Title

Learning word meaning with little means:An investigation into the inferential capacity of paradigmatic information

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 40(0)

Authors

de Deyne, Simon
Perfors, Amy
Navarro, Danielle J

Publication Date

2018

Learning word meaning with little means: An investigation into the inferential capacity of paradigmatic information

Simon De Deyne (simon.dedeyne@unimelb.edu.au)[a]
Amy Perfors (amy.perfors@unimelb.edu.au)[a]
Danielle J. Navarro (d.navarro@unsw.edu.au)[b]

[a] University of Melbourne, School of Psychological Sciences, 3005 VIC, Australia
[b] University of New South Wales, School of Psychology, 2052 NSW, Australia

Abstract

To what extent can the similarity structure of categories be inferred based on paradigmatic vs syntagmatic information? We explore this question in two studies that aim to capture paradigmatic information directly: first by having participants generate near-neighbors to exemplars from 15 basic categories, and second by having them partially rank the most similar exemplars. After constructing neighborhood graphs of the items in each category, we derived a local measure (based on direct neighbors) and a global measure (including indirect paths as well) of paradigmatic information. Both measures predict independently-obtained human pairwise similarities for each category, but incorporating indirect information substantially improves this prediction. In a third study, we contrast these measures with syntagmatic information obtained from a vast semantic network derived from 3 million judgments. The paradigmatic graphs are better predictors of similarity despite only encoding a fraction of these data. Broad implications for word learning and meaning are discussed.

Keywords: similarity; semantic networks; near neighbors; word associations; mental lexicon

Introduction

When we encounter new words, they usually do not come conveniently attached to definitions. Rather, we must somehow infer the meaning based on a small collection of cues. When the new words are learned through reading, the number of cues drops even more. Despite the difficulty of the task, people excel at solving this inductive problem and learning the meaning of words through text based on only a few examples. For example, people can infer geographical locations or social structures based on text (Louwerse & Connell, 2011), and even congenitally blind people have representations of concrete, visible entities that in many ways match those of sighted people (Lenci, Baroni, Cazzolli, & Marotta, 2013). What kind of information are they using? As a first step, what kind of information is most useful?

In considering this question, imagine that while perusing an interesting novel you see the new word *capybara*. The text provides some cues as to its meaning. You might notice its **syntagmatic** relationships – the words that co-occur together in close temporal proximity but have a syntactically distinct role. In this case, the sentence is "the capybara lives in the savannah", so *lives* and *savannah* are syntagmatic relations. Conversely, it also has **paradigmatic** relationships, which are related (e.g., in the same category as) and fulfill the same (syntactic) role in the sentence (e.g., *mouse* or *guinea pig*). [1]

From a developmental perspective, the formation of paradigmatic associations depends on the existence of syntagmatic associations and the former are expected to be better cues early in word learning compared to the latter (Sloutsky, Yim, Yao, & Dennis, 2017). Regardless of whether paradigmatic relations are acquired through some kind of higher-order associative learning (Ervin, 1961) or are directly encoded in language, the number of plausible paradigmatic relations is likely to be substantially less than the number of syntagmatic relations a word participates in at any point in the development. A word like *capybara* may co-occur with thousands of other words, whereas a much smaller number fill the same role. This logic suggests that paradigmatic information (which might be explicit in language) may be more informative about the underlying meaning of *capybara*: if there are relatively few paradigmatic relations for any given concept, it may imply that each one carries more inferential weight. If so, this suggests a possible resolution to the sparsity problem: people can learn so quickly from text because they have access to data – paradigmatic relations – that is highly informative as to category meaning.

The idea is intriguing, but the logic is only suggestive at best. In this study we therefore put it to the test. Our hypothesis is that paradigmatic relationships are far more informative about category structure than even orders of magnitude more syntagmatic data, at least when the information inherent in the paradigmatic data is appropriately extracted.

We evaluate this hypothesis by using a network approach to extract the full extent of structure inherent in purely paradigmatic relationships (measured in two different ways in Study 1 and Study 2 respectively). The network captures the notion that the meaning of each paradigmatically-related word depends on the meaning of all other paradigmatically-related words. More importantly, it also allows us to build an interconnected structure based on extremely sparse data in which each word is directly related to only a small number of other words. Inspired by Collins and Loftus (1975), we implement a simple mechanism of spreading activation which allows us to overcome this sparsity problem by considering not only paths between directly adjacent words, but also longer indirect paths between more distant words. In this way, we can build

[1] Although less typical, some paradigmatic relationships may *also* co-occur in language through conjunctions ("capybaras and guinea pigs") or similes ("capybaras, like chinchillas"): they are still paradigmatic because of the role they play.

up a rich representation based on sparse examples of only paradigmatic information.

For the purposes of this paper we restrict our investigation to words on the basic level. Thus, our question is to what extent word meaning can be inferred from paradigmatic relations in categories like *Birds* with basic-level exemplars (like *ostrich* or *sparrow*). We focus here because exemplars at the basic level are special in many ways: they are concrete, acquired early in life, and carry the most information (Rosch, Mervis, Grey, Johnson, & Boyes-Braem, 1976). However, it is still not entirely clear how they are learned or what data are most informative about their meaning.

The structure of this paper is as follows. We first describe the pairwise similarity data used for prediction in all of the subsequent studies. Then, in Study 1 we use a near-neighbor generation task to derive a network of paradigmatic relationships for 15 basic-level categories, as described above. We demonstrate that this information predicts the similarity data reasonably well. In Study 2, we show that this result is not a byproduct of the particular task; paradigmatic data obtained from a ranking task shows the same effect. In Study 3, we contrast these results with predictions based on a vastly larger quantity of syntagmatic-only information derived from a large-scale semantic network. Despite the fact that the syntagmatic information contains orders of magnitude more data, it predicts similarity structure worse than the (much sparser) paradigmatic information from Study 1 and Study 2.

Pairwise similarity data

All of the studies in this paper involve predicting human-rated pairwise similarity data. The data consists of 418 exemplars from 15 different semantic categories, each of which was rated for pairwise similarity by at least 15 participants in a previous study (De Deyne et al., 2008). The categories consisted of animals (30 *Birds*, 23 *Fish*, 26 *Insects*, 30 *Mammals*, 20 *Reptiles*), artifacts (29 *Clothing*, 33 *Kitchen utensils*, 27 *Music instruments*, 30 *Tools*, 30 *Vehicles*, 20 *Weapons*), and other categories (30 *Fruits*, 30 *Vegetables*, 30 *Sports*, 30 *Professions*). All items in this and subsequent studies were in Dutch, but for clarity are presented as their English translations.

Study 1: Near-neighbor Generation

The near neighbor generation (NNG) task is a production task similar to both word associations and feature generation. It differs from both because of how it is censored: in it, only paradigmatic (i.e., coordinate) responses are valid. By contrast, in the feature-generation task only syntagmatic responses are permitted, and a mixture is allowed in word associations. The NNG task thus yields a set of paradigmatic-only data.

Method

Participants. Participants were an opportunity sample of 363 first-year KU Leuven psychology students who participated voluntarily in a collective testing session exchange for course credit. As some students were not native Dutch speakers, an additional 41 volunteers were recruited resulting in a sample of 246 females, 158 males (mean age 20.6).

Stimuli & procedure. The stimuli consisted of each of the category exemplars for which we have similarity data from De Deyne et al. (2008). In this study, however, instead of rating pairwise similarities, participants were asked to generate as many similar words as possible to the one presented (the cue). People were shown an example using the cue word *museum* and asked to give buildings that are similar to a museum, like *gallery*, *library*, *exposition space*, *church*, *venue*, *archive*, *bank*, or *institute*. People were asked for as many similar responses that were also category members as possible and to avoid restricting themselves to only visual similarities. We also explained that some words might be harder than others, and in those cases, the number of responses they could generate might be smaller. Finally, the instructions highlighted that people could press “Unknown Word” if a cue word was unknown.

People generated responses like this for 15 random cues, each corresponding to an exemplar from a different category.[2] Thus, although any single participant saw only 15 cues, between everyone each of the exemplars served as a cue multiple times. During the experiment, the cue word was shown on top, followed by a question about other category members. For example, the question for *tiger* was "What other mammals are similar to tiger?"

Results

Data preparation. Each of the responses was spelling corrected and normalized to a canonical form. Diminutives, plurals and orthographic and dialect variants (e.g. *appelsien* and *sinaasappel*, both Dutch words for *orange*) were grouped and where possible matched to the word forms of our cues. To account for category size differences, a balanced dataset was derived in which all cues were judged by exactly 10 participants. Unknown cues, empty responses, or responses identical to the cue were removed (2% of the data). Next, we excluded responses that were not paradigmatic, including individuals (e.g., *Nemo*), incorrect responses (e.g. *jazz* as an exemplar of *Sports*) and syntagmatic responses (*tasty* as response to a *Fruit*): about 16% of the data total. The resulting exemplar dataset reveals that participants generated only a small number of neighbors, averaging 3.7 ($SD = 2.36$) responses per cue (median = 3, min = 1, max = 21). Finally, we selected for each category only those exemplar responses that also appeared as cues (retaining 52% of the exemplars).

Near neighbor graph. For each category, we derived a graph **G** based on the responses given in the NGG task. The graph was reduced to the largest strongly connected component, which means that only nodes that had both in- and out-going edges were retained. The weights of the directed edges were determined by the number of participants who generated a response for a specific cue divided by the total

[2]For technical reasons only the first 14 cues were stored during the collective session, which only impacted the total presentations.

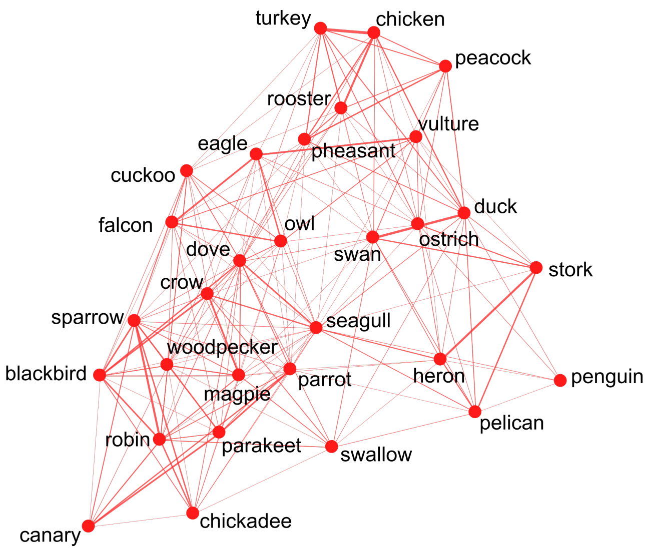

Figure 1: Example neighborhood graph for *Birds*. Line widths reflect the frequency with which each cue was generated in response to each other cue.

number of responses for that cue; that is, the weighted adjacency matrix corresponding to **G** was row-normalized to make **P**. Figure 1 shows the graph for the category *Birds*.

This graph was used to derive two different variables, described below, which we then use to predict pairwise similarity.

Direct neighbor (DN). We first derive a baseline measure that captures the purely local similarity between each exemplar in the weighted adjacency matrix in each category. For two exemplars i and j, the weighted vectors with their direct neighbors $\mathbf{P}_i$ and $\mathbf{P}_j$ yield a measure of cosine similarity:

$$cos(\mathbf{P}_i, \mathbf{P}_j) = \frac{\mathbf{P}_i \cdot \mathbf{P}_j}{\|\mathbf{P}_i\|_2 \|\mathbf{P}_j\|_2} \quad (1)$$

Spreading activation through random walks (RW). Meaning does not have to be determined solely by connections based on direct neighbors; indirect ones may also play a role. We test the utility of this sort of indirect information by deriving a measure of similarity between pairs of words based on the direct *and* indirect paths they share. For each node, this representation thus consists of a weighted sum of paths.

$$\mathbf{G_{rw}} = \sum_{r=0}^{\infty} (\alpha \mathbf{P})^r = (\mathbf{I} - \alpha \mathbf{P})^{-1}. \quad (2)$$

In this equation, r is the length of the path, $\mathbf{I}$ is the identity matrix, and α is a damping parameter that governs the extent to which similarity scores are dominated by short paths or by longer paths (Newman, 2010). Following previous work, we fix $\alpha = 0.75$ (De Deyne, Navarro, Perfors, & Storms, 2016; De Deyne, Perfors, & Navarro, 2016), although all our results are qualitatively identical if we instead use the best-fitting value. As before, we calculate semantic relatedness based on cosine similarity over this representation.

Results and Discussion. Table 1 shows for each category the correlations between the pairwise similarity judgments

Table 1: **Study 1**: Pearson r between pairwise similarity and similarity estimated based only on local paradigmatic information (direct neighbors, DN) or derived global paradigmatic information (random walks, RW). n_{ex} indicates the number of exemplars; n_r is the number of responses; k is the average number of direct neighbors (i.e. the node's out-degree), which captures the sparsity of each NN graph. The last two columns show confidence intervals (Δ CI) for the difference between the r's (RW - DN).

Category	n_{ex}	n_r	k	DN	RW	Δ CI	
Fruit	30	821	11	.51	.68	.12	.22
Vegetables	30	557	9	.31	.61	.23	.38
Birds	30	625	9	.62	.76	.10	.18
Fish	22	366	6	.70	.77	.02	.12
Insects	26	695	10	.55	.77	.16	.28
Mammals	29	513	7	.55	.78	.18	.28
Reptiles	20	536	10	.38	.67	.19	.40
Clothing	29	470	6	.51	.75	.19	.31
Kitchen utensils	31	367	5	.47	.69	.16	.28
Music instruments	27	586	8	.61	.80	.15	.24
Tools	23	195	4	.23	.46	.13	.33
Vehicles	30	548	7	.65	.83	.14	.23
Weapons	17	203	6	.48	.80	.22	.44
Sports	30	410	5	.48	.56	.02	.13
Professions	18	129	3	.26	.66	.26	.54
All	392	7021	7.1	.50	.71	.19	.22

and the two measures (DN and RW) derived from our near-neighbor graph built from paradigmatic relations only. The last line shows the results for all categories after standardizing the human similarity judgments. From here on, all correlations are significant at $p < 0.05$ (two-tailed) unless noted otherwise. Confidence intervals for the correlation difference of overlapping dependent variables for DN and RW were calculated using the procedure outlined in Zou (2007).

Several things are apparent from this. First, paradigmatic information alone is somewhat helpful in predicting similarity within categories, even when it is derived based on a relatively small number of exemplars: all correlations are significant, and most are above $r = .40$. Second, including indirect neighbors increases the correlations considerably, from an average of .50 to .71, and this difference was significant with none of the confidence intervals including zero. Although there is some variability between categories, it is clear that across the board, there is a great deal of latent, indirect structure between paradigmatically-related exemplars, and this structure is useful for predicting similarity. As a result, we find that only a very small number of neighbors (k ranging between 3 to 11) are sufficient to substantially determine the meaning of words within these categories.

Study 2: Near-neighbor Ranking

One problem with the generation task in Study 1 is that atypical exemplars tend not to be produced as responses. Since the exemplars in De Deyne et al. (2008) were chosen to cover both typical and atypical items, this meant that some items for which we have similarity data were not in the NNG dataset. As a consequence, the correlations in Table 1 may underestimate how much can be inferred from the paradigmatic network. To address this, Study 2 relied on a near-neighbor ranking task

Table 2: **Study 2**: Pearson r between pairwise similarity and similarity estimated based only on local paradigmatic information (DN) or derived global paradigmatic information (RW). n_{ex} equals the number of exemplars; n_r is the number of rankings; k is the average number of direct neighbors. The last two columns show the confidence interval (Δ CI) for the difference between the r's of RW minus DN..

Category	n_{ex}	n_r	k	DN	RW	Δ CI	
Fruit	30	900	15	.63	.80	.14	.21
Birds	30	900	15	.56	.78	.18	.27
Mammals	30	900	13	.67	.82	.11	.18
Vehicles	30	900	12	.69	.82	.10	.16
Professions	30	900	14	.49	.77	.23	.34
All	180	5400	13.8	.61	.80	.17	.21

(NNR) that asks participants to rank a set of items presented to them. It thereby avoids any limitations imposed by exemplar retrieval issues.

Participants. 81 native Dutch 1st-year economics students took part in exchange for credit (mean age 18.7 yr, 28 female).

Stimuli & procedure. Participants were given a cue word and asked to pick $m = 6$ most similar exemplars from a list of 30 exemplars of a category. As fewer subjects were available for this study, only five categories (*Birds*, *Fruit*, *Mammals*, *Vehicles* and *Professions*) were included, and each person judged four cue words from each category.

Data preparation and similarity evaluation. We excluded cues marked as "unknown", and obtained a final dataset where each exemplar was judged by exactly five participants. These data were used to construct near-neighbor graphs in exactly the same way as in the NNG task. For each cue, edge weights were calculated by counting the responses and normalized to sum to one. For each of the cues the similarity was calculated based on only direct neighbors (DN) or by including indirect paths using a random walk (RW).

The results are shown in Table 2. As can be seen from the average out-degree k, most exemplars were connected to at most half of the other exemplars in the category. As before, global similarity (RW) leads to a considerable improvement over the baseline based only on direct neighbors (DN). For each category, the correlation differences between DN and RW had confidence intervals excluding zero. The achieved RW correlations over all categories of around .80 are fairly impressive considering that they are only slightly smaller than the Spearman split-half reliability of the pairwise similarity ratings themselves (.87 *Fruit*; .90 *Birds*; .92 *Mammals*; .96 *Vehicles*; .91 *Professions*).

The role of data quantity One way to estimate the informativeness of this kind of paradigmatic information is to evaluate how the correlations to pairwise similarity judgment change when fewer neighbors are used. As Table 2 makes clear, using $m = 6$ ranked associates for each cue leads to a graph with 12 to 15 direct neighbors k of each node. What happens if people ranked different numbers of m exemplars?

The answer to this question, shown in Figure 2, reveals that

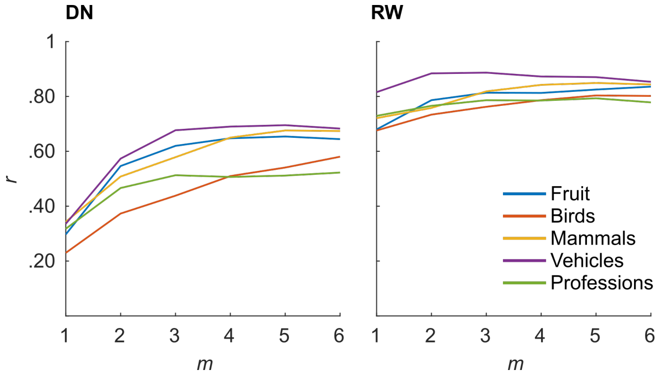

Figure 2: The effect of number of ranked responses m on correlations to pairwise similarity judgments. Both direct (DN) and indirect neighbor measures (RW) based on paradigmatic information predict reasonably well even with high data sparsity (low m). This is particularly apparent when the graph incorporates indirect links (RW).

correlations to pairwise similarity are reasonably strong even when they are based on relatively few ranked responses m. The effect is especially striking for the information based on global category structure: RW with only the first-ranked response outperforms DN with all six (Zou's test for correlation differences using all six categories: $\Delta r(2175) = .05, CI = [.03, .08]$). Given that the RW graph derived from $m = 1$ was based on only 17% of the data as the DN graph derived from $m = 6$ (i.e., 150 vs 900 total judgments), the fact that RW outperformed DN is an impressive testament to the power of inferred indirect associational information.

Study 3: Syntagmatic information

In Studies 1 and 2, we used the near neighbor task to construct semantic networks comprised solely of paradigmatic relations. Despite the extreme sparsity of these networks, we find that paradigmatic information is strongly correlated with similarity judgments. In Study 3 we take a complementary perspective, constructing networks based solely on syntagmatic relations. To do so, we rely on a large-scale semantic network constructed from Dutch word association data (see De Deyne, Navarro, & Storms, 2013). The complete data set includes 12,000 cue words and over 3 million responses in total. Using these data, we construct a semantic network that links words based on their associations, using a positive point-wise mutual information (PPMI) method to account for frequency biases in the word association task (see De Deyne et al., 2016). As with Studies 1 and 2, we consider two network measures of similarity, a direct-neighbor (DN) measure and random-walk (RW) measure. The included exemplars were matched closely to those included in Study 1 (see column 2, Table 1).

As typically constructed, semantic networks based on word association data incorporate both the syntagmatic and paradigmatic relations that underpin semantic representation. The approach we take here, however, uses the data from Study 1 supplemented with common variants (plurals, diminutives, alternate spellings, etc.) to remove as much of the paradigmatic information as possible. To that end, any edge that appeared

Table 3: **Study 3**: Pearson r between judged similarity and similarity estimates from a large-scale syntagmatic network. As before, we compare local information (DN) with indirect edges (RW). n_{ex} is the number of category exemplars and k is the number of direct neighbors. The last two columns show the confidence interval (Δ CI) for the difference (RW - DN).

Category	n_{ex}	k	DN	RW	Δ CI	
Fruit	30	62	.65	.58	-.11	-.05
Vegetables	30	65	.42	.31	-.16	-.06
Birds	30	67	.59	.59	-.03	.02
Fish	22	62	.65	.59	-.10	-.02
Insects	26	75	.78	.72	-.10	-.04
Mammals	30	69	.51	.52	-.02	.04
Reptiles	20	71	.75	.70	-.08	-.02
Clothing	29	78	.56	.53	-.06	.00
Kitchen utensils	31	72	.58	.52	-.11	-.02
Music instruments	27	70	.53	.35	-.22	-.14
Tools	24	82	.44	.43	-.04	.02
Vehicles	30	81	.74	.73	-.03	.01
Weapons	17	74	.77	.71	-.10	-.02
Sports	30	72	.73	.72	-.03	.01
Professions	18	80	.74	.79	.01	.09
All	394	72.0	.61	.57	-.06	-.04

in the paradigmatic networks from Study 1 was removed from the networks in Study 3, and the resulting networks can be assumed to be largely syntagmatic in nature. Doing so removed only a small amount of information: averaged over categories, on average 9 out of 72 edges were paradigmatic coordinates, which corresponds to 12% of all responses.

Results and Discussion

As shown in Table 3, networks based solely on syntagmatic information do provide a reasonable account of similarity judgments. However, despite the fact that these networks are based on a vastly larger data set, the correlations are *weaker* than those found in Studies 1 and 2 once indirect paths are considered. This suggests that even a small amount of paradigmatic information may provide as much semantic knowledge as a much larger quantity of syntagmatic information.

One interesting aspect of these results is that, in marked contrast to Study 1 and Study 2, the results for the DN baseline were somewhat better or at least similar to those from the RW measure that incorporated indirect paths. The only case in which the correlations were statistically higher was professions where the correlation difference confidence interval was positive (see column 6 and 7 in Table 3). This replicates previous findings that show a limit to the contribution of indirect paths within basic-level categories (De Deyne et al., 2016). In combination with the results from the previous two studies here, our findings suggest that there is a lot of indirect paradigmatic information within categories, but – at least with respect to encoding semantic knowledge – not a lot of indirect syntagmatic information within categories.

General Discussion

Understanding how the mind rapidly acquires and efficiently represents a massive amount of semantic knowledge is a fundamental question in cognitive science. Even in a small set of only 30 concepts there are 870 pairwise similarity relations that need to be encoded, a quantity that scales quadratically with the number of lexical entries. In this paper we find that most of this information can be encoded very efficiently using only a small number of links (about 2–3) per node in a semantic network of paradigmatic relations. The results extend previous work looking at how people represent the similarities between very dissimilar concepts (De Deyne et al., 2016). It demonstrated that spreading activation mechanisms over very sparse networks can capture a remarkable amount of the shared meaning between words that are not directly linked. Interestingly, however, a much larger network of syntagmatic relations does a much poorer job of encoding these similarity relations, suggesting that a relatively small part of the semantic network does most of the work in encoding this knowledge.

The results open up a number of questions about how semantic knowledge is encoded. First, in our studies the superior performance of paradigmatic relations only holds when indirect paths (i.e., RW) between words are included. When looking only at direct relationships (i.e., DN) the syntagmatic network performs comparably or even better than the paradigmatic network, albeit on the basis of a much larger training data set. A little bit of paradigmatic knowledge goes a long way, but only so long as it is combined with mechanisms that can exploit the structure of this knowledge.

Second, the poor performance of the random walk model over the syntagmatic network does not imply that indirect syntagmatic paths play no role: a ceiling level may potentially already be hit when a word has about 72 outgoing syntagmatic edges (see Table 3). However, this possibility seems unlikely: even when we re-ran the analyses from Study 3 using a much sparser version of the syntagmatic network (one that included only one third of the responses from the original data set), we still found that the random walk model did not improve the performance of the syntagmatic associations ($r = .53$ for DN, $r = .52$ for RW). More generally, in no version of our simulations did we find syntagmatic networks performing at the same level as the random walk model defined over the much sparser paradigmatic networks.

A third issue pertains to the trade-off between computation and storage. One way of contrasting spreading activation models with memory storage models comes from work on mediated priming with items like *lion* – (*tiger*) – *stripes* where storage accounts like the compound-cue theory have argued for a direct route between cue and target (McKoon & Ratcliff, 1992). A similar case could be made here: perhaps the mind actually stores a large number of direct paradigmatic relationships in long term memory, and the evidence for indirect (RW) effects is simply an artifact of methodological limitations. This explanation might be plausible with respect to Study 1 (where participants might have been limited in the number of responses they could retrieve). However, in Study 2 many of these limitations were relaxed, and it was still the case that the paradigmatic network was very sparse relative to the syntagmatic network. Indeed, in order to "censor" the word association network in Study 3, only a very small number of paradigmatic relations needed to be removed. Taken together, these considerations suggest that the paradigmatic links genuinely do much of the work in representing meaning.

Finally, there are also theoretical reasons why words can only be similar to a small number of other words. For example, similarity is often assumed to be exponentially decaying and only a small number of words can be near each other in high dimensional spaces (Tversky & Hutchinson, 1986). It is therefore perhaps to be expected that a paradigmatic network should be extremely sparse, with similarity relations encoded via an inferential process (like spreading activation) defined over that representation.

Future work

This paper only evaluated paradigmatic and syntagmatic inference using pairwise judged similarity. Mapping and inferring paradigmatic relations using NNG and NNR tasks might also provide a better account of other semantic tasks such as priming, or category induction. It might also account for both facilitatory and inhibitory effects in word processing depending on the density of semantic neighborhoods (Pexman, Hargreaves, Siakaluk, Bodner, & Pope, 2008). Our findings do suggest that generating near neighbors might be a very efficient context-independent way to assess semantic knowledge for large domains – more efficient than the data-intensive process of measuring pairwise similarity. As a result, the NNG and NNR tasks may be practical and tractable enough to acquire sufficient data to investigate individual differences in the mental lexicon.

That said, the current work still represents a first step and therefore has several limitations. First, because participants in Study 1 and Study 2 were opportunity samples, it is possible that any differences between them exist because of henceforth unnoticed differences between psychology and economics students. That said, these differences are likely to be minor considering the validity of the best-performing measures across both studies in predicting similarity ratings collected 10 years ago (De Deyne et al., 2008). Second, although the near neighbor generation task is highly efficient and complements other procedures like word association or semantic feature generation, it still has potential to be refined further. For example, in Study 1, despite explicit instructions to only produce category members, some participants provided other kinds of associations. Perhaps improved instructions or additional practice would improve the efficiency and accuracy of the data generated in these tasks. Finally, the current studies have only analyzed similarity to the category exemplars in De Deyne et al. (2008), which of course represents only a subset of potential category members. We expect that adding more commonly known exemplars could further improve estimates of similarity.

Conclusions

Altogether, this work makes two important contributions. First, it shows how word meaning in semantic categories can efficiently be approximated based on paradigmatic information acquired from a near-neighbor task. The efficiency is derived from the fact that a great deal of structure can be extracted from indirect neighbors as we did when we modeled global graph similarity. This might potentially lead to more efficient procedures to approximate the mental lexicon based on a single or small number of individuals.

Second, we show that it is theoretically possible to infer a substantial portion of the meaning of a word based on only a small amount of data encoded in language. To do so, learners must be attuned to paradigmatic relations and capable of finding the indirect structure between them. This finding has implications for how lexico-semantic models might solve Plato's problem: perhaps the heavy lifting is accomplished with a relatively very small portion of the data and specific kinds of relations.

Acknowledgments

ARC grants DE140101749 and DP150103280 supported SDD.

References

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*, 407–428.

De Deyne, S., Navarro, D. J., Perfors, A., & Storms, G. (2016). Structure at every scale: A semantic network account of the similarities between unrelated concepts. *Journal of Experimental Psychology: General*, *145*, 1228-1254.

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., Perfors, A., & Navarro, D. J. (2016). Predicting human similarity judgments with distributional models: The value of word associations. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics* (pp. 1861–1870). Osaka, Japan.

De Deyne, S., Verheyen, S., Ameel, E., Vanpaemel, W., Dry, M., Voorspoels, W., & Storms, G. (2008). Exemplar by Feature Applicability Matrices and other Dutch Normative Data for Semantic Concepts. *Behavior Research Methods*, *40*, 1030–1048.

Ervin, S. M. (1961). Changes with age in the verbal determinants of word-association. *The American Journal of Psychology*, *74*(3), 361–372.

Lenci, A., Baroni, M., Cazzolli, G., & Marotta, G. (2013). BLIND: a set of semantic feature norms from the congenitally blind. *Behavior Research Methods*, *45*, 1218–1233.

Louwerse, M. M., & Connell, L. (2011). A taste of words: Linguistic context and perceptual simulation predict the modality of words. *Cognitive Science*, *35*, 381–398.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1155.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press, Inc.

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*, 161–167.

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

Sloutsky, V. M., Yim, H., Yao, X., & Dennis, S. (2017). An associative account of the development of word learning. *Cognitive Psychology*, *97*, 1–30.

Tversky, A., & Hutchinson, J. (1986). Nearest neighbor analysis of psychological spaces. *Psychological review*, *93*, 3-22.

Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, *12*, 399–413.