $2*(N-l)$ is the number of positions of the two words that leave them at distance l within a sequence of length N. The term $\left[\begin{smallmatrix}N\\2\end{smallmatrix}\right] = \frac{N!}{(N-k)!}$ is the total number of positions that two words can occupy within the sequence.[8] This equation can be generalized to the probability of words happening within a window of size l. This is expressed as

$$P_{w_i,w_j}^{(\leq l)} = 2\sum_{i=1}^{l}\frac{N-i}{\left[\begin{smallmatrix}N\\2\end{smallmatrix}\right]} = \frac{2}{N(N-1)}\left(lN - \frac{l(l+1)}{2}\right), \tag{9}$$
$$1 \leq l < N.$$

The expression in Eq. 9 accumulates[9] the probabilities of words being distanced from 1 (consecutive words) to l ($l - 1$ intermediate words). Hence, the probability of two words happening in the same test and window, denoted by $P_{w_i,w_j}^{\text{linked}}$, is

$$P_{w_i,w_j}^{\text{linked}} = P_{w_i,w_j}^{test}P_{w_i,w_j}^{(\leq l)} = \frac{f_{w_i}f_{w_j}}{M\,M}\frac{2}{N(N-1)}\left(lN - \frac{l(l+1)}{2}\right),$$
$$1 \leq l < N.$$
$$\tag{10}$$

The mean cluster size found by Troyer et al. (1997) was 1.09 ± 0.54, where a cluster size of 1 had two words and so on. It basically means that most of the clusters made by participants contain no more than 3 words. Therefore, the expectations of getting semantic information for l greater

---

[7] A more individualized approach could be done by assessing individual test sizes instead.

[8] It is assumed that sequences, i.e. tests, do not contain repeated elements. In the unlikely event of finding a word repeated in a test, neighborhoods for all appearances are considered to obtain co-occurrences.

[9] It is straightforward to see that, when $l = N - 1$, $P_{w_i,w_j}^{(\leq l)} = 2\sum_{i=1}^{N-1}\frac{N-i}{\left[\begin{smallmatrix}N\\2\end{smallmatrix}\right]} = 1$.

than 2 are very reduced. Hence, we chose setting $l = 2$.[10] Given that N and l are 31.57 and 2, respectively, the calculated value for $P_{w_i,w_j}^{(\leq 2)}$ is 0.1246. This is, in our dataset, the basal probability of two words of a test being either consecutive or separated only by a third word by chance.

Afterward, for each pair of words, we obtained the confidence interval ($\alpha = 0.05$) for a binomial distribution given the number of attempts (number of tests) and the number of successes (co-occurrences according to parameter l). Such confidence intervals were computed using the Clopper and Pearson exact method (Clopper and Pearson 1934). The acceptance of an interaction or association between two words was based on whether $P_{w_i,w_j}^{\text{linked}}$ was smaller than the left confidence bound of the interval. This means that we can reject the hypothesis that the $P_{w_1,w_2}^{\text{linked}}$ obtained can be explained by chance. Although the Clopper and Pearson method is particularly appropriate for low-rate success experiments, it is certainly difficult to assess interaction significance for pairs of words with only one co-occurrence, specially when one of them has low frequency.[11]

Hence, we evaluated those pairs of words that co-occurred more than once. This implies that words that did not reach a co-occurrence greater than one with any other word were not included in the inference process (158 out of 400).[12] Additionally, it also implies that any pair of words included in the inference process with a co-occurrence equal to 1 is automatically not linked in the network. Further analyses were carried out in the giant component of the network.[13] The numerical representation of the inferred conceptual network (CN) is a 236 × 236 binary symmetric

---

[10] Setting $l = 1$ would only consider associations for strictly consecutive words, which are more likely to be related with respect to more distant concepts. The high-order variability naming related concepts requires of a large dataset to capture most relationships. A solution to overcome this issue consists of increasing parameter l. However, large windows provide more candidates for establishing relationships of words but at the same time, they reduce the significance of nearby concepts (method explained below) and are more likely to induce meaningless co-occurrences.

[11] For instance, a word named once would be automatically linked to any word named less than 32 times, considering that $N = 31.57$ and $l = 2$ in our dataset.

[12] Removing 39% of distinct words might seem a severe filtering, but they only represented 3.5% of all word occurrences within the tests as they were very low frequent items. Such small reduction of evidence is indeed one step ahead of previous works where semantic distance approaches have been applied to those words either said by a minimum of around 30% of participants or to most named words (threshold set around 12 occurrences) (Henley 1969; Chan et al. 1993; Aloia et al. 1996; Schwartz and Baldo 2001; Prescott et al. 2006).

[13] Those words with no significant interactions were not included in the network (4 words) since they represented isolated words that prevent a network analysis. Additionally, the isolated pair *eel-elver* was also removed for the same reason, leaving a total of 236 nodes in the network.

matrix *A* (see Sect. 2.1 for details). Such matrix contains all possible interactions among words. For every significant relationship between two concepts (w_i, w_j), the entries a_{ij} and a_{ji} were set to 1, and 0 otherwise. A Matlab implementation of the network inference process is available as electronic supplementary material (see "Appendix").

In order to compare our models with a basal co-occurrence approach, a thresholded co-occurrence strategy was also carried out. Using the same window ($l = 2$), different co-occurring thresholds from 1 to 10 were applied on the set of 236 concepts present in CN and ECN. On each case, pairs of concepts co-occurring below the threshold were classified as switching, and concepts co-occurring as many times as the threshold or above were classified as clustering. Table 1 describes four descriptive examples of how the methodology described in this section behaves with respect to an approach based on counting co-occurrences (prior step to thresholding). Examples such as *whale-mouse* co-occurring more frequently than *viper–cobra* in our verbal fluency dataset show the relevance of our approach for the inference process.

Conceptual network enrichment and topological evaluation

The recovery of missing links in inferred and experimental networks is a topic of crucial importance (Mestres et al. 2008) that has been addressed by taking advantage of the network topology, i.e., predicting real missed links based on those already observed (Yip and Horvath 2007; Clauset et al. 2008) and detecting both missing and spurious links (Guimera and Sales-Pardo 2009). In our case, the community structure of CN (i.e., the partition of the graph in modules) obtained by means of the GTOM algorithm (see Sect. 2.1 for details) was the basis of the enrichment

Table 1 Four examples of the concept–concept statistical analysis to decide whether each pair is associated and thus their nodes are linked in the network

Pair of concepts	$\hat{P}_{w_1}$	$\hat{P}_{w_2}$	$P^{\text{linked}}_{w_1,w_2}$	Hits	Interval	Linked
Monkey–horse	0.34	0.51	0.022	2	[0.0012, 0.035]	No
Whale–mouse	0.59	0.45	0.033	6	[0.011, 0.064]	No
Viper–cobra	0.04	0.04	0.0002	4	[0.0055, 0.0504]	Yes
Lion–tiger	0.73	0.59	0.054	91	[0.38, 0.52]	Yes

Pair of concepts indicates the pair studied; $\hat{P}_{w_1}$ is the frequency of the first concept (as defined in Eq. 6); $\hat{P}_{w_2}$ is the frequency of the second concept (as defined in Eq. 7); $P^{\text{linked}}_{w_1,w_2}$ is the value obtained according to Eq. 10; *hits* is the number of times that both concepts were named within a distance not greater than 2 (parameter *l*, see Eq. 8); *interval* is the confidence interval ($\alpha = 0.05$) for the binomial distribution considering the number of *hits* and the number of attempts (number of tests); a pair of concepts is *linked* in the conceptual network only when $P^{\text{linked}}_{w_1,w_2}$ is on the left of the *interval*, i.e., we can reject the hypothesis that the $P^{\text{linked}}_{w_1,w_2}$ obtained can be explained by chance

process in order to provide a reliable conceptual network model. Modules happened to be mostly ruled by semantic constraints, and thus, it is very likely that any node should be reachable from any other node of the same module in one step if there were not missing links. The integration of modular information was carried out setting in the adjacency matrix *A* a value of 1 for every pair of words found in the same module. Thus, every module became a fully connected set of nodes or *clique* (except auto-loops). This neighborhood enrichment produced the enriched conceptual network (ECN), and its visualization was carried out with Pajek (Batagelj and Mrvar 2002). A Matlab implementation of the enrichment process is available as electronic supplementary material (see "Appendix").

Network models used for switching and clustering classification

In order to evaluate CN and ECN as *in-silico* classifiers (evaluation via computer simulation) of clustering and switching transitions, animals not represented in the networks were removed from verbal fluency tests. The 200 tests were converted to binary vectors, where switching and clustering transitions were labeled according to CN and ECN. Every transition was labeled as clustering when both concepts were directly linked on the network and as switching otherwise (see Fig. 5 for a visual representation of the outputs produced by each classifier for all the tests). Those 21 out of 200 tests where more than 10% of concepts had to be eliminated were discarded for the classification task in order to avoid methodological biases. Finally, 20 of the 179 remaining tests were randomly selected. Two human raters[14] manually evaluated switching and clustering for the 600 transitions contained in the tests in order to provide an inter-rater agreement between human expertise, our unsupervised approach and a co-occurrences approach (BCON). Inter-rate agreements between every expert and *in-silico* outputs were measured by kappa coefficient (Cohen 1960).

Results

Verbal fluency data and inference of conceptual associations

The subjects produced a series of animals containing between 16 and 52 words (mean 31.57, SD 6.99). Overall,

---

[14] Raters had experience at the evaluation of verbal fluency tests in healthy controls and neurological patients. They were asked to judge whether each transition between two words was between animals from the same or different subcategories and had for guidance two articles with rules on how to evaluate clustering and switching (Troyer 2000; Villodre et al. 2006). Raters were blind to the results produced by the in-silico evaluations.
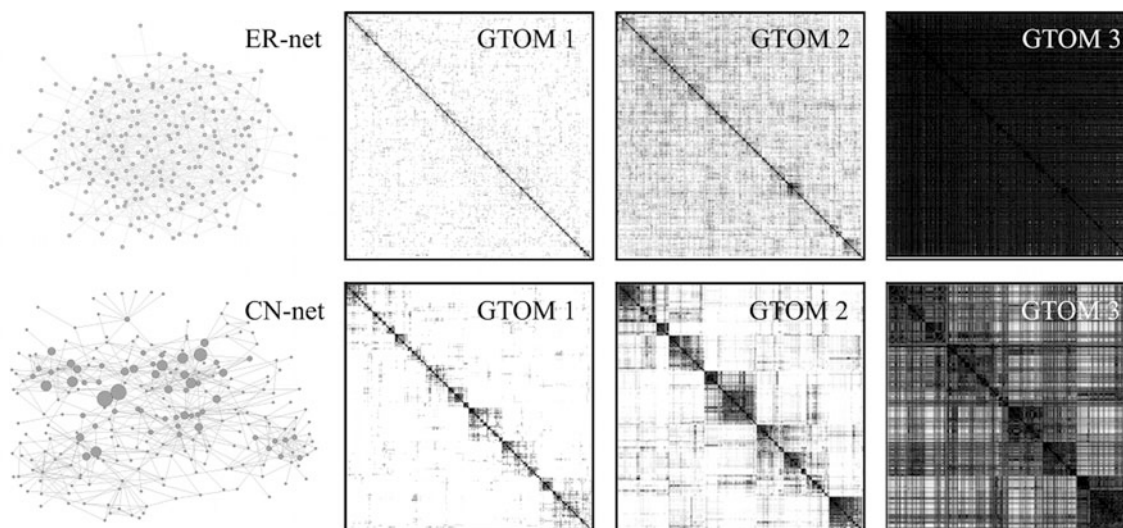
Fig. 2 GTOM orders from 1 to 3 for CN network and a random network (ER-net) with the same number of nodes and links created according to the Erdös-Rényi model. Results indicate the existence of high modularity in the conceptual network inferred, while no modularity appears in the random network

400 distinct animals were listed from which 115 animals appeared only once.[15] We used the previously described statistical approach in a novel fashion that permits the inference of concept associations from verbal fluency tasks taking into account the number of participants, mean test length, window length and word frequencies. The output of this method was an adjacency matrix of the CN. The topological characteristics of such network are summarized in Table 3, and its implications are described in Sect. 3.3.

Modularity analysis

It is widely accepted that semantic memory in general and natural categories in particular must be organized in subcategories. However, which and how many these subcategories are remains poorly understood. From a network perspective, the presence of such categorical organization should be related to the presence of modules in CN. Therefore, our next aim was to study the existence of modularity and, if present, its fundamentals and a characterization of each module. The clearest partition of the network in modules was obtained with GTOM2.[16] Figure 2

shows the absence of modularity in a random network with the same number of nodes and links. Regarding CN, GTOM1 shows the presence of several modules confirmed and better bounded when using GTOM2. For both networks, GTOM3 analysis showed a saturated overlap matrix indicating that no more generalizations were required to be evaluated.

The overlap measure matrix obtained with GTOM2 is represented in Figs. 2 and 3. On the top of the figure, we can see the hierarchical clustering performed on this matrix and the resulting modules colored. Once modules were defined, their content was qualitatively analyzed to report a brief description as inclusive as possible of each module. Table 2 summarizes the 18 modules found and their main characteristics.

In summary, we obtained the presence of 18 modules in an unsupervised manner (Fig. 3). The qualitative analysis of these modules confirmed that they were semantic in nature, contained elements with common attributes and their size was heterogeneous.

Conceptual network enrichment and topological evaluation

Modular semantic knowledge obtained in previous section was incorporated in the network by fully connecting nodes of the same module. Hence, every module became a clique connected with other modules. We refer to those nodes connecting different modules as *frontier animals*, i.e. nodes that have inter-module links. A representation of ECN can be seen in Fig. 4. The topological features before and after the enrichment (CN and ECN, respectively) are shown in Table 3. Enriching the network reduced the diameter from

---

[15] These figures are close to the results of 423 distinct animals, and 175 named only once obtained from 21 participants during 10 min somewhere else (Henley 1969) and might be indicating an average magnitude of the human lexicon size in the category of animals.

[16] The information regarding modularity provided by this matrix is the presence or absence of discrete blocks along the diagonal. When there is no modularity in a network, as it occurs in random graphs, no blocks appear independently of the number of neighborhood expansions until the graph represents itself one module. For those networks where modularity emerges, the selection of a hierarchical clustering cutoff (0.58 in our data) must separate those blocks as well as possible to get a feasible partition of the network in modules.
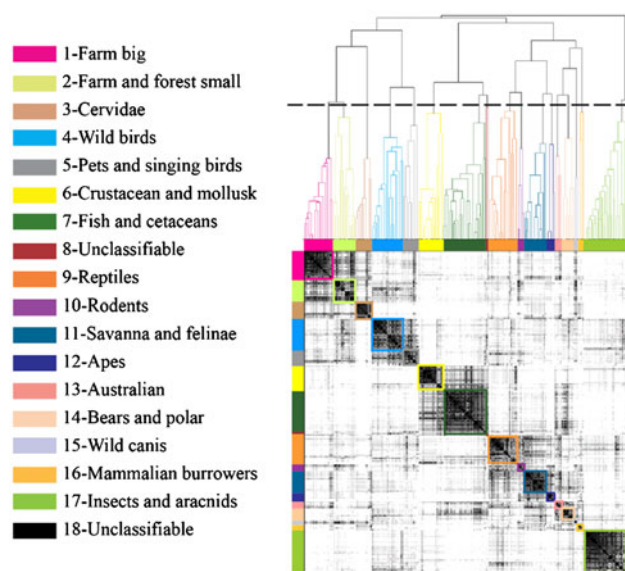
Fig. 3 Dissimilarity based on GTOM2 (gray scale) with a hierarchical clustering on it. Modules obtained correspond to the presence of black blocks along the diagonal of the matrix. On the left, a qualitative description of each module is also included. The two smallest modules (8 and 18) happened to be unclassifiable and they probably belong to other existing modules

Table 2 Description of the modules obtained by the GTOM2 technique applied to the CN network

Id	Description	n	Explored by	σ_{module}	Most frequent
1	Farm-big	21	0.83	0.16	Horse
2	Farm- and forest-small	16	0.85	0.15	Hen
3	Cervidae	12	0.35	0.05	Deer
4	Wild birds	23	0.86	0.10	Eagle
5	Pets and singing birds	11	0.95	0.33	Dog
6	Crustacean and mollusk	18	0.39	0.03	Octopus, crab
7	Fish and cetaceans	31	0.84	0.14	Whale
8	Unclassifiable	2	0.09	0.01	Manta ray
9	Reptiles	21	0.80	0.11	Snake
10	Rodents	5	0.55	0.18	Mouse
11	Savanna and felinae	16	0.93	0.23	Lion
12	Apes	6	0.41	0.12	Monkey
13	Australian	5	0.26	0.06	Kangaroo
14	Bears and Polar	9	0.47	0.11	Bear
15	Wild Canis	3	0.27	0.09	Wolf
16	Mammalian burrowers	4	0.17	0.03	Platypus
17	Insects and Arachnids	32	0.69	0.09	Fly
18	Unclassifiable	1	0.05	0.00	Ferret

Id stands for module position in the dendrogram; *n* is the number of nodes contained in each module; *Explored by* is the proportion of participants that named at least one concept of the module; σ_{module} is the standard deviation of concept frequencies of each module; *Most frequent* is the most cited concept of each module

9 to 6 (i.e. every animal can be attained from any other animal in no more than six steps along ECN) and the mean shortest path length from 4.40 to 3.24 (i.e. the shortest path length between every two nodes is on average shorter in ECN). Both network diameters were quite short due to a small-world phenomenon (Watts and Strogatz 1998) produced by *frontier animals* that act as *short-cuts* i.e. links that connect different regions of the network. Example of animals linking two or more modules are *monkey* and *crocodile*. *Crocodile* is part of the <Reptiles> module but has five links toward animals of <Savanna>, while *monkey* has three links toward animals of <Savanna> but conforms a module with other <Apes>. Finally, the conversion of every module to a clique multiplied by almost four the averaged degree of the network and increased the clustering coefficient from 0.33 to 0.87. As shown in Table 3, the high difference between $\langle C_{rand} \rangle$ and $\langle C \rangle$ for both networks showed the presence of high organization. In other words, concepts indirectly linked through a common neighbor are more likely to be directly linked, a phenomenon not observed when there is a random linkage of nodes in a network.

In-silico classifiers of switching and clustering

ECN aims to represent conceptual storage structure. There is a natural parallelism between the definition of clustering and switching and our conceptual model, where links connect related words and disconnected pairs imply that there is no relationship between the two concepts. Hence, we assessed whether ECN and CN could be used as reliable in-silico evaluators of verbal fluency transitions. Table 4 shows inter-rater agreements among *in-silico* and human judge expertise. With respect to human evaluations, CN is in good concordance with raters (0.71 and 0.70), while ECN shows even a higher agreement (0.82 and 0.83). Indeed, these figures are very close to the kappa coefficient found between the two human raters (0.88), which quantifies the inter-rater reproducibility. Hence, ECN is a conceptual representation closer to human evaluation than CN and represents an unsupervised reliable approach. This implies that the links added to ECN due to the network enrichment process were in benefit of a more accurate classification. Differences between CN and ECN evaluations for the complete dataset are shown in Fig. 5. Regarding a thresholded co-occurrence strategy, both ECN and CN overcome the best co-occurrence approach obtained (BCON), which showed low kappa coefficients with human raters (0.56 and 0.53) . Kappa coefficients obtained for a range of co-occurrence thresholds from 1 to 10, including BCON (threshold = 2) are shown in Fig. 6.
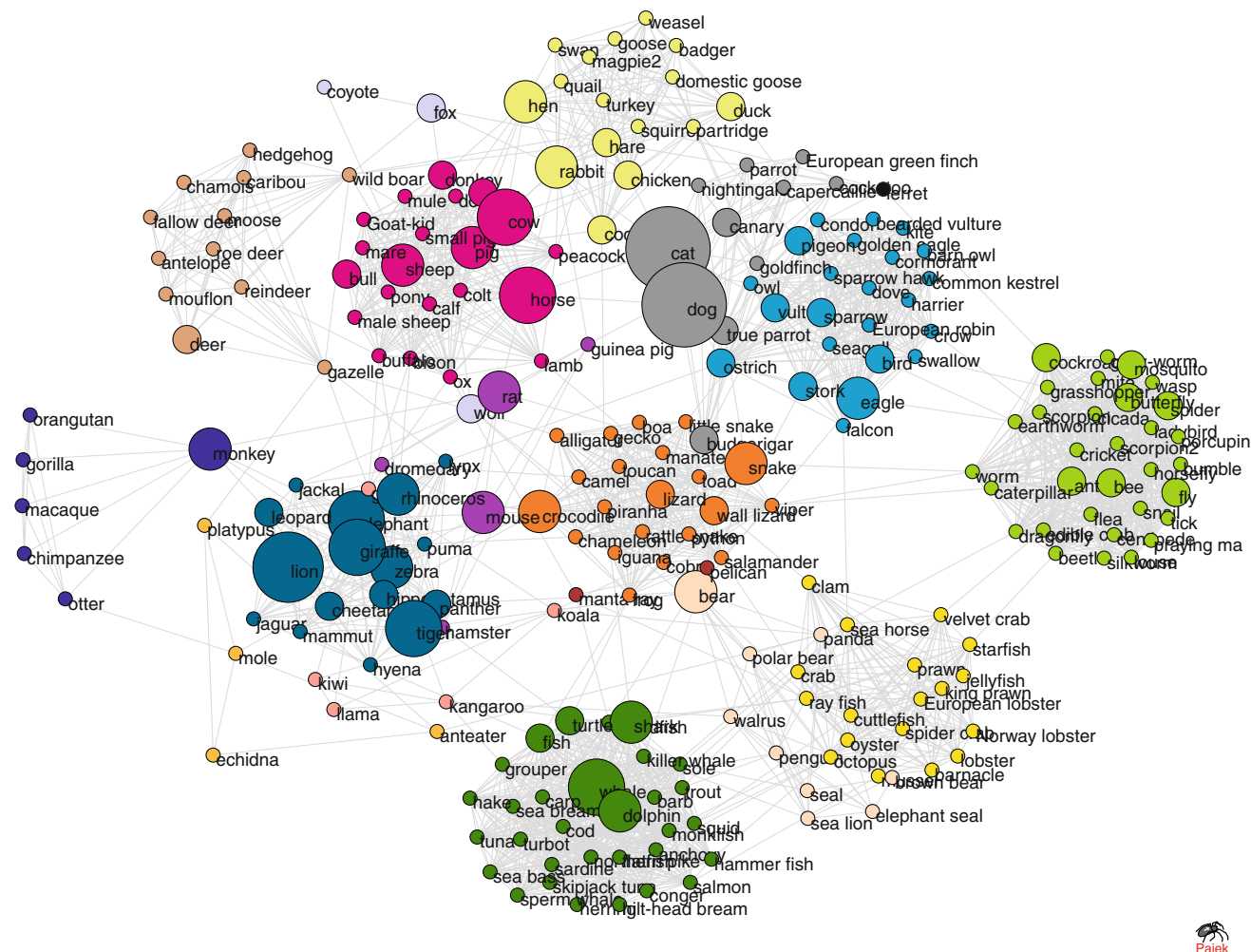
Fig. 4 Enriched conceptual network (*ECN*) is a conceptual organization model inferred from verbal fluency. Size of each node represents its frequency. Each module is identified with a different color in accordance with the color legend of Fig. 3. Links between nodes stand for concept associations and thus represent clustering transitions (related concepts). The absence of links between nodes indicate switching transitions (unrelated concepts, contextual change)

Table 3 Network analysis

Descriptor	CN	ECN	Description
N	236	236	Number of nodes
$|\Gamma|$	611	2,357	Number of interactions
D	9	6	Diameter
$\langle L \rangle$	4.40	3.24	Mean path length
$\langle k \rangle$	5.18	19.97	Average degree
$\langle C \rangle$	0.33	0.87	Average clustering coefficient
$\langle C_{rand} \rangle$	0.02	0.08	$\langle C \rangle$ Expected for a random network

Topological features of the conceptual network (CN) and the enriched conceptual network (ECN). A more detailed explanation of each measure can bee seen at Sect. 2.1

Table 4 Inter-rater agreement

	ECN	BCON	Rater1	Rater2
CN	0.85	0.62	0.70	0.71
ECN	–	0.58	0.82	0.83
BCON	–	–	0.56	0.53
Rater1	–	–	–	0.88

Kappa values among *in-silico* CN, ECN and BCON models and two experienced human raters

Discussion

Our study constitutes an attempt to tackle the complexity of semantic organization by means of network theory and verbal fluency data. By collecting verbal fluency tests from 200 individuals, we have been able to reconstruct a feasible network model of semantic memory, in particular for the natural category of animals. It has been common to use verbal fluency tests to extract representations of semantic memory. This has been usually done using the mean distance between pairs of words and including the most

Fig. 5 CN and ECN *in-silico*
evaluations of switching
transitions (*black*) and
clustering transitions (*white*).
Positions in gray indicate that
the test already ended, i.e., no
more animals were said by that
participant. The network
enrichment process introduced
some modifications in the
evaluation, i.e., some transitions
considered switching by CN
evaluator became clustering
under ECN evaluation
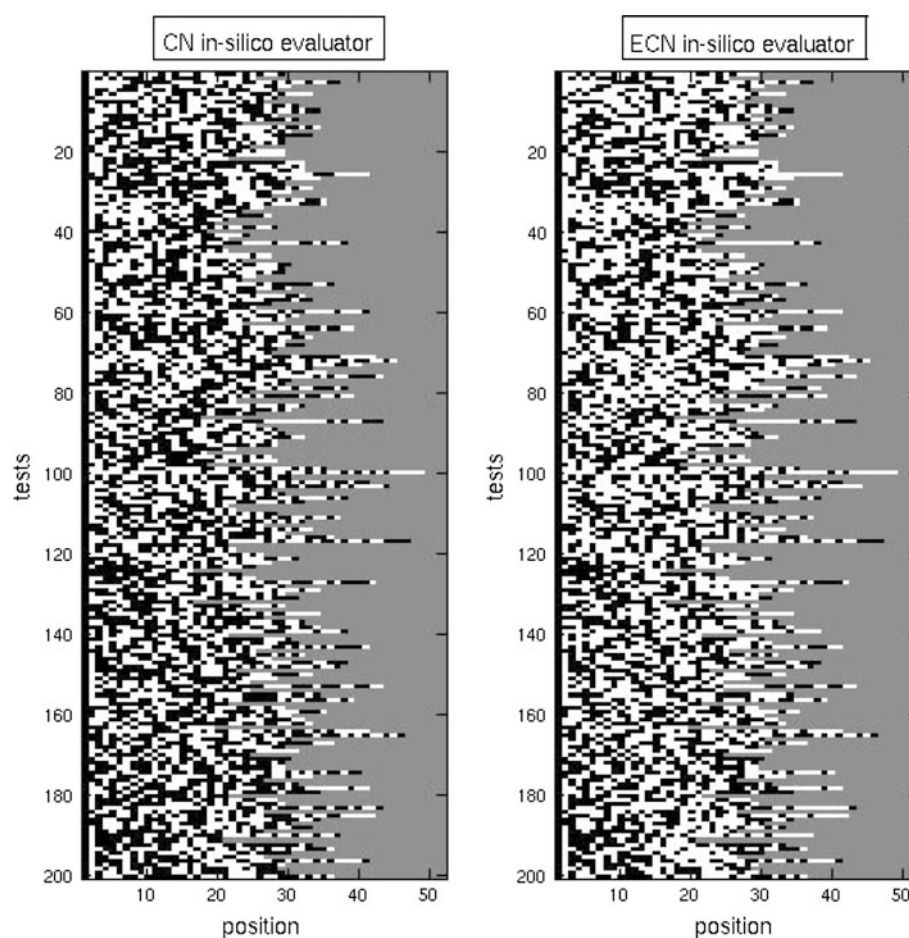
common elements in a multidimensional space (Henley
1969; Crowe and Prescott 2003; Schwartz et al. 2003).
Here, we have developed a methodology that produces a
novel representation of semantic memory as a graph. In our
case, nodes stand for concepts while links between nodes
represent that there is a semantic relationship between
them. Interestingly, the inferred network shows an orga-
nized structure characterized by a high modularity, which
seems to be ruled by a trade-off between conceptual con-
straints such as taxonomy, habitat and size of its concepts.
Additionally, connected and disconnected pairs of concepts
within ECN nicely match to clustering and switching
transitions, respectively, and thus gives rise to an accurate
in-silico classifier when compared to human expert
evaluation.

In Sects. 3.2–3.4, we respectively inferred a conceptual
network, extracted its modules and used them to enrich the
network. CN was obtained linking those concepts that co-
occurred significantly according to the methodology
described in Sect. 2.3. The detection of modules was car-
ried out with the GTOM algorithm and showed 18 modules
strongly addressed by semantic features. The community
structure obtained by the modularity analysis permitted us

to convert each module into a clique to create a final net-
work (ECN). This network connects any two concepts
found to be in the same module, and thus semantically
related, keeping at the same time the links between mod-
ules through frontier animals.

The validity of our model is demonstrated by the fact
that it could be used to classify transitions between words
into clustering or switching as proposed by Troyer. When a
person categorizes a transition as a clustering or switching,
he is making a dichotomous subjective judgment of the
feasibility of a semantic relationship between two words.
The high agreement between our networks and human
raters implies that our methodology was able to catch
important semantic properties that make a pair of concepts
to be subjectively connected. In addition, the outstanding
kappa coefficient obtained confirms the reliability of this
model as a classifier. It could be of use to the psychological
community to evaluate in a fast and reliable way verbal
fluency datasets, with the advantage of not dealing with
inter-rater differences derived from subjective judgments.

Between the two in-silico classifiers, using ECN was
clearly the most accurate. The main difference with respect
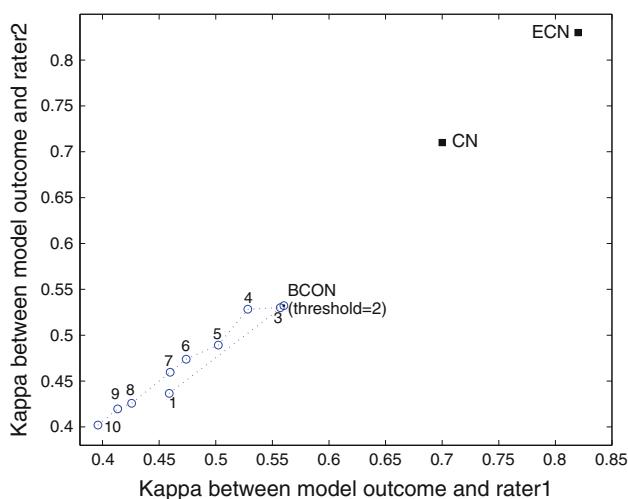to CN was that the modularity found had been exploited to

Fig. 6 Accuracy of the outcomes produced by the network models when compared to human expertise. CN and ECN are the networks inferred in the present study. Numbered points correspond to different thresholds (from 1 to 10) for a co-occurrences approach. Although thresholding at 2 led to the best co-occurring network (*BCON*), its accuracy is clearly overcome by CN and ECN networks

recover missing links between concepts. This points at an important property of semantic memory. It is not a disordered compendium of concepts but an ordered dataset, where it seems that every concept is included in a more general group. In our case, we make a specific proposal of a suitable classification into modules. A qualitative analysis of this classification indicates that most modules had semantic relevance, having their elements many features in common. However, by no means we propose that the modules found here are the only possible ones. A careful analysis of the dissimilarity matrix obtained from GTOM (see Fig. 3) shows evidence of certain hierarchical organization of the modularity, with highly connected submodules nested into bigger ones. Accepted theories on semantic representation and natural categories consider that cognitive categories do not have clear-cut frontiers. Elements are better or worse exemplars of their categories, conforming a typicality decay from the central concepts (Rosch 1974, 1975; Rosch and Mervis 1975). Although a limitation of a modular partitioning is that a concept only pertains to one module, in our approach those animals with a fuzzy module belonging still have links toward nodes of other modules.

A relevant issue addressed in this work is the design of an unsupervised statistical methodology that permits to extract co-occurrences above chance from verbal fluency data taking into account the frequency of each word, a window length, the number of participants and the mean length of the tests. The major advantage of an unsupervised approach is that concept relationships do not depend on expert judgment but only on empirical evidence and

allowed a reliable *in-silico* evaluation of switching and clustering. When compared to previous works of semantic distance (Henley 1969; Chan et al. 1993; Aloia 1996; Schwartz and Baldo 2001; Prescott et al. 2006), our approach does not need concepts to be named by a large proportion of participants and has the benefit of maximizing the final number of concepts taking part in the model. This methodology could be used in the future to explore different domains of semantic memory or to create syntactic networks from linguistic corpora, adding a confidence interval to methodologies already used (Ferrer i Cancho and Solé 2001).

It could be argued that creating dichotomous links between concepts (related vs. not related) is an oversimplification of the complexity of their relationships, which is not lost when using a multidimensional space approach. This is true since we can assume that there are concepts more related or more strongly connected than others. Nevertheless, it is important to remark that our aim was to obtain the underlying network of conceptual organization rather than measuring the semantic distance between concepts or gradients in their navigability (which might be represented by weighted graphs). In this sense, both approaches could be complementary for the study of verbal fluency data, where semantic distance and weighted links both intend to explain exploration phenomena. How this or other cognitive-related networks are explored and how this affects navigability and retrieval efficiency is a question of increasing interest (Boguñá et al. 2009; Goñi et al. 2010; Borge-Holthoefer and Arenas 2010a). Additionally, during verbal fluency tests, the human brain makes dichotomous classifications since switching and clustering, which are defined as mutually exclusive, have been shown to be originated at different neural locations (Troyer et al. 1998a). Similarly, when two people are asked to answer the question of whether two concepts are related or not during a verbal fluency task (i.e. to judge whether the transition has been a clustering or a switching), inter-rater reliability is very high, indicating an important level of agreement. Therefore, we can assume that a dichotomous model of semantic relationships is not opposed to the true reality of semantic memory but complementary to non-binary models. Such binary graphs are the natural output of using a statistical threshold, which ensures that the found relationships were true with a specific level of confidence. Additionally, our methodology, although it is somewhat less precise than non-binary models when qualifying links, is able to recover many reliable associations (note that semantic distance methodologies to date have only permitted to investigate the relationships among the most common items). Another limitation of this study is the use of a single semantic category. Future works could deal with other semantic categories, different in nature to the one

used here. An example could be non-living objects such as tools, which have been shown to activate different brain areas of animals during naming tasks (Chouinard and Goodale 2010). The use of a statistical approach could help to elucidate whether their representation differs from the clustered organization of animals.

How a system is organized greatly influences information retrieval mechanisms and efficiency (Noh and Rieger 2004). We have proposed here a model of semantic storage that could be used to further investigate the characteristics of human memory. The high clustering coefficient and the modular structure of ECN are a consequence of the high level of organization of the semantic storage. Both topological properties will impose severe restrictions on the navigability or exploration of the network. As there is not an unanimous model of semantic retrieval (Wixted and Rohrer 1994), the dynamical behavior of our semantic network when extracting information could be further studied. Investigating how possible retrieval mechanisms are influenced by the topological characteristics of our network would certainly provide interesting results that could give rise to new theories of semantic memory and specifically on how subjects produce semantic fluency outputs.

It is important to highlight that countless retrieval models can be created to explore a network such as ECN. Ideally, the interplay between the conceptual structure and the retrieval model should reproduce relevant features of verbal fluency, including the appearance of words in semantically related clusters (Troyer et al. 1997), the fact that some words are much more frequent than others (Overschelde et al. 2004), the tendency of subjects to produce more frequent words earlier in the test (Bousfield and Barclay 1950), and a similar effect of typicality (Rosch et al. 1976) or age of acquisition (Alvarez and Cuetos 2007) of words. Time effects, such as the appearance of words in spurts followed by silences (Wixted and Rohrer 1994), and the reduction in the production rate as a function of time (Bousfield and Sedgewick 1944) should also be accounted for. A commonly proposed model consists of randomly retrieving concepts (see Wixted and Rohrer (1994) for a review). When applying random graphs, a model whose output is the consequence of a random-walk through networks (Noh and Rieger 2004) can been proposed . While the former completely ignores the semantic structure, the latter totally depends on it. Random sampling models can hardly explain any of the listed effects, with the exception of the increasing silences (assuming that repeated elements manifest as silences). A random-walk on a highly modular graph (as it is the case of ECN) explains the presence of series of semantically related concepts but easily produces repetitions, due to persevering within the same module and thus producing silences from the

beginning. Partial combinations of both kinds of retrieval models are also possible and may overcome some of their limitations. Theoretical efforts in this direction have led to propose cognitive inspired strategies of graph exploration (Goñi et al. 2010). Nevertheless, the validity of a retrieval model when used on our storage representation (ECN) or any other would have to be tested confronting it with empirical data.

Future work could uncover new properties of semantic organization and retrieval in human cognition by applying similar or other topological analysis tools and studying other semantic categories on the networks inferred by this method. Furthermore, this methodology might be useful to better understand the evolution of semantic network acquisition and the relation between verbal fluency skills in neurodegenerative diseases from an unsupervised dual perspective, i.e. storage architecture degradation (Rogers et al. 2004) and impaired retrieval abilities.

Appendix

A Matlab (The Mathworks Inc., Natick, MA, USA) implementation of the methodology described in this study is available as electronic supplementary material. It starts with a verbal fluency dataset and at the last step obtains an enriched conceptual network. In order to ease the use of the code, all these files contain step-by-step explanations and references to sections and equations of this manuscript where appropriate. The script batch_verbal_fluency.m is also very helpful to comprehend the process in a global manner. The modular implementation of the different functions permits their independent use.

- **batch_verbal_fluency.m**: It is the general script that deals with the whole process from the verbal fluency data to the enriched conceptual network. It uses the functions described below.
- **count_words.m**: function that counts the number of words of each verbal fluency test contained in the dataset.
- **get_rel_frequencies.m**: function that gets the relative frequencies of each word included in the verbal fluency data.
- **getco_occurrences.m**: function that counts the number of co-occurrences of every pair words for a given maximum distance (parameter l)

- **get_statistical_co_occurrences.m**: function that performs the statistical approach described in the paper for the network inference.
- **get_components.m**: function that obtains the components of an undirected graph. This is used to obtain the giant component of the conceptual network.
- **computeGTOM.m**: function that performs the modularity analysis using the Generalized Topological Overlap Measure.
- **enrich_newtork.m**: function that performs the enrichment process of a network according to its modularity analysis (which is the output of computeGTOM in our case).
- **write_graph_links.m**: function that writes pairs of words that are linked in a graph according to a dictionary into a file. Each line consists of a pair *word,word*.

  The verbal fluency data of the 200 subjects used in this study are available in the file **data.mat**, which can be loaded typing *load data.mat* in a Matlab environment. The dictionaries of the 236 words included in the networks are available in **dictionaries.mat** (first column in Spanish, second column in English). In the case of Spanish, acute accents and dieresis were omitted and letter ñ was substituted by n.

  Finally, both CN and ECN graphs have been included in Spanish (original language of the tests) and English (translation made by the authors). These files include all the pair of words that are connected (i.e. links of the graph) in a comma separated value format (.csv). These files can be easily visualized as graphs with programs such as Pajek (Batagelj and Mrvar 2002) or Cytoscape (Shannon et al. 2003).
- **CN_spa.csv** is the conceptual network (CN) with animals written in English (graph with 236 nodes and 611 links).
- **CN_eng.csv** is the conceptual network (CN) with animals written in Spanish (graph with 236 nodes and 611 links).
- **ECN_spa.csv** is the enriched conceptual network (ECN) with animals written in Spanish (graph with 236 nodes and 2357 links).
- **ECN_eng.csv** is the enriched conceptual network (ECN) with animals written in English (graph with 236 nodes and 2357 links).

References

Albert R, Barabási AL (2002) Statistical mechanics of complex networks. Rev Modern Phys 74:47

Aloia M, Gourovitch M, Weinberger D, Goldberg T (1996) An investigation of semantic space in patients with schizophrenia. J Int Neuropsychol Soc 2(4):267–273

Alvarez B, Cuetos F (2007) Objective age of acquisition norms for a set of 328 words in spanish. Behav Res Methods 39(3):377–383

Anderson JR (1976) Language, memory and thought. Lawrence Earlbaum, Hillsdale

Anderson JR, Pirolli PL (1984) Spread of activation. J Exp Psychol Learn Mem Cogn 10(4):791–798

Ardila A, Ostrosky-Solís F (2006) Cognitive testing toward the future: the example of semantic verbal fluency (animals). Int J Psychol 41(5):324–332

Arenas A, Fernández A, Fortunato S, Gómez S (2008) Motif-based communities in complex networks. J Phys A Math Theor 41(5):224,001

Batagelj V, Mrvar A (2002) Pajek—analysis and visualization of large networks, vol 2265. Springer, Berlin

Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D (2006) Complex networks: structure and dynamics. Phys Rep 424:175–308

Boguñá M, Krioukov D, Claffy K (2009) Navigability of complex networks. Nat Phys 5(1):74–80

Borge-Holthoefer J, Arenas A (2010a) Categorizing words through semantic memory navigation. Eur Phys J B 74(2):265

Borge-Holthoefer J, Arenas A (2010b) Semantic networks: structure and dynamics. Entropy 12(5):1264–1302

Bousfield W, Barclay W (1950) The relationship between order and frequency of occurrence of restricted associative responses. J Exp Psychol 40(5):643–647

Bousfield W, Sedgewick C (1944) An analysis of sequences of restricted associative responses. J Gen Psychol 30:149–165

Budson AE, Price BH (2005) Memory dysfunction. N Engl J Med 352(7):692–699

Chan A, Butters N, Paulsen J, Salmon D, Swenson M, Maloney L (1993) An assessment of the semantic network in patients with alzheimers-disease. J Cogn Neurosci 5:254–261

Chouinard P, Goodale M (2010) Category-specific neural processing for naming pictures of animals and naming pictures of tools: an ale meta-analysis. Neuropsychologia 48(2):409–418

Clauset A, Moore C, Newman M (2008) Hierarchical structure and the prediction of missing links in networks. Nature 453(7191):98–101

Clopper C, Pearson S (1934) The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 26:404–413

Cohen J (1960) A coefficient of agreement for nominal scales. Educ Psychol Measure 20(1):37–46

Collins AM, Loftus EF (1975) A spreading-activation theory of semantic processing. Psychol Rev 82(6):407–428

Collins AM, Quillian MR (1969) Retrieval time from semantic memory. J Verbal Learn Verbal Behav 8(2):240–248

Collins AM, Quillian MR (1970) Does category size affect categorization time? J Verbal Learn Verbal Behav 9(4):432–438

Crowe S, Prescott T (2003) Continuity and change in the development of category structure: insights from the semantic fluency task. Int J Behav Dev 27(5):467–479

Danon L, Duch J, Arenas A, Díaz-Guilera A (2007) Large scale structure and dynamics of complex networks: from information technology to finance and natural science. World Scientific, Singapore, pp 93–113

Eguíluz V, Chialvo D, Cecchi G, Baliki M, Apkarian A (2005) Scale-free brain functional networks. Phys Rev Lett 94(1):018,102

Erdös P, Rényi A (1960) On the evolution of random graphs. Publ Math Inst Hung Acad Sci 5:17–61

Ferrer i Cancho R, Solé RV (2001) The small world of human language. Proc R Soc Lond Ser B Biol Sci 268(1482):2261–2265

Galeote M, Peraita H (1999) Memoria semántica y fluidez verbal en demencias. Revista Española de Neuropsicol 1(2–3):3–17

Goñi J, Martincorena I, Corominas-Murtra B, Arrondo G, Ardanza-Trevijano S, Villoslada P (2010) Switcher-random-walks: a

cognitive inspired strategy for random exploration on networks. Int J Bifurc Chaos 20(3):913–922

Griffiths TL, Steyvers M, Tenenbaum JB (2007) Topics in semantic representation. Psychol Rev 114(2):211–244

Gruenewald P, Lockhead G (1980) The free recall of category examples. J Exp Psychol Hum Learn Mem 6(3):225–240

Guimera R, Sales-Pardo M (2009) Missing and spurious interactions and the reconstruction of complex networks. Proc Natl Acad Sci USA 106:22,073–22,078

Hayes-Roth B (1977) Evolution of cognitive structures and processes. Psychol Rev 84(3):260–278

Henley NM (1969) A psychological study of the semantics of animal terms. J Verbal Learn Verbal Behav 8:176–184

Jeong H, Mason S, Barabási AL, Oltvai Z (2001) Lethality and centrality in protein networks. Nature 411(6833):41–42

Lerner A, Ogrocki P, Thomas P (2009) Network graph analysis of category fluency testing. Cogn Behav Neurol 22(1):45–52

Lezak M (1995) Neuropsychological assessment, 3rd edn. Oxford University Press, New York

Lund K, Burgess C (1996) Producing high dimensional semantic spaces from lexical co-ocurrence. Behav Res Methods Instrum Comput 28(2):203–208

Mestres J, Gregori-Puigjané E, Valverde S, Solé R (2008) Data completeness-the achilles heel of drug-target networks. Nat Biotechnol 26(9):983–984

Motter AE, de Moura AP, Lan YC, Dasgupta P (2002) Topology of the conceptual network of language. Phys Rev E 65(065102)

Newman M (2003) The structure and function of complex networks. SIAM Rev 45:167–256

Noh JD, Rieger H (2004) Random walks on complex networks. Phys Rev Lett 92(11)

Overschelde JPV, Rawson K, Dunlosky J (2004) Category norms: an updated and expanded version of the battig and montague (1969) norms. J Mem Lang 50:289–335

Patterson K, Nestor PJ, Rogers TT (2007) Where do you know what you know? the representation of semantic knowledge in the human brain. Nat Rev Neurosci 8:976–987

Prescott TJ, Newton LD, Mir NU, Woodruff PW, Parks RW (2006) A new dissimilarity measure for finding semantic structure in cathegory fluency data with implications for understanding memory organization in schizophrenia. Neuropsychology 20(6):685–699

Quillian MR (1967) Word concepts: A theory and simulation of some basic semantic capabilities. Behav Sci 12(5):410–430

Raskin S, Sliwinski M, Borod J (1992) Clustering strategies on tasks of verbal fluency in parkinson's disease. Neuropsychologia 30(1):95–99

Ravasz E, Somera A, Mongru D, Oltvai Z, Barabási AL (2002) Hierarchical organization of modularity in metabolic networks. Science 297(5586):1551–1555

Rips LJ, Shoben EJ, Smith EE (1973) Semantic distance and the verification of semantic relations. J Verbal Learn Verbal Behav 12:1–20

Rogers T, Lambon Ralph M, Garrard P, Bozeat S, McClelland J, Hodges J, K KP (2004) Structure and deterioration of semantic memory: a neuropsychological and computational investigation. Psychol Rev 111(1):205–235

Rosch E (1974) Linguistic relativity. In: Silverstein A (eds) Human communication: theoretical perspectives. Halsted Press, New Sork

Rosch E (1975) Cognitive representations of semantic categories. J Exp Psychol Gen 104(3):192–233

Rosch E, Mervis CB (1975) Family resemblances: studies in the internal structure of categories. Cogn Psychol 7:573–605

Rosch E, Simpson C, Miller RS (1976) Structural bases of tipicality. J Exp Psychol Hum Percept Perform 2(4):491–502

Rosvall M, Bergstrom CT (2008) Maps of random walks on complex networks reveal community structure. Proc Natl Acad Sci USA 105(4):1118–1123

Schwartz S, Baldo J (2001) Distinct patterns of word retrieval in right and left frontal lobe patients: a multidimensional perspective. Neuropsychologia 39(11):1209–1217

Schwartz S, Baldo J, Graves RE, Brugger P (2003) Pervasive influence of semantics in letter and category fluency: a multidimensional approach. Brain Lang 87:400–411

Shannon P, Markiel A, Ozier O, Baliga N, Wang J, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13(11):2498–2504

Sigman M, Cecchi G (2002) Global organization of the wordnet lexicon. PNAS 99(3):1742–1747

Sloman SA (1998) Categorical inference is not a tree: the myth of inheritance hierarchies. Cogn Psychol 35:1–33

Solé RV, Corominas-Murtra B, Valverde S, Steels L (2010) Language networks: their structure, function and evolution. Complexity

Sporns O, Chialvo D, Kaiser M, Hilgetag C (2004) Organization, development and function of complex brain networks. Trends Cogn Sci 8:418–425

Steyvers M, Tenenbaum JB (2005) The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. Cogn Sci 29:41–78

Tröster A, Fields J, Testa J, Paul R, Blanco C, Hames K, Salmon D, Beatty W (1998) Cortical and subcortical influences on clustering and switching in the performance of verbal fluency tasks. Neuropsychologia 36(4):295–304

Troyer AK (2000) Normative data for clustering and switching on verbal fluency tasks. J Clin Exp Neuropsychol 22(3):370–378

Troyer AK, Moscovitch M, Winocur G (1997) Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. Neuropsychology 11(1):138–146

Troyer AK, Moscovitch M, Winocur G, Alexander MP, Stuss D (1998a) Clustering and switching on verbal fluency: the effect of focal frontal- and temporal-lobe lesions. Neuropsychologia 36(6):499–504

Troyer AK, Moscovitch M, Winocur G, Leach L, Freedman M (1998b) Clustering and switching on verbal fluency tests in alzheimer's and parkinson's disease. J Int Neuropsychol Soc 4(2):137–143

Villodre R, Sánchez-Alfonso A, Brines L, Nunez A, Chirivella J, Ferri J, Noe E (2006) Fluencia verbal: estudio normativo piloto según estrategias de agrupación y saltos de palabras en población española de 20 a 49 años. Neurología 21(3):124–130

Voy B, Scharff J, ADPerkins, Saxton A, Borate B, Chesler E, Branstetter L, Langston M (2006) Extracting gene networks for low-dose radiation using graph theoretical algorithms. Plos Comput Biol 2(7):e89

Wagner G, Pavlicev M, Cheverud J (2007) The road to modularity. Nat Rev Genet 8(12):921–931

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, Cambridge

Watts D, Strogatz S (1998) Collective dynamics of 'small-world' networks. Nature 4(393):440–442

Wixted J, Rohrer D (1994) Analyzing the dynamics of free recall: An integrative review of the empirical literature. Psychon Bull Rev 1(1):89–106

Yip AM, Horvath S (2007) Gene network interconnectedness and the generalized topological overlap measure. BMC Bioinform 8(22)