Assessing the time course of the influence of featural, distributional and spatial representations during reading

Ernesto Guerra[1,2] (ernesto.guerra@mpi.nl)
Falk Huettig[2] (falk.huettig@mpi.nl)
Pia Knoeferle[1] (knoeferl@cit-ec.uni-bielefeld.de)

[1]Cognitive Interaction Technology Excellence Cluster and Department of Linguistics,
Bielefeld University, Inspiration I, 33615, Bielefeld, Germany

[2]Max Planck Institute for Psycholinguistics, Wundtlaan 1,
Nijmegen, 6525 XD, The Netherlands

Abstract

What does semantic similarity between two concepts mean? How could we measure it? The way in which semantic similarity is calculated might differ depending on the theoretical notion of semantic representation. In an eye-tracking reading experiment, we investigated whether two widely used semantic similarity measures (based on *featural* or *distributional* representations) have distinctive effects on sentence reading times. In other words, we explored whether these measures of semantic similarity differ qualitatively. In addition, we examined whether visually perceived spatial distance interacts with either or both of these measures. Our results showed that the effect of featural and distributional representations on reading times can differ both in direction and in its time course. Moreover, both featural and distributional information interacted with spatial distance, yet in different sentence regions and reading measures. We conclude that featural and distributional representations are distinct components of semantic representation.

Keywords: semantic similarity, featural representations, distributional representations, spatial distance, eye tracking, reading.

Introduction

In the context of semantic representation of concepts, two perspectives have dominated research in the cognitive sciences. On one view, semantic representation is based on the perceived physical characteristics of objects (e.g., shape, color, etc.), but also the functional knowledge gained through direct interaction with them (e.g., is-edible, used-to-cut, etc., see Cree & McRae, 2003; McClelland & Rogers, 2003; McRae & Boisvert, 1998; McRae, de Sa, & Seidenberg, 1997; McRae et al., 2005; Rogers & McClelland, 2004, 2008; Vigliocco et al, 2004). For example, the word *sheep* refers to something that bleats, is covered with soft wool, is white or brown, has four legs, and eats grass. This sort of information is generally acquired through the senses. To put it in Andrews and colleagues' words (see Andrews, Vigliocco & Vinson, 2005, 2007, 2009), this kind of representational information can be described as extra-linguistic, featural and experiential. We will refer to this sort of data as *featural* representations for the rest of the paper.

On a different view, semantic representation can be captured by examining the statistical dependencies between words across corpora of spoken and written language. Such corpora could include novels, essays, or articles from newspapers and scientific journals, but also transcribed spoken conversations. Latent semantic indexing (LSI, see Deerwester, Dumais, Landauer, Furnas, Harshman, 1990; Landauer & Dumais, 1997), for instance, is a method that reduces the dimensionality of a language corpus by decomposing each text in a frequency matrix, or text-document. In this model, the statistics are derived by a decomposition of the term frequencies in each of texts. Thus, this data can be described as intra-linguistic, disembodied and *distributional*, as we will refer to it for the rest of the paper.

Indeed, both distributional and featural representations alone can produce models of semantic representation capable of accounting for human behavioral data (McRae et al., 1997; Landauer & Dumais, 1997; Lund & Burgess, 1996; Vigliocco et al., 2004). For instance, McRae et al. (1997) used feature-based similarity cosines to predict a number of human behavioral responses such as reaction times and similarity ratings. Similarly, Landauer and Dumais (1997) used distributional similarity cosines to predict performance both of non-native speakers in an English synonym test and of native speakers in a word-sorting task. Such studies, however, have concentrated on one of these sources of information, often neglecting the other.

More recently, evidence from machine learning has showed that models integrating both featural and distributional information can outperform featural- or distributional-only models (Andrews et al., 2005, 2007, 2009). For instance, Andrews et al. (2007) trained three Bayesian models using either a combination of both featural and distributional representations, or featural or distributional representations alone. The three models were then compared on their predictive power in modeling human data from three semantic tasks; word association norms from, reaction times from a lexical priming experiment, and picture-word interference latencies. Overall, the combined model was the best predictor of human performance in the three tasks.

Andrews and colleagues advocate that featural and distributional representations are both critical for language acquisition and that both contribute to different aspects of semantic representation. Moreover, they argue that theories of semantic representation that exclude either of these two types of representations in favor of the other, would be inevitably limited in their scope (see Andrews et al., 2009: p. 466). Featural representations can be thought to contribute to semantic representation via direct interaction with things in the environment. For instance, a *sheep* is known to bleat, to be soft, white and to have four legs, all attributes perceptible through the senses. A clear limitation of featural representations, however, is that they can only be acquired for concepts that have a perceptible referent in the physical world. Concepts like *peace* or *war* are difficult to describe in terms of perceptible features (but see Kousta et al., 2011 for a focus on introspective information). Distributional representations, in turn, do not distinguish between levels of concreteness of concepts, equally treating *sheep* and *war* as vectors in a matrix. On the other hand, distributional representations from current linguistic corpora focus on the statistical patterns of words in a linguistic context and cannot say anything about the physical world. Thus, featural and distributional representations alone can only provide limited insight into semantic representations.

Against this background, it seems plausible to argue that featural and distributional representations are in fact qualitatively different. Semantic priming and picture-word interference experiments can indeed provide measures of human behavior related to semantic similarity. However, response times in such tasks cannot reveal potential differences in the time course with which featural and distributional representations come into play during online semantic interpretation. Furthermore, if these two kinds of representations are qualitatively distinct, it is likely that they distinctively interact with other cues such as spatial distance (see, Casasanto, 2008; Guerra & Knoeferle, 2012). We examined these two questions using eye tracking.

The present study

The purpose of the present research is two-folded; first, we examined the <u>time course</u> of the effects of featural and distributional representations during real-time sentence interpretation, and, second, we examined whether these semantic similarity measures <u>interact with spatial distance.</u> The second question is based on recent evidence from eye movements during reading showing that spatial distance between objects (far vs. close) modulated online semantic interpretation of sentences expressing similarity (or difference) between abstract nouns (Guerra & Knoeferle, 2012).

Contrasting with those results, we first need to establish whether spatial distance could in fact interact with sentences expressing similarity between <u>concrete nouns</u>. Extending previous findings from abstract- to concrete-noun comparison would allow us to further investigate the *kind* of similarity that can interact with spatial distance, which in turn can further our understanding about the different components (e.g., featural, distributional, visuo-perceptual) that contribute to semantic representation and processing.

Method

Participants Twenty-eight native speakers of German with normal or corrected-to-normal vision were recruited to take part in the experiment. None of them learnt a second language before age six. They all gave informed consent and received a monetary compensation of six euros for participating.

Materials and Design We constructed 60 German sentences, expressing either similarity (1) or difference (2) between two concrete nouns. Words that differed between the two versions of an item were matched for frequency and length. A translated example item is presented below (critical regions in bold font).

(1) 'A goat$_{NP1}$ and$_{coord.}$ a **sheep**$_{NP2}$ are$_{VP1}$ indeed$_{ADV}$ **similar**$_{ADJ}$, **argued**$_{VP2}$ the teacher$_{NP3}$'.

(2) 'A rat$_{NP1}$ and$_{coord.}$ a sheep$_{NP2}$ are$_{VP1}$ indeed$_{ADV}$ **different**$_{ADJ}$, **argued**$_{VP2}$ the teacher$_{NP3}$'.

Spatial distance was introduced into the design by presenting two playing cards in a visual context preceding each sentence. On critical trial, the two cards moved either close together or far apart. We combined each level of similarity (similar vs. different) and spatial distance (close vs. far) in a 2 x 2 within-subjects within-items Latin square design. This resulted in four experimental conditions, namely, Close-Similar, Far-Similar, Close-Dissimilar, and Far-Dissimilar. Each participant saw one condition of each item, and the same number of items per condition.

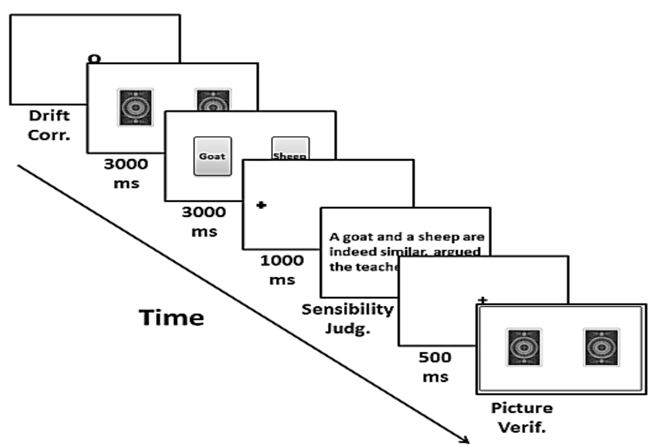

Figure 1: Schematic representation of an experimental trial.

Procedure Before the eye-tracking experiment, the experimenter performed a 9-point calibration procedure. Next, participants completed ten practice trials. After practice, the experimenter re-calibrated the eye tracker and the experiment started. Figure 1 depicts an example experimental trial. On each trial, participants inspected a visual context with two playing cards that moved to different positions and turned around after three seconds, showing two nouns that appeared in the ensuing sentence. Subsequently, participants read a sentence and made a sensibility judgment. Finally, they verified whether a picture of two playing cards matched or mismatched the two playing cards presented before the sentence.

Data Analysis 1

Based on previous findings (see Guerra & Knoeferle, 2012), we defined three critical regions in the sentences (i.e., the NP2, ADJ and the VP2). Before computing reading measures, we removed all fixations longer than 1200 or shorter than 80 ms (cf. Sturt, Keller, & Dubey, 2010). Subsequently, we calculated three reading measures for each critical sentence. First-pass reading time was defined as the sum of all fixations from first entering a region until moving to another region. Regression path duration was defined as the sum of all fixations from first entering a region until moving to the right of that region. Finally, total reading time was defined as the sum of all fixations in a region during the trial (see Rayner, 1998). We analyzed the log-transformed reading measures using a “maximal” linear mixed effects regression (LMER; see Barr et al., 2013) including random intercepts for participants and items, and random slopes for the fixed effects (i.e., similarity, spatial distance) and their interaction.

Results

Data analysis revealed no reliable effects at the NP2 region (all *p*-values > 0.1) but a reliable main effect of spatial distance emerged at the ADV region ($p < .03$). More importantly, reliable interaction effects between spatial distance and similarity emerged in first-pass reading times ($p < 0.004$) and regression path duration ($p < .03$) at the VP2 region. Additionally, a reliable main effect of similarity emerged in this region in total reading times ($p = .01$). Figure 2 presents the pattern of interaction for first-pass reading times and regression path duration for VP2.

Discussion

Previous findings showed that spatial distance could rapidly interact with the interpretation of semantic similarity between abstract nouns (e.g., *war*, *peace*) as reflected by sentence reading times (Guerra & Knoeferle, 2012). In the present study, we extended these results from abstract to concrete nouns. As can be seen in Figure 2, first-pass and regression path reading times were longer for sentences expressing similarity between concrete nouns when preceded by a visual context with objects far apart (vs. close together), while reading times for sentences expressing dissimilarity were longer when preceded by object close together (vs. far apart). We thus replicated the previously-observed interaction effects. Next, we evaluated the influence of our two similarity measures (based on featural versus distributional representations) on reading times, and assessed how these two measures interact with spatial information.

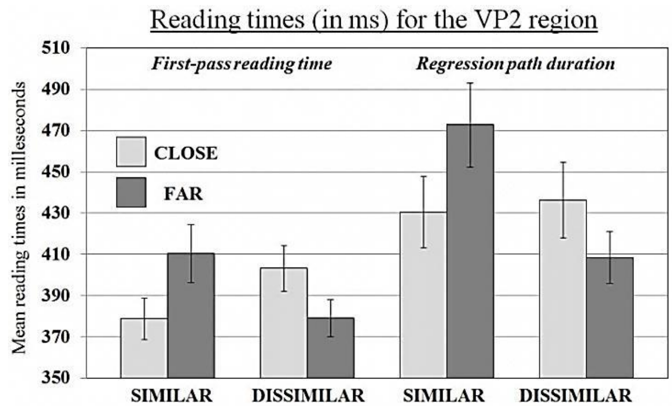

Figure 2: Mean first-pass reading time (on the left) and regression path duration (on the right) in milliseconds for the VP2 region as a function of sentence type (similar vs. dissimilar) and spatial distance (light-gray bars and dark-grey bars represent close and far conditions, respectively). Error bars represent standard errors of the mean.

Data Analysis 2

We performed latent semantic indexing (LSI, see Deerwester et al., 1990; Landauer & Dumais, 1997) for each noun of the 60 concrete-noun triplets (e.g., *goat*, *rat*, *sheep*). A corpus of 20,000 Wikipedia articles in German was used to train our set of triplets. From that initial corpus, a number of 1,618 articles contained words from our item set and was used to calculate LSI vectors of length (K) 10,000. Similarity between the LSI vectors representing the words was calculated by means of the cosine similarity of the LSI vectors. Additionally, for a subset of our items (N=37) we obtained comparable cosine values between concrete nouns from feature-based similarity norms (McRae et al., 2005). Figure 3 presents the correlation between featural and distrubutional cosine values.

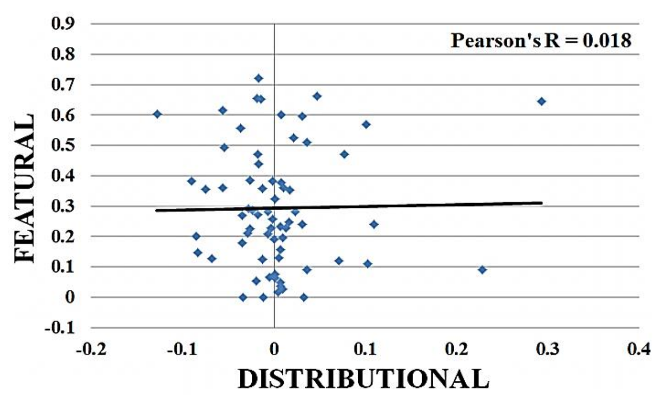

Figure 3: Pearson correlation between featural (McRae et al., 2005) and distributional (LSI) cosine values

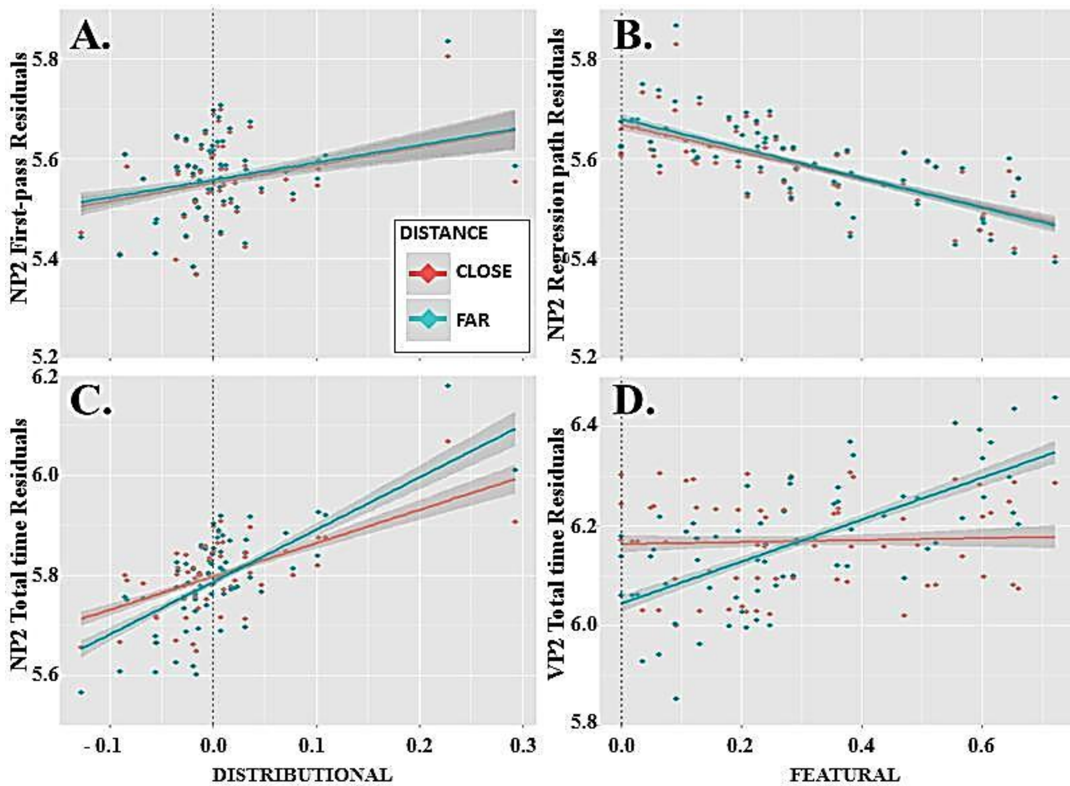

Figure 4: Scatter plots of the LMER model residuals for four reading measures where statistically significant effects of distributional or featural representations were observed. On the y-axis, the reading time residuals from the LMER models are plotted. The x-axis plots the cosine values based on distributional representations (graphs on the left), and based on featural representations (graphs on the right). Panel **A.** depicts the main effect of the distributional factor on first-pass reading times at the NP2 region for both the far and the close conditions. Panel **B.** shows the main effect of the featural factor on regression path duration at the NP2 region for both spatial distance conditions. Panel **C.** presents the interaction effect between distributional factor and spatial distance in total reading times at the NP2 region. Panel **D.** shows the interaction effect between the featural factor and spatial distance in total reading times at the VP2 region.

The effect of distributional and featural representations on log-transformed reading times, and their interaction with spatial distance, was evaluated on the subset of items for which we had both similarity cosine values using LMER models. Such models included, for each reading measure and critical region, distributional and featural cosines values as fixed effects and their interaction with spatial distance. Following the "maximal" logic, we also included random intercepts for participants and items, and random slopes for the fixed effects and their interaction.

Results

At the NP2 region, the LMER analysis revealed reliable main effects of distributional representations in first-pass and an interaction with spatial distance in total reading times (all *p*s < 0.03). Moreover, main effects of featural representations were marginal in first-pass ($p = 0.05$) and statistically significant in regression path ($p = 0.03$). No effects of distributional or featural representations emerged at the ADJ region (all *p*s > 0.1).

By contrast, at the VP2 region, reliable interaction effects between spatial distance and featural similarity emerged in first-pass, regression path (both *p*-values < 0.003) and total reading times ($p < 0.05$), in addition to reliable main effects of distance in first-pass and regression path (both *p*-values < 0.03). Yet, no interaction between spatial distance and distributional similarity emerged in this region. Figure 4 illustrates the influence of distributional and featural representations, and their interaction with spatial distance, on sentence reading times.

Discussion

In our second analysis, we examined whether featural and distributional representations had a distinctive signature on reading times for sentences expressing either similarity or difference between two concrete nouns, and moreover, whether they interact with spatial distance.

The results of the LMER showed that both featural and distributional representations have an early (at the NP2 in first-pass and regression path, respectively) but distinctive influence on reading times. As it can be seen when comparing panels A and B in Figure 4, reading times were positively correlated with the cosine values based on distributional representations (A), but negatively correlated with the cosine values based on featural representations (B). Moreover, the results showed reliable interaction effects between spatial distance and both featural and distributional representations. Yet, such effects emerged in different sentences regions and different measures; distributional representations interacted with spatial distance only in a late reading measure (i.e., total times) while we observed a pervasive interaction effect at the VP2 region between spatial distance and featural representations in early and late reading measures.

Compared to the contrasting main effects of these semantic similarity measures on reading times at the NP2 region, any reliable interaction effects followed the same pattern for both kinds of similarity measures. Overall, reading times increased as the cosine values (or the degree of association between words) increased, however this main effect was significantly reduced when sentences were preceded by objects close together compared to far apart. These findings are in coherence with the results from Analysis 1, and with previously reported interaction effects between spatial distance and semantic similarity during sentence comprehension (see Guerra & Knoeferle, 2012).

General Discussion

One major goal for the cognitive sciences is to elucidate the nature of human semantic representations. We described two main ways of measuring semantic similarity, and briefly discussed how, in most cases, each of them has been used to understand semantic representation in isolation. Indeed, both cosines based on statistical patterns of words among words (the *distributional* tradition) and cosines based on concepts features (the *featural* tradition) are good predictors of human-based behavioral data (see, e.g., Landauer & Dumais, 1997; McRae et al. 1997). However, recent computational-modeling evidence showed that the combination of both kinds of representations has better predictive power compared those using either of them (e.g., Andrews et al., 2007).

We examined whether featural and distributional representations have qualitatively different effects in the context of sentence comprehension. Participants' eye movements were recorded as they read sentences expressing similarity (or dissimilarity) between two concrete nouns. In addition, two objects either far apart or close together preceded each critical item sentence. In two planned analyses, we assessed the effects of featural, distributional, and spatial information respectively on sentences reading times.

The first analysis showed that spatial distance between objects in the visual context modulated sentence-reading times as a function of sentence meaning. This replicated previously-reported effects of spatial distance on reading times and extended them to concrete (rather than abstract) nouns. The second analysis revealed distinctive effect patterns of featural and distributional representations on reading times. At an early sentence region (i.e., NP2) and in early measures, both featural and distributional similarity cosines affected reading times significantly. Interestingly, these effects had opposite directionality: The stronger the association based on distributional representations, the longer the reading times; the stronger the association based on featural representations, the shorter the reading times. Qualitative differences between featural and distributional representations can thus be captured using online measures with fine temporal resolution. In addition to the early main effects, interaction effects between distance and both semantic similarity measures emerged in the reading times. While their time course was different for featural and distributional representations, the direction of the effect was the same. Spatial distance modulated the effect of distributional representations at the NP2 in a late measure (i.e., total times); in turn, it modulated the effect of featural representations at the VP2 region in early and late measures.

Overall then, featural and distributional representations have distinct effects on semantic processing. In agreement with the proposal by Andrews and colleagues, our results suggest that these two indexes of semantic similarity are in fact qualitatively different. The outcome of our experiment offers two main contributions. First, our study examined the fine-grained time course of the effects of these two distinct similarity measures. In doing so, we showed that these measures can rapidly (e.g., in first-pass) and incrementally (as the sentence unfolds) index semantic interpretation. Moreover, examining these effects in the context of sentence reading provides a further evaluation of the influence of these measures on human semantic processing. Second, our study enabled further investigation of the relation between semantic similarity and spatial distance. Moving from abstract to concrete nouns permitted us to evaluate whether spatial distance modulated the effects of either or both featural and distributional information on reading times. In this regard, the interaction patterns observed in Data Analysis 1 were largely driven by the overlap of features between the two concrete concepts, rather than the statistical dependencies between the two words in corpora. However, the reliable interaction effect observed at the NP2 region in total reading times, suggests that spatial distance can also interact with distributional information although in later measures compared to featural representations.

A limitation of the present study is that the feature-based similarity cosines were obtained from a norming study conducted in English (McRae et al., 2005), while the language of the reading experiment was German. Future research should address this open issue, either by collecting feature-norms from German speakers or by conducting a similar eye-tracking reading study in English. We chose McRae's norms, since it included a larger number of our set of word pairs compared to other existing semantic features norms (cf. Vinson & Vigliocco, 2008). Moreover, while the present results contribute to the understanding of semantic representations, we are at this stage hesitant to speculate on the directionality of these early effects and its explanation.

Future research should further explore the consistency of these patterns.

Conclusion

In sum, the results from the present study offer a number of new insights into the role of featural, distributional and spatial information for the semantic interpretation of similarity. First, we extended previous results on spatial distance-similarity interactions from abstract nouns to concrete nouns. Second, based on our results, it is clear that both featural and distributional representations index rapid and incremental effects during sentence processing. Third, such effects were qualitatively distinct, both in terms of their time course and directionality. Finally, we demonstrated that both representational sources can interact with spatial information, but that their interactions differ in the time course.

In conclusion, our results support the idea that these two similarity measures represent qualitatively different aspects of semantic representation and that models that combine both representational sources can reveal their differential influence on human semantic interpretation.

Acknowledgments

This research was funded by the Cognitive Interaction Technology Excellence Cluster (CITEC, German Research Foundation, PK) and by a PhD scholarship awarded to EG by the Ministry of Education, Government of Chile. We thank John McCrae and Phillip Cimiano at the Semantic Computing Group (CITEC, Bielefeld University) for their valuable help in calculating the Latent Semantic Indexation for our training set and Clara Matheus at the Language and Cognition Lab (Bielefeld University) for matching our German nouns to the similarity norms of the English nouns.

References

Andrews, M., Vigliocco, G., & Vinson, D. (2005). The role of attributional and distributional information in semantic representation. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 127–132). Austin, TX: Cognitive Science Society.

Andrews, M., Vigliocco, G., & Vinson, D. (2007). Evaluating the contribution of intra-linguistic and extra-linguistic data to the structure of human semantic representations. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 767-772). Austin, TX: Cognitive Science Society.

Andrews, M., Vigliocco, G., & Vinson, D. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychl Review, 116*, 463-498.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random-effects structure for confirmatory hypothesis testing: Keep it maximal. *JML, 68*, 255-278.

Casasanto, D. (2008) Similarity and proximity: When does close in space mean close in mind. *Memory and Cognition, 36*, 1047-1056.

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *JEP: General, 132*, 163-201.

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., & Harshman, R. A. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science, 41*, 391-407.

Guerra, E., & Knoeferle, P. (2012). Abstract language comprehension is incrementally modulated by non-referential spatial information: evidence from eye-tracking. *Proceedings of the 34th Annual Meeting of the Cognitive Science Society* (pp. 1620–1625). Austin, TX: Cognitive Science Society.

Kousta, S. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: why emotion matters. *JEP: General, 140*, 14-34.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*, 211-240.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28*, 203-208.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4*, 310-322.

McRae, K., & Boisvert, S. (1998). Automatic semantic similarity priming. *JEP: LMC, 24*, 558-572.

McRae, K., Cree, G., Seidenberg, M., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, Instruments, & Computers, 37*, 547-559

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *JEP: General, 126*, 99-130.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372-422.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach.* The MIT Press.

Rogers, T. T., & McClelland, J. L. (2008). Précis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences, 31*, 689-749.

Sturt, P., Keller, F., & Dubey, A. (2010). Syntactic priming in comprehension: Parallelism effects with and without coordination. *JML, 62*, 333-351.

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

Vinson, D. P., & Vigliocco, G. (2008). Semantic feature production norms for a large set of objects and events. *Behavior Research Methods, 40*, 183-190.