Evidence for thematic representational structure in the mental lexicon

Simon De Deyne (simon.dedeyne@adelaide.edu.au)[a]
Steven Verheyen (steven.verheyen@ppw.kuleuven.be)[b]
Amy Perfors (amy.perfors@adelaide.edu.au)[a]
Daniel J. Navarro (daniel.navarro@adelaide.edu.au)[a]

[a] University of Adelaide, School of Psychology, 5005 Adelaide, Australia
[b] University of Leuven, Department of Psychology, Tiensestraat 102, 3000 Leuven, Belgium

Abstract

Semantic structure in the mental lexicon is often assumed to follow a hierarchical taxonomic structure grouping similar items. This study uses a hierarchical network clustering analysis of a massive word association dataset that does not primarily focus on concrete noun categories, but includes the majority of the words used in daily life. At this scale, we found widespread overlap between thematically organized clusters, arguing against a discrete categoric view of the lexicon. An empirical analysis focusing on taxonomic categories confirmed the widespread thematic structure even for concrete noun categories in the animal domain. Overall, this suggests that applying network clustering to word association data provides valuable insight into how large-scale semantic information is represented. This analysis leads to a different, more thematic topology than the one inferred from idealized small-scale approaches that sample only specific parts of the lexicon.

Keywords: semantic networks; thematic roles; taxonomies; clustering.

One of the most influential ideas in psychological theories about the representation of semantic knowledge holds that concepts are grouped together on taxonomic grounds. In this taxonomic view, the entities that constitute a category are comparable because they have the same function in categories like TOOLS or VEHICLES, or look the same and/or share certain biological properties in categories like FRUIT or INSECTS (Lin & Murphy, 2001). The view that concepts are organized as a taxonomy of categories (e.g., BIRDS or ANIMALS) on the basis of entity feature overlap has been challenged as some studies have attributed a larger role to thematic relations between words that perform complementary roles in the same scenario or event (Gentner & Kurtz, 2005; Lin & Murphy, 2001). Others stress the role of affect in structuring word meaning (Niedenthal, Halberstadt, & Innes-Ker, 1999). These findings are suggestive, but we still lack a comprehensive account of the extent to which these different principles structure lexical meaning and whether the same principles hold at different hierarchical levels that group together entities. Moreover, not all aspects of these different proposals are compatible, nor is it clear at what level of abstraction the proposed principles operate. In addition, many studies of semantic structure are biased towards concrete noun categories (Medin, Lynch, & Solomon, 2000; Medin & Rips, 2005) and most only focus on a small part of the lexicon. This focus on small data sets is of particular concern because many characteristics of networks, including information retrieval, are qualitatively different at different sizes (the *more is different principle*, discussed in Baronchelli, i Cancho, Pastor-Satorras, Chater, & Christiansen, 2013).

The goal of this paper is to explore how the mental lexicon is semantically structured, using a large-scale semantic network that covers the majority of words in the lexicon. In the first part of the paper we use a flexible clustering technique to investigate how the lexicon as a whole is organized, and find evidence for widespread thematic structure. In the second part, we present a more focused empirical test, that investigates whether categories are organized taxonomically. We find that even within concrete noun categories like BIRDS and TOOLS, most of the organizational structure appears to be thematic.

Study 1: Network Clustering

Constructing the semantic network. The network we used was derived from a large scale word association study described in detail in De Deyne, Navarro, and Storms (2013).[1] The study involved a total of 71,380 native Dutch speakers and over 12,400 cue words, and employed a multiple response design in order to better approximate weak associations between words. For the purposes of this investigation, non-dominant word forms (e.g., *apples*) were removed if a dominant form (e.g., *apple*) was also present in the corpus. This resulted in a corpus that contained 2.41 million responses to 11,252 cue words. From this corpus we constructed a network in which each word is a node, and two words are connected by an edge if one word is an associate of the other. In order to avoid over-weighting high-frequency edges between words, response frequencies were transformed to reflect the mutual information between two words (see De Deyne, Verheyen, & Storms, in press; Bullinaria & Levy, 2007, for details). The resulting network captures the semantic relations in the lexicon (De Deyne et al., 2013, in press).

Extracting clusters from networks. Given a semantic network, the goal is to use statistical methods to work out how it is organized. A standard way to do this is to extract clusters of words that are more highly interconnected within a cluster than between clusters. By examining which words tend to form dense clusters, we are able to get a sense of how the lexicon as a whole is structured.

In recent years, network clustering methods have been developed that can handle large networks. One such method is the *Order Statistics Local Optimization Method* (OSLOM) developed by Lancichinetti, Radicchi, Ramasco, and Fortunato (2011). Using OSLOM, clusters can be identified by

[1] The word association project is ongoing and can be accessed at `http://www.smallworldofwords.com`.

Table 1: Overview of the hierarchical cluster structure showing five levels (Level 1 is broadest, Level 5 is most precise). The statistics include total number of clusters N, average cluster size $\langle N_c \rangle$ and its standard deviation, number of homeless nodes $N_{homeless}$, number of nodes member of multiple clusters $N_{overlapping}$, maximum overlap for any node and the average p-value $\langle p \rangle$.

	1	2	3	4	5
N	2	7	37	161	506
$\langle N_c \rangle$	8588	3049	515	112	25
$sd(N_c)$	2112	973	364	66	12
$N_{homeless}$	18	18	39	86	380
$N_{overlapping}$	5943	6956	5263	4717	1676
max(overlap)	2	7	8	10	6
$\langle p \rangle$	0	0.062	0.04	0.035	0.051

evaluating the likelihood that such a cluster would arise in a comparable random network. A full discussion of OSLOM is beyond the scope of this paper, but for the current purposes, it suffices to note that OSLOM allows clusters to overlap and it automatically determines a hierarchical solution if the structure of the network supports this. It is also flexible enough to exclude words from any cluster ("homeless nodes") if they are not sufficiently well connected to others. Finally, each cluster is associated with a p-value, allowing the statistical significance of individual clusters to be assessed. Following Lancichinetti et al. (2011) the choice for p at 0.25 was determined on a case-by-case basis where in this study the worst clusters were still interpretable (see further). For more details on OSLOM and how we applied it to our data, see the supplementary materials.[2]

Results

Applying OSLOM to the semantic network resulted in a solution with five hierarchical levels. An overview of this solution is shown in Table 1. There was a large degree of variability in the number of clusters across the five different levels. On average, the p-value of the extracted clusters was low (.051), indicating that the obtained clusters were unlikely to arise in a comparable random network.

There were few homeless nodes at any level, indicating that most words were reliably attributed to a specific cluster. There was also a considerable degree of overlap at all levels relative to the size of the clusters; clusters were more distinct at the more precise levels, where more clusters were obtained. For instance, at the lowest level 1,676 words appeared in multiple clusters, compared to 5,943 at the highest level.

The fact that many words appear in multiple clusters argues against the idea of a discrete representation in categories. Inspection of these overlapping nodes allows us to grasp how different clusters are related at the same hierarchical level, which might also explain why certain clusters are grouped together at higher hierarchical levels. Membership of multiple clusters often reflects various related senses a word might have. For example, the word *language* (see Figure 1) was attributed to four different clusters related to nationality, speech, language education, and communication, thus explaining important relations between these different clusters. In other cases, overlap might point to separate senses. For instance, the Dutch homonym *bank* (meaning either *financial institution* or *couch*) belonged to both a cluster indicating finance and a cluster for furniture and sitting. Both cases confirm that the obtained solution derives overlapping clusters in a sensible way by accounting for polysemy and homonymy.

[2]The supplementary materials together with the clustering solution at the different hierarchical levels can be downloaded from `http://www.smallworldofwords.com/data/cogsci2015/`

Table 2: First ten (alphabetical) words from four example clusters consisting of N_c members found for different values of p at Level 5 of the hierarchy. Each column shows elements of thematic structure.

$p < 0.0001$, $N_c = 14$	$p < 0.0001$, $N_c = 21$	$p = 0.247$, $N_c = 35$	$p = 0.243$, $N_c = 29$
aikido	envy	anthropology	earthly
combat	envious	dino	real
martial art	dislike	dinosaur	existence
belt	disapproval	dodo	to exist
Japan	aversion	evolution	fact
combat	hate	fossil	factual
kimono	to hate	history	present
to wrestle	grudge	cave	now
judo	bitter	bludgeon	is
judoka	resentment	mammoth	sober

Figure 1 shows the hierarchical clusters obtained by our analysis. At each level the most prototypical examples of clusters are shown, where the typicality of any word is measured by the weighted sum of all links it receives from other words in that cluster. At the most general level (Level 1), there are only two distinct clusters, one of which appears to contain words with negative connotations and one with positive ones.[3] In order to verify whether this interpretation is supported statistically, we used the valence judgments reported by Moors et al. (2012), which are applicable to 3,642 non-overlapping words in our clusters. The valence judgments differed significantly between our two clusters according to an independent t-test ($t(3640) = 7.367$, $CI = [0.190, 0.327]$). This supports the interpretation, and agrees with other research that suggests valence is the most important dimension in semantic space (De Deyne, Voorspoels, Verheyen, Navarro, & Storms, 2014; Samsonovic & Ascoli, 2010) and that category structures are affect-based (Niedenthal et al., 1999).

At lower levels in the hierarchy, the meaning associated with each cluster becomes more concrete. For instance, Level 2 differentiates among the "negative" cluster words at Level 1 – making a distinction between a purely negative cluster (with words like *negative* and *sadness*) and clusters with central nodes like *school*, *religion*, and *money*. The subdivisions of the "positive" cluster involve the central nodes *nature*, *music*, *sports*, and *food* which might be interpreted as covering sensory information and natural kinds.

Inspecting over 500 derived clusters revealed a widespread thematic structure, grouping together entities that could cooc-

[3]Unlike other results reported in the paper, this seems to be slightly dependent on parameter choices: we found that some choices of parameter values produced more than two clusters. However, the positive vs negative distinction was always present.

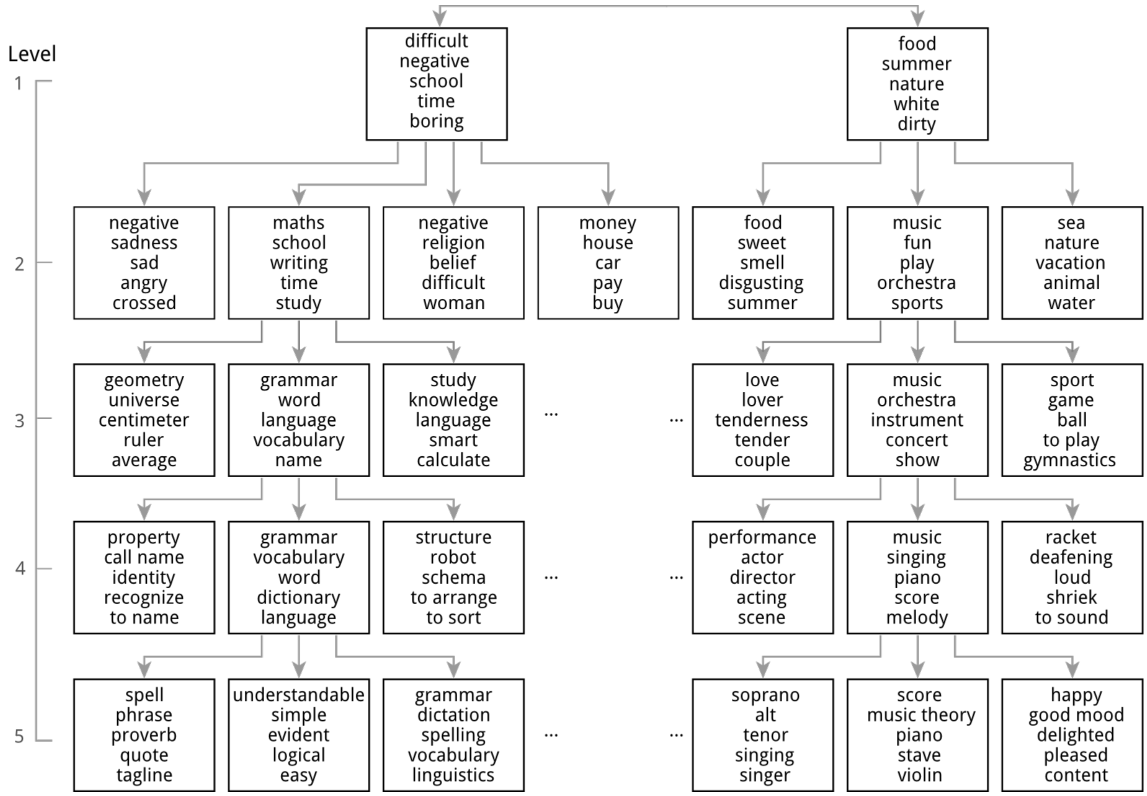

Figure 1: Hierarchical tree visualization of clusters in the lexicon with five most central members.

cur in the same scenario or event more often than what would be expected under a strict taxonomic organization (e.g., *poet* and *poem*). This was the case regardless of the size of the clusters and how significant they were. For example, in the case of low values for p this resulted in clusters like the left column in Table 2, which has taxonomic elements (e.g., *judo*, *karate*, and *aikido*) but also thematic ones (*Japan*, *belt*, *kimono*, *to wrestle*). The inclusion of thematic elements occurred even in clusters focused around non-concrete words, as in the second and last column of Table 2. For example, in the second column we observe mostly abstract negative sentiments and the inclusion of a mixture of adjectives, nouns, and verbs (which do not reflect a pure taxonomy, but also allow for properties and actions). It is also evident that even clusters with fairly high p values were meaningful, as is shown in columns 3 and 4 of Table 2. These clusters have a sensible interpretation, although many words would be appropriate in different clusters as well.

Study 2: Searching for taxonomic categories

Our results so far suggest that clustering of the mental lexicon reveals widespread thematic organization and considerable overlap between clusters. There would seem to be little evidence that taxonomic categories play an important role in organizing the lexicon. However, the findings in the previous section were based on a subjective interpretation of over 500 clusters in an exploratory clustering analysis rather than on an explicit search for taxonomic structure. To redress this, this section presents analyses that conduct exactly such a search.

Method and Procedure

Data from an exemplar generation task by Ruts et al. (2004) were used to identify members of the most commonly used taxonomic categories in the literature (see for example Rosch, Mervis, Grey, Johnson, & Boyes-Braem, 1976; Hampton, 1979). In this task, 100 participants generated as many exemplars they could think of for six artifact categories (CLOTHING, KITCHEN UTENSILS, MUSICAL INSTRUMENTS, TOOLS, VEHICLES, and WEAPONS) and seven natural kinds (FRUIT, VEGETABLES, BIRDS, INSECTS, FISH, MAMMALS, and REPTILES). This resulted in a total of 789 words, of which 588 were included in the word association data (the missing words tended to be low frequency words like *komodo* or *shiruken*). The names of the categories and the number of exemplars obtained through this procedure are shown in the first two columns of Table 3.

The critical question is whether the clusters extracted in Study 1 include anything that might correspond to concrete noun categories like BIRDS and VEGETABLES. For each of our categories, we found the best matching cluster and calculated the precision and recall in terms of the F-measure for clustering performance (see Ball, Karrer, & Newman, 2011, for more details regarding the use of F in clustering). A taxonomic-like organization would be evident in clusters with high precision and recall, resulting from many true positives and few false positives and false negatives. For instance, if the cluster corresponding to the category BIRDS contained *robin* (a true positive) and did not contain *spoon* (a true negative), that would increase the F-score. Conversely, if it contained *guitar* (a false positive) or did not contain *ostrich* (a false negative), that would decrease the F-score. This way, high

Table 3: F-values and cluster sizes for items generated for 13 concrete noun categories. N_{human} is the category size based on the exemplar generation task; N_c is the size of the best-matching cluster; F captures precision and recall according to the human categories for the full network. F' is calculated from a network that excluded potential thematic information. F-values are fairly low, indicating a rather low correspondence between the clusters and the taxonomic categories. Excluding thematic information results in F' values that do capture taxonomic information.

Category	N_{human}	N_c	F	F'
FRUIT	40	50	0.47	0.84
VEGETABLES	35	58	0.50	0.90
BIRDS	53	63	0.53	0.90
INSECTS	40	34	0.46	0.68
FISH	37	48	0.57	0.91
MAMMALS	61	21	0.20	0.76
REPTILES	21	22	0.65	0.51
Mean	41	42	0.48	0.79
CLOTHING	46	70	0.35	0.80
KITCHEN UTENS.	71	18	0.20	0.66
MUSICAL INSTR.	46	24	0.37	0.89
TOOLS	73	56	0.25	0.76
VEHICLES	46	28	0.16	0.73
WEAPONS	46	25	0.37	0.88
Mean	55	37	0.28	0.79

F-scores should reflect categories that are not overly specific (many false negatives) or general (many false positives).

Results

The best matching clusters were found at the lowest level in the hierarchy (Level 5). As Table 3 shows, for natural kinds the average number of generated exemplars (41) was similar to the average number of cluster members (42). However, for artifacts the clusters contained fewer members (55 vs 37). The F-values were on average 0.48 for the natural categories and 0.28 for the artifacts. Exceptions like FISH notwithstanding, the results indicate only limited support for the presence of a taxonomic organization. Moreover, the difference between both domains is consistent with previous work finding that artifact categories do not have as clear of a delineation as natural kinds do (Ceulemans & Storms, 2010). Overall the F-values are low, suggesting that the network structure of the mental lexicon does not support a general and strict taxonomic organization.

To explain why even the best-matching clusters provide poor approximations to taxonomic categories, Table 4 lists the five most central false positives. In some cases these words could in fact be interpreted taxonomically. For instance, in several cases the category label was included in the cluster. However, for the most part the intrusions are thematic in nature: *beak*, *egg*, *nest*, and *whistle* appear in the BIRDS cluster; in the case of FRUIT, intrusions included *juicy*, *pick*, and *summer*. In other words, the thematic interpretation of entities at the lowest level in Figure 1 (e.g., *score*, *music theory*, *piano*, *stave*, *violin*) is now confirmed empirically for a total of 13 frequently used categories.

Overall, our inability to find a taxonomic organization even for biological categories, combined with the widespread thematic structure across nearly all clusters, strongly suggests that multiple factors contribute to structure in the mental lexicon, and thematic relations are a major one of them.

Finding taxonomies by restricting the network. One potential response to the previous analyses relates to the nature of the data upon which they are based. Perhaps the word association task simply fails to capture taxonomic information, and if so, the results of these analyses are simply an artifact of the choice of task. Alternatively, perhaps the "failure" arises because the word association task is more general than the tasks typically used to study taxonomic categories.

There is some evidence that a different choice of task would produce different choices. For instance, much of the work on taxonomic organization relies on tasks in which participants are asked to list features of entities (e.g., Ruts et al., 2004). One could argue that feature generation is a constrained version of the word association task, and the key difference is the number of thematic responses one gets in both procedures. Similarly, feature generation stimuli are usually restricted to concrete nouns, which places restrictions on what words *can* be grouped together. In other words, the tendency to find taxonomic categories may be a result of restricting the task.

To test this idea, we used the word association data to construct a network that included *only* those 588 words that belonged to one of the taxonomic categories. Moreover, in order to approximate the "shared features" measure that is more typical of feature generation tasks, we computed the cosine similarity between pairs of words. That is, words that have the same associates are judged to be more similar, and this similarity was used to weight the edges in the restricted network.[4] We then applied the clustering procedure to this restricted network and repeated the analysis from the previous section. The F-statistics from this analysis are reported as the F'-values in Table 3. This time, the results of the clustering show a high degree of agreement with the taxonomic organization, with an average F-value of 0.79. The only exception was REPTILES, which upon inspection appears to reflect a failure to distinguish REPTILES from INSECTS.

The success of this analysis suggests two things. First, the word association task *does* encode taxonomic information, as evidenced by the fact that we are able to reconstruct taxonomic categories. However, the fact that the only way to do so is to mimic all the restrictive characteristics of a feature generation task (e.g., limited word set) is revealing. Taxonomic information is not the primary means by which the mental lexicon is organized: if it were, we should not have to resort to such drastic restrictions in order to uncover taxonomic categories.

[4] Note that one could also derive such a similarity-based network for the complete lexicon, which would reflect the similarity between cues rather than their weighted associative strength. We did in fact do this. It produced results similar to the original analysis.

Table 4: Top 5 false positives ordered by cluster in-strength per category. Most of the false positives are thematic in nature. For instance, false positives for the BIRDS category include *beak*, *egg*, *nest*, and *whistle*.

Category	1	2	3	4	5
FRUIT	fruit	juicy	pit	pick	summer
VEGETABLES	vegetable	healthy	puree	sausage	hotchpotch
BIRDS	bird	beak	nest	whistle	egg
INSECTS	insect	vermin	beast	crawl	animal
FISH	fish	fishing	rod	slippery	water
MAMMALS	rodent	gnaw	tail	pen	marten
REPTILES	reptile	scales	animal	tail	amphibian
CLOTHING	clothing	fashion	blouse	collar	zipper
KITCHEN UTENSILS	cooking	kitchen	stove	cooker hood	burning
MUSICAL INSTR.	wind instrument	to blow	fanfare	orchestra	harmony
TOOLS	tools	carpenter	carpentry	wood	drill
VEHICLES	speed	drive	vehicle	motor	circuit
WEAPONS	sharp	stab	blade	point	stake

General Discussion

The prominent view that the mental lexicon is organized in taxonomic categories remains highly influential. It dates back to knowledge ontologies proposed by Aristotle and Linnaeus and lives on in current encyclopedias like Wikipedia or search engines like Yahoo! (Inc). While it might be useful or at least economical to retrieve information from an encyclopedia by agreeing on a single taxonomic index, information stored in the mental lexicon seems accessible in a different way.

Despite the ubiquity of the taxonomic view, we found only limited evidence (at best) for a dominant taxonomic organization in a large-scale semantic network derived from human word-association judgments; instead, thematic organization was widespread. The pervasiveness of thematic organization occurred even in a typical taxonomic domain like animals. In previous work the pervasive contribution of thematic or relational knowledge may have been overlooked due to a selection bias stemming from a focus on certain concepts (nouns, mostly concrete) and semantic relations (mainly taxonomically defined).

In other work we have performed similar analyses that suggest that the thematic organization observed here does not depend strongly on the type of data (word associations) from which the network is constructed. Our results showed widespread thematic organization when using networks based on primary responses only (unlike this work, which used three responses) as well as models derived from syntactic dependency relations in written and spoken corpora (De Deyne, Verheyen, & Storms, 2015). The main difference observed in the latter situation is that models derived from text resulted in fewer clusters as a consequence of the smaller signal-to-noise ratio in the text-based models compared to word association based ones. For the word association network based on primary responses only, the evidence for a taxonomic organization was slightly stronger because the category labels itself were more often given as a primary than as a secondary or tertiary response (e.g, *alligator – reptile*; De Deyne & Storms, 2008). Even these networks were predominantly thematically organized, though. Altogether, these findings add support to the idea that our results are robust against differences in the amount, type, and quality of the data from which the lexicon is derived.

Broader Implications. While our results do not exclude the role of taxonomic grounds for grouping entities in special cases like FISH, thematic structure was widespread, showing up in nearly all investigated clusters at all depths of the derived hierarchy. The finding that many words from domains like animals (which are traditionally considered taxonomic) are thematically clustered, even at the lowest level of the hierarchy, supports the idea that the networks are organized primarily along thematic rather than taxonomic or categorical lines. Other studies that evaluate semantic structure in the mental lexicon also suggest that the network is thematic. It can account better for thematic relatedness judgments than taxonomic relatedness judgments (De Deyne et al., in press) and has been shown to facilitate word processing when thematic but not coordinate prime-target pairs are used (Hutchison, 2003). This converges with recent evidence that highlights the role of thematic representations even in domains such as animals (Gentner & Kurtz, 2006; Lin & Murphy, 2001; Wisniewski & Bassok, 1999) and the fact that a taxonomic organization of knowledge might be heavily culturally defined (Lopez, Atran, Coley, Medin, & Smith, 1997), a consequence of formal education (Sharp et al., 1979), or reflect different levels of expertise (Medin, Lynch, Coley, & Atran, 1997).

While other reserachers have argued for a role for thematic relations, these relations haven't received nearly as much attention as taxonomic structure. One explanation is that in contrast to previous studies, our network has a wide coverage of all kinds of words – words that vary widely in terms of their abstractness, emotional connotation, and part of speech (verbs, adjectives, and nouns). By not restricting the type of words in the network, the risk of a selection bias towards concrete nouns is reduced and the likelihood of identifying thematic representations increases (Medin et al., 2000; Medin & Rips, 2005). In addition, it is quite likely that the thematic organization reflects an inherent property of language: most words are taxonomically related to only a small number of other words, but can occur in a variety of thematic settings. This is in line with previous findings showing that Zipf's law reflects the tendency to avoid excessive synonymy in semantic networks (Manin, 2008).

Our findings have implications that affect other domains of language processing as well. They might explain what kind of semantic information is likely to become activated in tasks like priming or word recognition. In the case of priming, for example, it could explain the distinct effects found for thematically and taxonomically related pairs (Hutchison, 2003). Similarly, in word recognition it may provide a way to understand the processing advantages associated with words that have "rich" semantics (Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008), which were previously explained in terms of the number of entity features a word has or the number of contexts in which it occurs.

A final point worth noting is that the present work focuses on semantic representations; this could highlight other properties than modal-specific representations. For instance, language-based representations might underestimate the richness of certain perceptual properties (Pecher & Zwaan, 2005). It is possible that a similar selection bias pertains to non-linguistic concept representations as well. For instance, an encounter with a certain bird or interaction with a specific musical instrument involves more than just encoding of the perceptual features of each – at the least there is other contextual information, like habitat or venue, that ought to be part of modal-specific proposals. Overall, our work suggests that the classic taxonomic view is unnecessarily restrictive if the goal is to understand how semantic or concept information is stored or retrieved in a wide range of tasks.

Acknowledgments

This research has been supported by an ARC grant DE140101749 awarded to the first author. SV is a postdoctoral fellow at the Research Foundation - Flanders. Salary support for AP and DJN was provided by ARC grants DE120102378 and FT110100431, respectively.

References

Ball, B., Karrer, B., & Newman, M. (2011). Efficient and principled method for detecting communities in networks. *Physical Review E*, *84*(3), 036103.

Baronchelli, A., i Cancho, R. F., Pastor-Satorras, R., Chater, N., & Christiansen, M. H. (2013). Networks in cognitive science. *Trends in Cognitive Sciences*, *17*(7), 348–360.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*, 510–526.

Ceulemans, E., & Storms, G. (2010). Detecting intra-and inter-categorical structure in semantic concepts using HICLAS. *Acta Psychologica*, *133*(3), 296–304.

De Deyne, S., Navarro, D. J., & Storms, G. (2013). Better explanations of lexical and semantic cognition using networks derived from continued rather than single word associations. *Behavior Research Methods*, *45*, 480–498.

De Deyne, S., & Storms, G. (2008). Word Associations: Network and Semantic properties. *Behavior Research Methods*, *40*, 213–231.

De Deyne, S., Verheyen, S., & Storms, G. (2015). Structure and organization of the mental lexicon: a network approach derived from syntactic dependency relations and word associations. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, & B. Job (Eds.), *Towards a theoretical framework for analyzing complex linguistic networks.* Berlin/New York: Springer.

De Deyne, S., Verheyen, S., & Storms, G. (in press). The role of corpus-size and syntax in deriving lexico-semantic representations for a wide range of concepts. *Quarterly Journal of Experimental Psychology*.

De Deyne, S., Voorspoels, W., Verheyen, S., Navarro, D. J., & Storms, G. (2014). Accounting for graded structure in adjective categories with valence-based opposition relationships. *Language and Cognitive Processes*, *29*(5), 568–583.

Gentner, D., & Kurtz, K. J. (2005). Relational categories. In W. K. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman, & P. W. Wolff (Eds.), *Categorization inside and outside the lab.* (pp. 151–175). American Psychology Association.

Gentner, D., & Kurtz, K. J. (2006). Relations, objects, and the composition of analogies. *Cognitive Science*, *30*, 609–642.

Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, *18*, 441–461.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? *Psychonomic Bulletin and Review*, *10*, 785–813.

Lancichinetti, A., Radicchi, F., Ramasco, J. J., & Fortunato, S. (2011). Finding statistically significant communities in networks. *PloS one*, *6*(4), e18961.

Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, *1*, 3–28.

Lopez, A., Atran, S., Coley, J. D., Medin, D. L., & Smith, E. E. (1997). The tree of life: Universal and cultural features of folkbiological taxonomies and inductions. *Cognitive psychology*, *32*(3), 251–295.

Manin, D. Y. (2008). Zipf's law and avoidance of excessive synonymy. *Cognitive Science*, *32*(7), 1075–1098.

Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to rome? *Cognitive psychology*, *32*(1), 49–96.

Medin, D. L., Lynch, E. B., & Solomon, K. O. (2000). Are there kinds of concepts? *Annual review of psychology*, *51*(1), 121–147.

Medin, D. L., & Rips, L. J. (2005). Concepts and categories: memory, meaning, and metaphysics. In K. Holyoak & R. Morrison (Eds.), *The cambridge handbook of thinking and reasoning* (pp. 37–72). Cambridge University Press, Cambridge, UK.

Moors, A., Houwer, J. D., Hermans, D., Wanmaker, S., van Schie, K., Harmelen, A.-L. V., ... Brysbaert, M. (2012). Norms of valence, arousal, dominance, and age of acquisition for 4,300 dutch words. *Behavior research methods*, 1–9.

Niedenthal, P. M., Halberstadt, J. B., & Innes-Ker, Å. H. (1999). Emotional response categorization. *Psychological Review*, *106*(2), 337.

Pecher, D., & Zwaan, R. A. (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge University Press.

Pexman, P. M., Hargreaves, I. S., Siakaluk, P. D., Bodner, G. E., & Pope, J. (2008). There are many ways to be rich: Effects of three measures of semantic richness on visual word recognition. *Psychonomic Bulletin & Review*, *15*(1), 161–167.

Rosch, E., Mervis, C., Grey, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*, 382–439.

Ruts, W., De Deyne, S., Ameel, E., Vanpaemel, W., Verbeemen, T., & Storms, G. (2004). Dutch norm data for 13 semantic categories and 338 exemplars. *Behaviour Research Methods, Instruments, and Computers*, *36*, 506–515.

Samsonovic, A. V., & Ascoli, G. A. (2010). Principal semantic components of language and the measurement of meaning. *PloS one*, *5*(6), e10921.

Sharp, D., Cole, M., Lave, C., Ginsburg, H. P., Brown, A. L., & French, L. A. (1979). Education and cognitive development: The evidence from experimental research. *Monographs of the society for research in child development*, 1–112.

Wisniewski, E. J., & Bassok, M. (1999). What makes a man similar to a tie? *Cognitive Psychology*, *39*, 208–238.