Untangling Semantic Similarity: Modeling Lexical Processing Experiments with Distributional Semantic Models.

Farhan Samir
Department of Computer Science
University of Toronto
(fsamir@cs.toronto.edu)

Barend Beekhuizen
Department of Language Studies
University of Toronto, Mississauga
(barend.beekhuizen@utoronto.ca)

Suzanne Stevenson
Department of Computer Science
University of Toronto
(suzanne@cs.toronto.edu)

Abstract

Distributional semantic models (DSMs) are substantially varied in the types of semantic similarity that they output. Despite this high variance, the different types of similarity are often conflated as a monolithic concept in models of behavioural data. We apply the insight that word2vec's representations can be used for capturing both paradigmatic similarity (substitutability) and syntagmatic similarity (co-occurrence) to two sets of experimental findings (semantic priming and the effect of semantic neighbourhood density) that have previously been modeled with monolithic conceptions of DSM-based semantic similarity. Using paradigmatic and syntagmatic similarity based on word2vec, we show that for some tasks and types of items the two types of similarity play complementary explanatory roles, whereas for others, only syntagmatic similarity seems to matter. These findings remind us that it is important to develop more precise accounts of what we believe our DSMs represent, and provide us with novel perspectives on established behavioural patterns.

Keywords: semantic similarity; priming; word2vec; distributional semantics; semantic neighbourhood density.

Introduction

The cognitive processes and representations involved in word meaning are a central area of interest for cognitive science. A successful starting point for modeling these phenomena computationally is the distributional hypothesis (Firth, 1957): the idea that a word's linguistic contexts are highly informative of its meaning. Distributional semantic models (DSMs) operationalize this idea by representing the meaning of a word as a vector induced from the set of contexts the word occurs in. Such models have widespread use for modeling psycholinguistic phenomena regarding meaning.

One such phenomenon that has been studied extensively using DSMs is semantic priming, in which the prior presentation of a 'prime' word facilitates the processing of a related 'target' word (Meyer & Schvaneveldt, 1971). When studied at the level of individual prime–target pairs, the semantic similarity between prime and target words has been found to be an explanatory factor of the variance in reaction times (Ettinger & Linzen, 2016; Mandera et al., 2017). However, these item-level studies treat semantic similarity as a monolithic relation, a position that is at odds with much previous work (e.g. Lund et al., 1995; Jones et al., 2006; Günther et al., 2016) that proposes distinctions between semantic similarity in a narrow sense (featural overlap) and associative similarity (between-word associations based on words' conceptual meaning or distributional co-occurrence). In these studies, DSMs thought to instantiate one but not the other kind of similarity have been found to explain different priming effects on an aggregate level (as opposed to an item level).

The underlying conception of semantic versus associative similarity, however, has been disputed (Ettinger & Linzen, 2016; McRae et al., 2012), as has one of the main ways in which it is operationalized (Günther et al., 2016). In this paper, we instead follow another distinction, based on distributional properties (e.g. Schütze & Pedersen, 1993), namely, that between syntagmatically related words (words that occur in each other's near proximity, such as *drink–coffee*), and paradigmatically related words (words that can be substituted for each other such as *book–novel*). We propose that word2vec (Mikolov et al., 2013), a model that has been used to produce monolithic 'semantic similarity' scores in item-level semantic priming studies like Ettinger & Linzen (2016) and Mandera et al. (2017), can isolate the syntagmatic and paradigmatic similarity of word pairs by using its representations in ways most applications of this model do not do.

We will demonstrate that effects of semantic similarity presented as a holistic concept can be broken down into paradigmatic and syntagmatic similarity, where the relative contributions of each depends on the nature of the task and stimuli. We do so, first, by modeling the facilitatory effect of each of these kinds of word similarity on semantic priming. After validating the intuition of our construct, we show that the relative explanatory importance of the two kinds of similarities differs across stimulus types, but that syntagmatic similarity is an overall more reliable factor.

To show the potential of our approach in modeling processing effects beyond semantic priming, we also look at another semantic factor that has been shown to facilitate word recognition, namely semantic neighbourhood density (the degree of similarity of a word to its nearest neighbours in a DSM, cf. Buchanan et al., 2001). By separately considering nearest neighbours according to paradigmatic similarity vs. syntagmatic similarity, we show that these two measures do not equally contribute to the explanation of the reported effects.

These two case studies (1) remind us that it is important to develop accounts of the kinds of constructs we believe our DSM-based measures are representing, and (2) provide us with novel ways of looking at experimental results, thus motivating further inquiry into the semantic representations and processes involved in lexical processing.

Our Approach

Distributional Semantic Model. The CBOW algorithm of word2vec (Mikolov et al., 2013) is unusual in providing a straightforward way to obtain word vectors that encode both paradigmatic and syntagmatic similarity. This is due to its optimization process that yields *two* distinct sets of vectors, within the same semantic space. We can then consider similarity of word representations both *within* one set of vectors, or *across* the two sets of vectors, as we explain here.[1]

At a high level, word2vec optimizes a one-hidden layer neural network to predict a target word (on the output layer) on the basis of the words in its context (on the input layer). This yields two sets of word vectors (of the same dimension): those encoded in the columns of the input matrix (between the input layer and the hidden layer), and those encoded in the rows of the output matrix (between the hidden layer and the output layer). In what follows, we refer to the input vector for a word w as v^I_w and the output vector as v^O_w. The optimization procedure is designed to bring the representation of context words in the input matrix close (in the same vector space) to the representation of their target words in the output matrix. That is, the cosine similarity for $v^I_{w_c}$ and $v^O_{w_t}$ will be high if w_c is a frequent context word of w_t.

Although most uses of word2vec only use input vectors for performing tasks such as analogical reasoning (Mikolov et al., 2013), using both sets of vectors can have advantages (Levy et al., 2015; Roller & Erk, 2016; Beekhuizen et al., 2019). In this work, we build on the insight of Levy et al. (2015) that using both sets of word2vec vectors can yield two measures of similarity, which separately reflect syntagmatic similarity and paradigmatic similarity. Following Grefenstette (1994), we refer to the two measures of similarity as first-order similarity and second-order similarity, respectively.[2]

First-order similarity (FOS). A measure of syntagmatic similarity should be high for words that frequently co-occur with each other, such as phrasal associates (*drink – coffee*, *book – cover*). Such pairs will frequently occur as (context, target) pairs in the training corpus of word2vec and will therefore be predictive of each other. Because of this we expect that word2vec's training procedure will maximize their similarity *across* the two matrices. Suggestive evidence for this intuition comes from Levy et al. (2015) who provide a proof that the inner-product of vectors across the two matrices in the SkipGram variant of word2vec approximates measures of word co-occurrence. Given this reasoning, we calculate the first-order similarity (FOS) between two words (w_1, w_2) as follows :

$$FOS(w_1,w_2) = cos(v^I_{w_1}, v^O_{w_2}) = \frac{v^I_{w_1} \cdot v^O_{w_2}}{\|v^I_{w_1}\|\|v^O_{w_2}\|}.$$

word	metric	nearest neighbours
school	FOS	med, high, grad, graduate, elementary, at, boarding, graduated, reform, dropout
	SOS	harvard, yale, gym, nyu, class, stanford, academy, lawndale, university, princeton
book	FOS	comic, read, wrote, write, writing, written, reading, published, poems, novel
	SOS	diary, novel, script, journal, poem, bible, newspaper, manuscript, column, poems
theory	FOS	quantum, based, basic, physics, psychological, relativity, overriding, chemistry, scientific, sexual
	SOS	strategy, method, concept, phenomenon, character, personality, perception, technique, flaw, knowledge
excited	FOS	very, about, so, getting, too, pretty, terribly, extremely, 'm, awfully
	SOS	upset, uptight, nervous, depressed, anxious, worried, touchy, thrilled, annoyed, embarassed

Table 1: Ten nearest neighbours for four target words, based on First-order (FOS) and Second-order similarity (SOS).

Second-order similarity (SOS). A measure of paradigmatic similarity should be high for two words that are structurally substitutable (cf. Lyons, 1977, 241), such as near-synonyms (*book – novel*), antonyms (*high – low*), co-hyponyms (*adjective – adverb*), or hyponym–hypernym pairs (*Africa – continent*). Each of the two should occur in similar contexts, without necessarily appearing in the vicinity of one another (Schütze & Pedersen, 1993; Grefenstette, 1994). Considering word2vec, this means that these words will be the targets of (many of) the same context words, and will be the contexts of (many of) the same target words. We can then expect the word2vec training algorithm to learn similar representations of these two words *within* each of the input and output matrices. Focusing on the input matrices here (but noting that analogous results are obtained when looking at the output matrices instead), we define the second-order similarity (SOS) between two words (w_1, w_2) as follows:

$$SOS(w_1,w_2) = cos(v^I_{w_1}, v^I_{w_2}) = \frac{v^I_{w_1} \cdot v^I_{w_2}}{\|v^I_{w_1}\|\|v^I_{w_2}\|}.$$

First exploration. Table 1 illustrates our intuition that FOS captures syntagmatic and SOS captures paradigmatic relations between words. Neighbours with high FOS are frequent collocates of the target word (e.g., *elementary* and *dropout* for

[1] For our experiments, we train the CBOW algorithm of word2vec on the SUBTLEX$_{\text{US}}$ corpus (Brysbaert & New, 2009). We used the implementation of the algorithm in gensim (Řehůřek & Sojka, 2010) with 15 negative samples, a window size of 10, and default settings for all other hyperparameters.

[2] Strictly speaking, Grefenstette (1994) denotes the strength of word relations as *affinities* rather than *similarities*. Others denote the relations as *co-occurrences* (Jurafsky & Martin, 2014) or *associations* (Jones et al., 2006). For reasons of simplicity, alignment with most common usage in distributional semantics, and because both measures are derived from the cosine similarity, we use the term *similarity*.

school) whereas neighbours with high SOS are substitutable words (e.g., *Harvard* or *academy* for *school*).

Previous works have recognized the separation of these two categories of relations (Günther et al., 2016; Jones et al., 2006), and have used HAL and LSA to model paradigmatic and syntagmatic relations, respectively. However, Günther et al. (2016) report a correlation of $r \sim .90$ between the two measures which raises the question of whether these models are truly capturing independent relations. Conversely, our computational measures of the two relations have a correlation of $r = 0.36$ ($p < 0.005$) providing initial validation that our measures are capturing distinct constructs.

Similarity Effects in Semantic Priming

The degree of semantic similarity between a prime–target pair in a DSM has been shown to factor into item-level regression models of semantic priming (Günther et al., 2016; Mandera et al., 2017; Ettinger & Linzen, 2016). However, such works rarely elucidate the notion of similarity they claim to tap into. Here, we study how our two types of semantic similarity contribute to the explanation of semantic priming effects. We hypothesize that our measures will capture different yet complementary types of semantic similarity, because FOS will capture similarity in cases of priming of frequently co-occurring words (cf. Hutchison et al., 2013), and SOS will do so for cases of priming of semantically related words that do not co-occur (cf. McRae & Boisvert, 1998). We test our hypothesis on the Semantic Priming Project dataset (SPP, Hutchison et al., 2013), a collection of 6,644 prime–target pairs, with their reaction time (RT) in a primed lexical decision task (LDT).

Sensitivity of our Measures

Before looking into the online measures of priming, we use the SPP to validate that our two similarity scores capture the intended concepts. Prime–target pairs in the SPP are divided into 12 categories of relations, which we group into the two overarching categories of paradigmatic and syntagmatic relations. The paradigmatic relations that we use are: synonymy, antonymy, superordination (i.e. hypernymy), and category coordination (i.e. co-hyponymy). The syntagmatic relations are forward-phrasal association and backward-phrasal association. The remaining 6 categories (perceptual property, functional property, instrument, script, unassociated, action) are not easily classified, so we exclude them from this analysis.

If SOS captures paradigmatic relations and FOS syntagmatic ones, we expect paradigmatic prime–target pairs to be more similar in SOS than in FOS, and syntagmatic ones more similar in FOS than in SOS. To test this, we look at the relative similarity between a prime and its target by ranking all 1,661 target words (in the Semantic Priming Project) by similarity to the prime word and comparing the rank of the actual target given FOS to the rank of the actual target given SOS. We expect pairs in paradigmatic categories ($n = 1397$) to display lower ranks (higher similarities) for the target in SOS-based rankings than FOS-based ones, and vice versa for pairs in syntagmatic categories ($n = 562$). For each of the six aforementioned categories, we plot the average ranking for every target given its prime with both FOS and SOS.

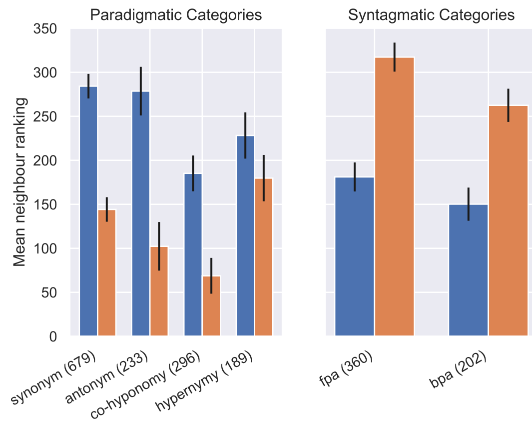

Figure 1: Average rank of targets for primes (lower is better) in each category [with number of items], using FOS vs. SOS. Blue bars represent rankings by FOS and orange bars represent SOS. One-sided paired t-tests show significant differences in the predicted direction between FOS and SOS ranks for all categories (for hypernymy: $p < .01$, for all others: $p < .001$).

Fig. 1 confirms this intuition: for each of the relations, the difference between rankings based on FOS and those based on SOS is significant in the predicted direction. This pattern validates our intuition about the FOS and SOS measures.

Modeling Semantic Priming Project Data

Next, we test our hypothesis by fitting linear regression models to primed RT for all of the (prime, target) pairs in the SPP data, using our two measures of semantic similarity as the predictive variables of interest. We expect that more related (prime, target) pairs should be judged as more similar, by FOS and SOS, than less (or un-) related pairs, and predict a lower primed RT – i.e., the priming effect. As in Mandera et al. (2017); Günther et al. (2016), but in contrast to Hutchison et al. (2008), we carry out linear regressions on the primed RT, rather than the difference between the primed and unprimed RT, to ensure that our measures are sufficiently sensitive to capture the priming effect (or lack thereof, for unrelated pairs) on an itemwise basis.

In our regressions here, we address several shortcomings of previous work using DSM similarity measures on the SPP (Günther et al., 2016; Ettinger & Linzen, 2016), including those using word2vec (Mandera et al., 2017). First, we use both FOS and SOS measures derived from our DSM, which enables a finer-grained exploration of the role of different types of similarity. Second, we include non-primed lexical decision RT of the target as a covariate because it captures influences on primed RT that do not pertain to the relation between the prime and the target. Whereas others have subtracted the unprimed RT from the primed RT and used the

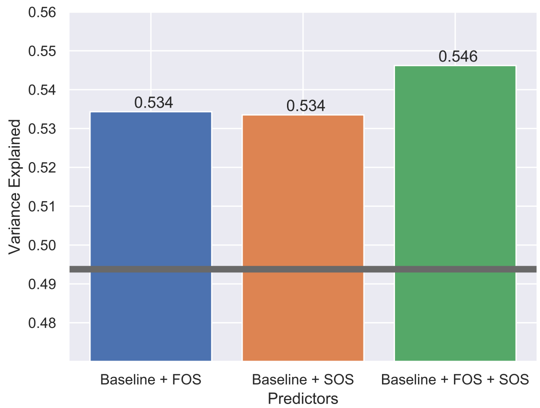

Figure 2: Performance of three linear regression models over the baseline model (grey line) on the SPP.

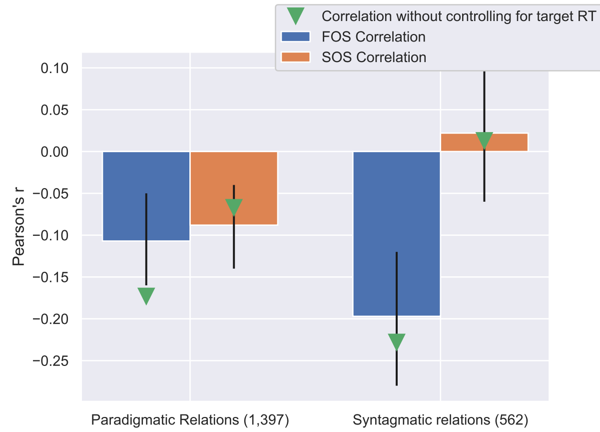

Figure 3: Partial correlation of semantic similarity measures (FOS and SOS) with RTs for paradigmatically and syntagmatically related words. $p < 0.001$ in all categories except SOS correlation with syntagmatically-related pairs. See footnote 4 for discussion of the green triangles.

difference as the dependent variable (Hutchison et al., 2008), or ignored the unprimed RT entirely (Mandera et al., 2017; Günther et al., 2016; Ettinger & Linzen, 2016), we use unprimed RT as a covariate.

Setup. We conduct four regression analyses over all the pairs in the SPP in modelling LDT primed RT, averaging over the two stimulus onset asynchronies of 200ms and 1200ms. As a baseline model, we include the lexical characteristics of both target and prime (length, orthographic neighbourhood density and log-frequency in SUBTLEX$_{\text{US}}$), and the unprimed RT (from the Semantic Priming Project). The second and third analyses include one of FOS and SOS, respectively, while the fourth includes both variables as predictors. We exclude any item that has missing data for any of the independent variables, leaving us with 6,477 (out of 6,644) items.

Results & Discussion. Fig. 2 shows that adding either measure of similarity, FOS or SOS, results in a clear improvement of R^2 over the baseline. Further, adding both measures – i.e., considering both paradigmatic and syntagmatic similarity – results in a further small but significant increase in R^2 to .546 $(t_{6,466} = -13.45, p < 0.001)$ over the models with only one of the DSM measures.

It is apparent from Fig. 2 that syntagmatic similarity in word2vec is a significant factor. Using only paradigmatic similarity in this DSM, as Mandera et al. (2017); Ettinger & Linzen (2016) do, means missing out on a significant source of explanation of behavioural effects. Since FOS and SOS are representative of syntagmatic and paradigmatic similarity, we conclude that semantic priming is facilitated by both substitutability of the prime-target as well as their co-occurrence.

Analysis on Related Pairs

Having shown that both FOS and SOS are predictive of priming effects across related and unrelated pairs, we next raise the question of whether the two measures capture different kinds of priming effects in the data. In particular, we ask whether FOS is a better predictor of the priming effect on syntagmatically related pairs, and SOS is a better predictor of the priming effect on paradigmatically related pairs.

Set-up. In these regressions, we use the same split of the SPP prime–target pairs into syntagmatically related and paradigmatically related pairs as we did in our earlier ranking analysis. For both categories, we calculated two partial correlations: between primed RT and FOS, and between primed RT and SOS. We use the same covariates as above (length, orthographic neighbourhood density, log-frequency, and unprimed RT). We also partial out the the forward association strength and backward association strength of each pair; these are two common covariates used in semantic priming research because of their known effect on the priming effect (Hutchison et al., 2008; Mandera et al., 2017).[3]

After partialing out these variables, we investigate the unique contribution of each of our two measures of similarity. The unique effect of SOS is determined by first partialing out the FOS and then determining the correlation with the remaining variance, and vice versa. We run each of these analyses per group of words: once for paradigmatically related words (where we expect SOS to be the more powerful predictor) and once for the syntagmatically related words (where we expect FOS to be the more powerful predictor).

Results & Discussion. Fig. 3 shows that, as expected, only FOS has a significant partial correlation with primed RT $(r = -0.197, p < 0.001)$ for syntagmatically related words, and SOS does not $(r = 0.022, p = 0.6)$. On the other hand, both FOS $(r = -0.107, p < 0.001)$ and SOS $(r = -0.08, p < 0.001)$ provide significant unique contributions to the predic-

[3] The forward association strength refers to the proportion of subjects that responded with the target word given the prime word as a cue in the Nelson et al. (2004) association norms. Conversely, the backward association strength refers to the proportion of subjects that responded with the prime word given the target word as a cue.

tion of primed RT for paradigmatically related words. That is, both degree of substitutability and degree of co-occurrence contribute to an explanation of primed lexical processing time for paradigmatically related words. Presumably, the processing time of *fork* primed by *spoon* is facilitated by the high substitutability of the words as well as their frequent co-occurrence (Justeson & Katz, 1991). Conversely, the processing time of *fork* primed by *eat* is only facilitated by their frequent co-occurrence.[4]

This result is interesting because it contrasts with other analyses of priming with distributional models, which have suggested that measures of paradigmatic similarity are important in capturing priming effects of paradigmatic relations while measures of syntagmatic similarity are not (Jones et al., 2006; Günther et al., 2016).

Semantic Neighbourhood Density Effects

In our second analysis, we apply our approach of separating semantic similarity into paradigmatic and syntagmatic similarity to the study of another semantic effect in lexical processing. Buchanan, Westbury, & Burgess (2001) introduce the concept of semantic neighbourhood density, measured by the mean distance between the target word and its 10 nearest neighbours in a DSM (i.e., its semantic neighbourhood). Using HAL (Lund et al., 1995) as their DSM, they show that words with a high semantic neighbourhood density are processed faster than words with a low semantic neighbourhood density, all else being equal. The rationale for this effect is that words with a higher density have a highly concentrated area of spreading activation, resulting in quicker processing.

Here, we explore the nature of the semantic neighbourhood effect further by asking what kind of semantic similarity drives it. We first define paradigmatic semantic neighbourhood density (PSND) as the mean SOS between a target word and its 10 most SOS-similar neighbours. This score indicates how many highly substitutable words there are for some target word. Next, we define syntagmatic semantic neighbourhood density (SSND) as the mean FOS between a target word and its 10 most FOS-similar neighbours. This indicates how many frequently co-occurring words a target word has. These two indices reflect different views on the nature of such semantic neighbourhoods, with PSND representing a word having many substitutes, and SSND representing a word having many transitional possibilities.

We hypothesize that both PSND and SSND facilitate processing in an LDT, as tasks such as free association (Nelson et al., 2004) indicate that cue words can activate both paradigmatically and syntagmatically related words, and that high degrees of activation of either (as measured through neighbourhood density) can be expected to facilitate lexical processing.

[4] As Fig. 3 shows, not controlling for the effect of the unprimed target RT results in a considerable increase of the the effect of FOS $r = -0.107$ to $r = -0.173$. Controlling for the unprimed target RT is crucial, lest the effects of DSM similarity be overstated.

Setup. We use RTs for words in the (non-primed) visual lexical decision task (LDT) of the SPP. In order to isolate any potential influence of PSND and SSND from well-understood low-level effects, we partial out the effects of word length, log frequency, and orthographic neighbourhood size and keep the residual variance. On this residual variance, we study the unique explanatory role of PSND by first partialing out SSND, and that of SSND by first partialing out PSND. We perform the analysis on the 3,814 words in the SPP that have values for the aforementioned properties.

Results & Discussion. Partialing out the effect of PSND, we find that SSND is significantly negatively correlated with RT ($r = -.23, p < 0.001$), meaning that the higher the average FOS is between a target word and its 10 nearest neighbours, the lower its reaction time in visual lexical decision, in line with the findings of Buchanan et al. (2001). However, when we partial out the effect of SSND, we do not find a significant correlation ($r = -.01, p = 0.35$). This result suggests that in the single-word lexical decision task, a word that has a relatively high level of FOS with its closest neighbours, such as *school* (having a standardized SSND of 1.11) , will be recognized more quickly than words that do not (e.g., *theory*, having a standardized SSND of only 0.33 to its nearest neighbours; cf. Table 1 for examples of both words).

Thus, there appears to be a clear effect of frequent co-occurrence with other words in facilitation of word recognition in an LDT. Meanwhile, substitutability does not seem to have an independent contribution to the explanation of the density effect. Whereas the approach of Buchanan et al. (2001) does not, and in fact cannot, consider these alternatives separately, the architecture of CBOW and our similarity measures defined over it provide a way to study the effects of different kinds of semantic similarity in isolation.

General Discussion

Semantic similarity in distributed semantic models (DSMs) is frequently invoked as an explanation in lexical processing studies. However, such similarity is often taken to be a monolithic concept (Ettinger & Linzen, 2016; Mandera et al., 2017), in contrast to prior work demonstrating the relevance of distinguishing different *kinds* of similarity (Lund et al., 1995; Jones et al., 2006; Günther et al., 2016). We propose a novel way of using the representational potential of the successful word2vec DSM (Mikolov et al., 2013), used elsewhere as a source of monolithic semantic similarity (Mandera et al., 2017; Ettinger & Linzen, 2016), to capture a well-known distinction between syntagmatic, or first-order similarity (frequent co-occurrence in a narrow context), and paradigmatic, or second-order similarity (substitutability) (Schütze & Pedersen, 1993; Grefenstette, 1994). Following Levy et al. (2015), we use word2vec's representation of each word in two distinct matrices to operationalize first-order similarity in word2vec as between-matrix similarity, and second-order similarity as within-matrix similarity (within the input matrix).

In our first experiment, we show that both first- and second-order similarity uniquely predict variance on a primed lexical decision task on the data of the Semantic Priming Project (Hutchison et al., 2013). We furthermore find that, when splitting prime–target pairs into syntagmatically related ones and paradigmatically related ones, an asymmetry emerges: whereas both first-order similarity and second-order similarity are unique predictors of primed RT for paradigmatically related pairs, only first-order, but not second-order similarity, uniquely predicts syntagmatically related pairs.

Next, we generalize our approach by looking at the facilitatory effect on unprimed lexical decision that Buchanan et al. (2001) attribute to semantic neighbourhood density (the average similarity between a target word and its ten nearest neighbours in a DSM). Calculating semantic neighbourhood density based on first-order similarity vs. second-order similarity shows that it is the former that uniquely explains the effect Buchanan et al. found.

Our analysis illustrates the importance of considering the type of semantic similarity a DSM measure is sensitive to and the necessity to consider more than one type of semantic similarity to explain behavioural results. While such issues have previously been raised for other models (Jones et al., 2006; Günther et al., 2016), we demonstrate their relevance for word2vec as well: the most frequently used semantic similarity measure in word2vec arguably better captures word substitutability than frame or association-based similarity.

Whereas the modeling study for priming shows that both kinds of similarity play a role, the asymmetry between first-order similarity and second-order similarity is striking. While the dominance of first-order similarity may be due to task or design effects in the experimental set-up, it raises an interesting question for follow-up studies, namely whether knowledge of simple co-occurrence patterns is the true cause of (some) semantic priming effects and semantic neighbourhood density effects. That is: should we, in these cases, attribute the observed effects to the processing expectations associated with words rather than patterns of substitutability?

A further important topic for future research is understanding the relation between the similarities in the various DSMs better. A fuller comparison of the various DSMs and their variants (e.g., LSA, HAL, BEAGLE) on the same sets of data would be a first step in the direction of a fuller understanding of how distributional properties correspond to psycholinguistic constructs of word meaning.

Additionally, future research should be directed at comparing and combining DSM-based approaches discussed in this work with graph-theoretical approaches in investigating lexical processing phenomena (Kenett et al., 2017; Steyvers & Tenenbaum, 2005). Previous work has opted for constructing semantic graphs using results from word association tasks instead of DSMs, partly motivated by observations that DSMs do not model syntagmatic relations such as *red – roses* well (Kenett, 2019). As shown in this work, such a limitation is not present with the computational measures of FOS and SOS derived from the word2vec model.

Acknowledgments

FS and SS are supported by an NSERC Discovery Grant RGPIN-2017-06506 to SS, and FS and BB are supported by an NSERC Discovery Grant RGPIN-2019-06917 to BB. We thank the anonymous reviewers for their constructive comments.

References

Beekhuizen, B., Cui, C. X., & Stevenson, S. (2019). Representing lexical ambiguity in prototype models of lexical semantics. In *Proceedings CogSci.*

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*(3), 531-544.

Ettinger, A., & Linzen, T. (2016). Evaluating vector space models using human semantic priming results. In *Proceedings of the 1st Workshop on Evaluating Vector-Space Representations for NLP* (pp. 72–77).

Firth, J. (1957). A synopsis of linguistic theory 1930-1955. *Studies in Linguistic Analysis*, 1–32.

Grefenstette, G. (1994). Corpus-Derived First , Second and Third-Order Word Affinities. In *Proceedings of the Sixth Euralex International Congress.*

Günther, F., Dudschig, C., & Kaup, B. (2016). Predicting lexical priming effects from distributional semantic similarities: A replication with extension. *Frontiers in Psychology*, *7*.

Hutchison, K. A., Balota, D. A., Cortese, M. J., & Watson, J. M. (2008). Predicting semantic priming at the item level. *Quarterly Journal of Experimental Psychology*, *61*(7), 1036–1066.

Hutchison, K. A., Balota, D. A., Neely, J. H., Cortese, M. J., Cohen-Shikora, E. R., Tse, C.-S., . . . Buchanan, E. (2013). The semantic priming project. *Behavior Research Methods*, *45*(4), 1099–1114.

Jones, M. N., Kintsch, W., & Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552.

Jurafsky, D., & Martin, J. H. (2014). *Speech and language processing. vol. 3.* Pearson London London.

Justeson, J. S., & Katz, S. M. (1991). Co-occurrences of antonymous adjectives and their contexts. *Computational linguistics*, *17*(1), 1–19.

Kenett, Y. N. (2019). What can quantitative measures of semantic distance tell us about creativity? *Current Opinion in Behavioral Sciences*, *27*, 11–16.

Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1470.

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving Distributional Similarity with Lessons Learned from Word Embeddings. *Transactions of the Association for Computational Linguistics*, *3*, 211–225.

Lund, K., Burgess, C., & Atchley, R. A. (1995). Semantic and associative priming in high-dimensional semantic space. In *Proceedings CogSci.*

Lyons, J. (1977). *Semantics* (Vol. I). Cambridge: Cambridge Univeristy Press.

Mandera, P., Keuleers, E., & Brysabert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57-78.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(3), 558–572.

McRae, K., Khalkhali, S., & Hare, M. (2012). Semantic and associative relations: Examining a tenuous dichotomy. In V. F. Reyna, S. B. Chapman, M. R. Dougherty, & J. Confrey (Eds.), *The Adolescent Brain: Learning, Reasoning, and Decision Making* (pp. 39–66). Washington, DC: APA.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings NeurIPS.*

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

Řehůřek, R., & Sojka, P. (2010, May 22). Software Framework for Topic Modelling with Large Corpora. In *Proceedings LREC* (pp. 45–50).

Roller, S., & Erk, K. (2016). PIC a different word: A simple model for lexical substitution in context. In *Proceedings NAACL* (pp. 1121–1126).

Schütze, H., & Pedersen, J. (1993). A Vector Model for Syntagmatic and Paradigmatic Relatedness. In *Making Sense of Words: Proceedings of the Conference* (pp. 104–113).

Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, *29*(1), 41–78.