# Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words

Paul Hoffman · Matthew A. Lambon Ralph · Timothy T. Rogers

**Abstract** Semantic ambiguity is typically measured by summing the number of senses or dictionary definitions that a word has. Such measures are somewhat subjective and may not adequately capture the full extent of variation in word meaning, particularly for polysemous words that can be used in many different ways, with subtle shifts in meaning. Here, we describe an alternative, computationally derived measure of ambiguity based on the proposal that the meanings of words vary continuously as a function of their contexts. On this view, words that appear in a wide range of contexts on diverse topics are more variable in meaning than those that appear in a restricted set of similar contexts. To quantify this variation, we performed latent semantic analysis on a large text corpus to estimate the semantic similarities of different linguistic contexts. From these estimates, we calculated the degree to which the different contexts associated with a given word vary in their meanings. We term this quantity a word's *semantic diversity* (SemD). We suggest that this approach provides an objective way of quantifying the subtle, context-dependent variations in word meaning that are often present in language. We demonstrate that SemD is correlated with other measures of ambiguity and contextual variability, as well as with frequency and imageability. We also show that SemD is a strong predictor of performance in semantic judgments in healthy individuals and in patients with semantic deficits, accounting for unique variance beyond that of other predictors. SemD values for over 30,000 English words are provided as supplementary materials.

**Keywords** Semantic cognition · Semantic ambiguity · Polysemy · Linguistic context

P. Hoffman (✉) · M. A. Lambon Ralph
Neuroscience and Aphasia Research Unit (NARU), School of Psychological Sciences, University of Manchester,
Zochonis Building, Oxford Road,
Manchester M13 9PL, UK
e-mail: paul.hoffman@manchester.ac.uk

T. T. Rogers
University of Wisconsin,
Madison, WI, USA

Language is a remarkably complex and versatile system of communication. With a finite vocabulary, people can express a near-infinite number of sentiments, and this is true in part because most words can express a range of different ideas in different situations. This flexibility in word usage is undoubtedly useful but comes with a cost: In some circumstances, the correct interpretation of a word is ambiguous. Sometimes the ambiguity is obvious and extreme, as in the case of homonyms with multiple unrelated meanings. *Bark*, for instance, can refer to the sound made by a dog or to the covering of a tree, and the correct interpretation depends entirely on the linguistic and situational context in which the word appears. In the case of polysemous words, the ambiguity is more graded and subtle. For example, *chance* can denote a situation governed by luck ("It's down to chance"), an opportunity that may arise in the future ("I'll do it when I get a chance"), or a risky option ("Take a chance"). These different senses are distinct but clearly related.

Semantic ambiguity can have a strong influence on language processing and comprehension, particularly when words appear in impoverished contexts. In semantic tasks, such as judging the relatedness of words or making categorization judgments, participants are required to focus on a specific interpretation of each word, and ambiguous words suffer a processing disadvantage as a consequence (Piercey & Joordens, 2000; Hino, Lupker & Pexman, 2002). In contrast, in single-word recognition tasks such as lexical decision or reading aloud of single words, meaning provides important feedback to phonological and orthographic levels of representation, but it is not necessary to settle on a

specific meaning. In these cases, an ambiguity *advantage* is often observed (Borowsky & Masson, 1996; Rodd, Gaskell, & Marslen-Wilson, 2002; Rodd, 2004; Woollams, 2005; Yap, Tan, Pexman, & Hargreaves, 2011).

The critical question that we consider in this study is the following: What makes a word ambiguous? The commonly accepted answer is that words are ambiguous if they have more than one sense or meaning. This raises a further question, however, to which the answer is inevitably subjective: How different must two uses of a word be for them to qualify as separate senses? The answer seems straightforward for words like *bark* that have two entirely separate meanings, but it is more troublesome for polysemous words like *chance*. Some lexicographers are "lumpers," who classify each word using as few senses as possible, whereas others are "splitters," who are more willing to accept variation in usage as indicating distinct senses. In either case, both theorists assume that each word can be captured by a limited number of discrete senses. Accordingly, most psycholinguistic studies of ambiguity classify words as either ambiguous or unambiguous on the basis of the number of senses that the words have in dictionaries or on participants' ratings of the number of meanings they have (Kellas, Ferraro, & Simpson, 1988; Borowsky & Masson, 1996; Hino & Lupker, 1996; Hino et al., 2002; Rodd et al., 2002; Rodd, 2004).

In this article, we describe an alternative measure of semantic ambiguity that discards the assumption that words have a discrete number of distinct senses or meanings. Instead we assume, in common with a variety of different computational approaches to language processing (Cruse, 1986; McClelland, St. John, & Taraban, 1989; Elman, 1990; Landauer, 2001; Rodd, Gaskell, & Marslen-Wilson, 2004; Rogers & McClelland, 2004), that a word's meaning varies continuously depending on the particular contexts in which it appears (Cruse, 1986; Landauer, 2001). According to this view, two uses of the same word are never truly identical in meaning, as their precise connotations in each case depend on the immediate linguistic and environmental context. Some words will tend to be used only in quite similar contexts, and so their meanings in each usage will vary only subtly. Other words will appear in a broad variety of quite different contexts, and so will have quite diverse meanings. The degree to which a word's meaning seems ambiguous thus depends on the diversity of contexts in which it can appear. Viewed in this way, the small number of discrete senses that lexicographers assign to words are an attempt to segment this continuous, context-dependent variation.

To illustrate this idea, consider the words *perjury* and *predicament*, neither of which are polysemous (according to the WordNet lexical database). Though both words have a single definition, they vary in the degrees to which they are tied to a particular context. *Perjury* occurs in a very restricted range of situational and linguistic contexts relating to courtrooms and legal proceedings. From the word alone, one can infer a great deal of information about the situation:

for instance, that the agent committing perjury is a witness in a trial and has told a lie under oath, likely under questioning by a lawyer and in front of a judge. In contrast, *predicament*, though it always refers to a difficult dilemma, can be used in a wide variety of contexts: The word is just as appropriate describing a cat stuck in a tree as it is describing a world leader caught in a diplomatic crisis. Without additional context, one can infer very little about the situation in which the word has been deployed. Although most people would judge the core meaning of *predicament* to be the same in the two cases just described, each has a rather different semantic "flavor," since the context alters the way that we interpret the word. This difference in contextual variability between *perjury* and *predicament* is not captured by the traditional definition of semantic ambiguity, since neither word is thought to have multiple senses. One way of measuring these differences is to estimate the degrees to which the different contexts in which a given word appears vary in their meanings: The contexts in which *perjury* occurs likely are more similar in meaning overall, whereas the contexts in which *predicament* occurs likely differ substantially. We have termed this quantity *semantic diversity* (SemD; Hoffman, Rogers, & Lambon Ralph, 2011).

The goals of this article are to present one way of estimating SemD through the use of latent semantic analysis (LSA; Landauer & Dumais, 1997) and to investigate how this SemD measure relates to other psycholinguistic variables associated with semantic ambiguity. We also demonstrate that SemD can predict performance in semantic judgments in healthy and semantically impaired individuals more successfully than can traditional ambiguity measures. Finally, we provide as supplementary materials the SemD values for 31,739 English words.

## Related prior work

We are not the first researchers to consider the contextual influences on lexico-semantic processing. Schwanenflugel and Shoben (1983) proposed that that *context availability*— the ease with which a word brings to mind a particular context—influences word recognition (see also Galbraith & Underwood, 1973; Schwanenflugel, Harnishfeger, & Stowe, 1988). They asked participants to rate this quality for various words on a Likert scale and found that words for which it was easy to generate contexts were recognized more quickly. It seems reasonable to suppose that the availability of a word's context may be related to the variability of the contexts in which it appears: Presumably, when these contexts are more restricted (as in the case of *perjury*), contextual details become more readily available. Yet this relationship is indirect and not perfectly transparent. For instance, a word like *obstacle* may bring to mind a particular concrete context

very strongly (e.g., a herd of cows standing in the middle of a road—a reasonably common occurrence in rural England). This word can also, however, be used in a variety of more abstract situations (e.g., obstacles to career advancement) that do not lend themselves so readily to the construction of a mental image. So the availability of a particular context does not necessarily preclude the existence of a wide range of other potential contexts that do not come to mind as easily. Approaches based on subjective ratings have additional disadvantages more generally. First, data collection is resource-intensive if one is interested in obtaining this quantity across a pool of many thousands of words. Second, it is not clear exactly how participants arrive at their decisions in such studies, particularly when introduced to an unfamiliar concept like contextual availability or variability. It could be the case, for example, that when participants are asked how many different contexts they associate with a word, they assume that more frequently occurring words appear in more diverse contexts and base their decision on their overall familiarity with the word.

Adelman, Brown, and Quesada (2006) adopted a more objective, corpus-based approach to measuring contextual variability. They simply counted the number of documents in a text corpus that contained a given word. The authors called this measure *contextual diversity*, and they demonstrated that it is a better predictor of word recognition speed than is log word frequency. Though this quantity is useful, it is unlikely to reflect the context-dependent variation in word meaning that we are interested in for present purposes (and we note that the authors made no such claim for it). One clue to this is the very strong correlation between contextual diversity and simple (log) word frequency. These two measures are correlated at $r > .95$, suggesting that contextual diversity is principally associated with a word's frequency of occurrence rather than its meaning. The reason why contextual diversity, as defined by Adelman et al., is unlikely to capture variability in meaning is that it does not take into account the degree of similarity between the various contexts in which a word appears. *Tax*, for example, might appear in a large number of documents in total, but if all of those documents relate to similar financial matters, then the word is not associated with a particular breadth of semantic information. Another word might appear in fewer documents overall yet be more diverse, if those documents spanned a broader range of topics.

McDonald and Shillock (2001) also developed a corpus-based method, defining a measure that they called *contextual distinctiveness*, which measures the predictability of a word's immediate context. For each appearance of a particular word in their corpus, they analyzed the distribution of words occurring within a ten-word window. The central idea here was that a word's immediate neighborhood is partially predictable when the word occurs in only a narrow range of local contexts, but is less predictable if the local neighborhood contains many different contexts. For instance, the

neighborhood around the word *amok* is partially predictable because it almost always contains the word *run*. Unlike the Adelman et al. (2006) measure, this approach takes into account the similarity of a word's local contexts. With regard to measuring contextual variation in a word's meaning, however, the ten-word window seems too narrow to encompass the broader theme and topic of each context. Instead, the measure is strongly influenced by the particular phrases in which words occur (like the "run amok" example above). This issue is illustrated by number words (e.g., *five*), which occur in a highly constrained set of contexts according to this measure, because they often appear in the immediate vicinity of other numbers and units of measure. When considering the broader linguistic context, however, we would expect number words to be highly diverse, because they can be used in a wide range of different situations.

Our approach to measuring semantic diversity adapts the objective corpus approaches pioneered by Adelman et al. (2006) and McDonald and Shillock (2001), with the specific goal of measuring the degree to which the various contexts associated with a given word are similar in their general meaning. Toward this end, we developed an approach based on latent semantic analysis (LSA), one of a number of techniques that use patterns of word co-occurrence to construct high-dimensional semantic spaces (Lund & Burgess, 1996; Landauer & Dumais, 1997; Griffiths, Steyvers, & Tenenbaum, 2007). The LSA method utilizes a large corpus divided into a number of discrete contexts, where each context is a sample of text from a particular source. LSA then tabulates a co-occurrence matrix registering which words appear in which contexts. Each word is represented as a vector, with each element of the vector corresponding to its frequency of occurrence in a particular context. The underlying structure in the co-occurrence matrix is then extracted using a data-reduction technique called *singular value decomposition* (SVD). This important step reveals *latent* higher-order relationships between words, based on their patterns of co-occurrence. SVD returns a lower-dimensional vector for each word (typically about 300 elements long), with the similarity structure of these vectors approximating the latent similarity structure of the original co-occurrence matrix. Thus, word representations can, for this approach, be viewed as points in a high-dimensional space, with the proximity between words indicating the degree to which the words appear in similar contexts (and, thus, the extent to which they have similar meanings).

Importantly for the present purposes, the LSA process also places each individual context in the same high-dimensional semantic space, with the proximity of any two contexts reflecting their similarity in content. To calculate SemD for a particular word, we examined all of the contexts in which the word appeared and calculated their average

similarity to one another. When the contexts were very similar to one another on average, this suggested that the word was associated with a fairly restricted set of meanings and was relatively unambiguous. When the contexts associated with a given word were quite dissimilar to one another, this suggested that the meaning of the word was more ambiguous. In a recent study, we demonstrated that SemD is an important factor in predicting the success of comprehension judgments in aphasic patients with comprehension impairments following stroke (Hoffman, Rogers, et al., 2011). The purposes of the present study were to describe our method for computing SemD in more detail, to explore the relationship between SemD and other psycholinguistic measures of semantics and ambiguity, and to make available the SemD values for 31,741 English words. We also extended our previous analyses by demonstrating that SemD successfully predicts semantic processing in healthy participants as well as neuropsychological patients.

## Method for calculating SemD

The method is summarized in Table 1. Our implementation of LSA used the written text portion of the British National Corpus (BNC; British National Corpus Consortium, 2007). We selected this corpus partly because our participants are British, but primarily because, with a total size of 87 million words, even rather low-frequency words occur in it many times. This is critical, because in order to accurately assess the contextual variability of a word, it had to appear in enough different contexts to ensure that a few idiosyncratic or unrepresentative uses would not unduly distort the results. We applied a threshold of 40 contexts as the minimum number that a word should appear in to be included in the analysis. The BNC is made up of 3,125 separate documents, but many of these are very long (e.g., whole newspaper editions or book chapters). We therefore subdivided

each document into separate contexts 1,000 words in length, on the assumption that these smaller chunks would be more likely to be tightly focused on a single topic than would an entire document. This created a total of 87,375 contexts. LSA was carried out using the general text parser (Giles, Wo, & Berry, 2004). Words were only included in the co-occurrence matrix if they appeared in at least 40 contexts and at least 50 times in the corpus as a whole. 38,544 words met these criteria. Morphological variants of the same lemma (e.g., *kick*, *kicks*, and *kicked*) were treated as separate words. Prior to SVD, values in the matrix were log-transformed. The logs associated with each word were then divided by that word's entropy (*H*) in the corpus:

$$H = \sum_c p_c \log(p_c),$$

where *c* indexes the different contexts in which the word appears and $p_c$ denotes the word's frequency in the context divided by its total frequency in the corpus. These standard transformations were performed to reduce the influence of very high-frequency function words whose patterns of occurrence were not relevant in generating the semantic space (Landauer & Dumais, 1997). SVD was then used to produce a solution with 300 dimensions, which is in the region of optimum dimensionality for LSA models. The result of this process was two sets of vectors. First, there was a vector for each word in the corpus, describing its location in the semantic space. These are the vectors typically used in applications of LSA; similarity in the vectors of two words is thought to indicate similarity in their meanings. Second, there was a vector for each context that we analyzed, describing its location in the semantic space. We hypothesized that the similarity between the vectors of two contexts would indicate their similarity in semantic content. These context vectors were used in the calculation of SemD, described below.

*Testing the word vectors in the LSA model*

Our novel application of LSA involved using the context vectors to compute SemD values. Before proceeding with this stage, it was important to confirm that the LSA model that we generated was an accurate approximation of semantic relationships—that is, that the patterns of word similarity extracted from the corpus did indeed reflect genuine patterns in the English language. The word vectors of LSA spaces are typically assessed using standardized human comprehension tests. Landauer and Dumais (1997) tested their model with the synonym judgment section of the Test of English as a Foreign Language and found that its accuracy (64 %) was comparable to that of nonnative English speakers applying to American universities. Using the same approach, we tested our model using a neuropsychological test of synonym judgment

**Table 1** Steps in calculating semantic diversity (SemD)

| Step | Procedure |
| --- | --- |
| 1 | Divide each document in the British National Corpus into "contexts" of 1,000 words. |
| 2 | Perform latent semantic analysis (LSA) procedure on the 1,000-word contexts. This gives LSA vectors for words (as typically used in LSA applications) and vectors for contexts. |
| 3 | For word *x*, find all of the contexts in which *x* appears at least once (if more than 2,000 contexts, randomly select 2,000 for analysis). |
| 4 | Using the LSA vectors for contexts, compute the similarity of all pairwise combinations of contexts containing *x* by taking the cosine of their vectors. |
| 5 | Take the mean of the cosines, then take the natural log of this and reverse the sign to give the semantic diversity value for *x*. |

(Jefferies, Patterson, Jones, & Lambon Ralph, 2009, which we will return to later in the article). Each trial consisted of a probe word and three choices, one of which had a similar meaning to the probe. We calculated the cosine between the vector representing the probe word and each of the three choices (the standard method for measuring distance between words in the LSA space) and judged the model to have "chosen" whichever choice had the highest cosine with the probe. The model selected the correct option on 82 % of trials. Native English speakers perform slightly better on this test— scores around 95 % are typical. However, many of the model's errors occurred on trials with the highest error rates and longest RTs in human participants, indicating that the model's judgments of word similarity were a good approximation of human performance.[1]

Computing SemD from the LSA space

Having created and tested the LSA semantic space, we calculated SemD by making use of the semantic vectors generated for contexts. Just as a high cosine between the vectors representing two words is thought to indicate that they are related in meaning, a high cosine between the vectors for two contexts would indicate that they were similar in their topics. For a given word, we took all contexts in which it appeared and calculated the cosine between each pair of contexts in the set. We then took the mean of these cosines to represent the average similarity between any two contexts containing the word.[2] The resulting values were log-transformed, and their signs were reversed so that words appearing in more diverse contexts had higher values.

Using this method, we calculated SemD for all 38,544 words in the corpus. Figure 1 shows a histogram depicting the distribution of the words in the corpus across SemD values. Below the bars are three representative examples of words from each SemD band. The lowest values were obtained for words with highly specific meanings that were likely to occur in a restricted range of contexts. *Gastric*, for example, only appeared in medical discussions relating to the digestive system, and these contexts were represented as being very similar to one another in the LSA space. Higher

values were obtained for words that could be used in a wide range of contexts and whose precise interpretation might be highly ambiguous.

To test the reliability of the technique, we also calculated SemD values using two alternative LSA spaces. The first was generated using the BNC, but with each document divided into contexts of 250, rather than 1,000, words in length. The smaller context size resulted in a fourfold increase in the number of contexts to be analyzed, which meant that it was necessary to truncate the corpus slightly to ensure that the SVD could be calculated successfully. We did this by only analyzing the first 50,000 words of each BNC document, which reduced the overall size of the corpus to 79 million words. The SemD values derived from the 1,000- and 250-word BNC spaces were strongly correlated ($r = .90$, $p < .0001$; $N = 5,004$). We also created an LSA space using the TASA corpus (Zeno, Ivens, Millard, & Duvvuri, 1995). This corpus contains just over 13 million words and is designed to reflect the reading experience of American students from third grade up to college level. The documents in this corpus are much shorter than those in the BNC (average length = 278 words) so each document was treated as a single context. Due to the smaller size of this corpus, we relaxed our criteria for word inclusion: All words occurring in the corpus at least four times in two different contexts were included in the LSA computation, and SemD values were subsequently calculated for all words occurring in at least ten contexts. Despite the values being derived from somewhat divergent corpora (i.e., British vs. American English and everyday vs. educational reading experience), we found a strong correlation between the BNC- and TASA-derived SemD values ($r = .76$, $p < .0001$; $N = 4,631$).

Relationship to other psycholinguistic measures

In this section, we investigate how our notion of SemD relates to existing approaches to measuring ambiguity, as reviewed in the introduction. We also investigated the relationship between SemD and the key psycholinguistic measures of frequency, concreteness, and imageability. Intuitively, one would expect words with variable, ambiguous meanings to be higher in frequency, because there would be more situations in which they could be employed. Indeed, in previous studies we have argued that the greater semantic diversity of high-frequency words is the principal reason for the absence of frequency effects in the comprehension of patients with semantic deficits following stroke (Hoffman, Jefferies, & Lambon Ralph, 2011; Hoffman, Rogers, et al., 2011). It has also long been argued that abstract words have inherently more variable and context-dependent meanings than do concrete or highly imageable words (Bransford & Johnson, 1972; Kieras, 1978; Saffran, Bogyo,

---

[1] English speakers produced a correct-response rate of 97 % for the trials that the model completed correctly. In contrast, they were only 85 % accurate for trials on which the model made an error. Likewise, the mean reaction time for the trials that the model completed correctly was 1,986 ms, whereas it was 2,726 ms for the trials on which the model erred.

[2] In fact, the process of calculating cosines for *all* pairwise combinations of contexts would have been computationally prohibitive for very high-frequency words that appeared in many thousands of contexts. To make the process more tractable, we analyzed a maximum of 2,000 contexts for each word. When a word appeared in more than 2,000 contexts, we randomly selected 2,000 for the analysis.

**Fig. 1** Distribution of words as a function of semantic diversity (SemD)
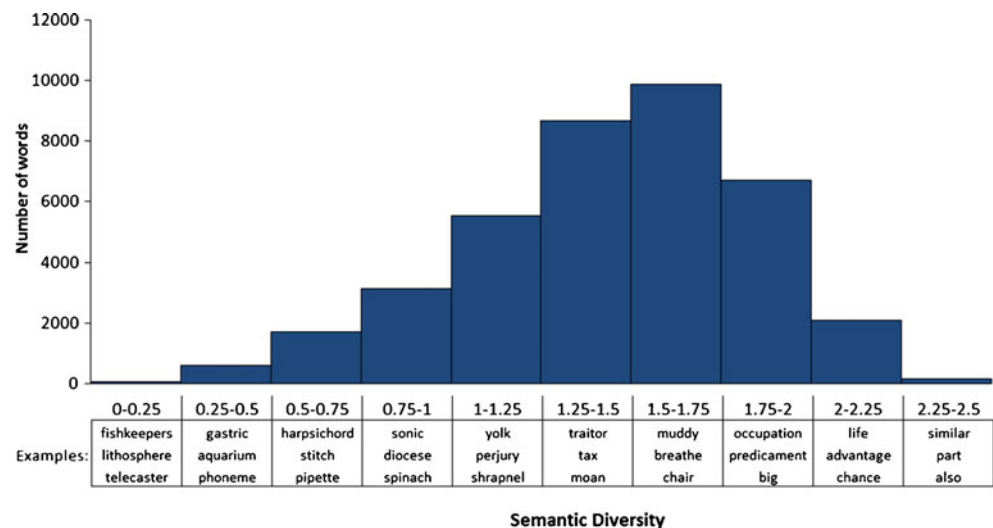


Schwartz, & Marin, 1980). Schwanenflugel and colleagues made the argument that concrete words are generally processed more efficiently than abstract words because they have greater context availability—that is, it is easier to link them to specific contexts (Schwanenflugel & Shoben, 1983; Schwanenflugel et al., 1988).

Method

We analyzed the properties of 325 words for which concreteness, imageability, and context availability ratings were collected by Altarriba, Bauer, and Benvenuto (1999). Log-transformed frequency values for all words were obtained from the BNC, and SemD was calculated using the method described above. We also obtained values for the following variables, which we will refer to as *ambiguity measures*, because they all measure different aspects of contextual variability.

*Contextual diversity* Following Adelman et al. (2006), this was computed for each word by calculating the total number of contexts in the BNC that contained the word (from a possible total of 87,375) and log-transforming this value.

*Number of senses* We looked up the total number of senses for each word in the WordNet lexical database (Miller, 1995). This is a commonly used source for ambiguity studies based on dictionary definition counts (McDonald & Shillock, 2001; Rodd et al., 2002; Yap et al., 2011). As the distribution of the number of senses was somewhat skewed (most words had ten or fewer senses, but a small number of words had 20 or more senses), these values were log-transformed.

*Contextual distinctiveness* Values for this measure were obtained from the database reported by McDonald and Shillock (2001). Only 279 of the 325 words were present in this database. Contextual distinctiveness values are smallest for those words whose local (ten-word) contexts are highly predictable—that is, those that are least ambiguous. Thus, this scale operates in the opposite direction from the other ambiguity measures.

Results

A correlation matrix for all variables is shown in Table 2, and Fig. 2 shows scatterplots depicting the relationship between SemD and each of the other variables. SemD was significantly correlated with all of the other context/ambiguity measures, indicating convergence between the differing approaches to variation in meaning or contextual usage. Words high in SemD tended to have many different senses, to occur in a large number of contexts (indexed by the contextual diversity measure), and to be low in contextual distinctiveness, indicating that their local contexts were highly unpredictable. Note, however, that there was considerable variation in SemD values even amongst words that had only one sense in WordNet (log senses = 0), suggesting that SemD might provide a more fine-grained and continuous measure of ambiguity than the traditional, definition-based approach. For example, the two words *knowledge* and *aeroplane* both had only one sense, but they differed substantially in their SemD values. In addition, context availability ratings indicated that participants found it harder to generate a particular context for high-SemD words, presumably as a result of the high variability in their contexts.

We also found that SemD was strongly correlated with frequency (as was previously reported by Hoffman, Rogers, et al., 2011). In fact, all of the ambiguity measures were correlated with frequency, demonstrating that higher-frequency words are associated with a higher degree of variability in their contextual usage. SemD was also strongly correlated with imageability and concreteness, supporting previous assertions that abstract words are associated with greater contextual variability (Saffran et al.,

**Table 2** Correlations between lexical–semantic and ambiguity measures

| Variable | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|
| 1. Semantic diversity | .49*** | −.51*** | −.48*** | −.26*** | .56*** | −.35*** | .36*** |
| 2. Log word frequency | – | −.07 | −.20*** | .26*** | .99*** | −.68*** | .52*** |
| 3. Concreteness | | – | .75*** | .62*** | −.15** | −.02 | .01 |
| 4. Imageability | | | – | .34*** | −.24*** | .03 | −.10 |
| 5. Context availability | | | | – | .20*** | −.20** | .16** |
| 6. Contextual diversity | | | | | – | −.68*** | .53*** |
| 7. Contextual distinctiveness | | | | | | – | −.44*** |
| 8. Log number of senses | | | | | | | – |

* $p < .05$. ** $p < .01$. *** $p < .001$. $N = 325$ for all variables except contextual distinctiveness, for which $N = 279$.

1980; Schwanenflugel & Shoben, 1983). However, the other ambiguity measures were only weakly correlated with concreteness/imageability (in the case of contextual diversity) or were not correlated with these variables at all (with number of senses and contextual distinctiveness).

We explored the relationship between the measures in more detail by performing a principal components analysis using SPSS. This revealed two factors with eigenvalues greater than one, together accounting for 73 % of the total variance. The Varimax-rotated factor loadings are shown in Table 3. Note, first of all, that frequency loads very strongly on the first factor and that concreteness and imageability both load strongly on the second factor. The three existing ambiguity measures (number of senses, contextual diversity, and contextual distinctiveness) all load strongly on the frequency factor but not on concreteness/imageability. Therefore, these measures all capture the tendency for higher-frequency words to be more ambiguous, but not the hypothesized greater ambiguity of more abstract/less imageable words. In contrast, context availability loads heavily on the concreteness/imageability factor,

but not on the frequency factor. Interestingly, SemD is the only measure that loads evenly on both factors. Thus, this analysis suggests that the SemD measure aligns with two intuitive and widely held beliefs about semantic ambiguity: that word meanings are more likely to be contextually variable if the words are high in frequency *and* if they are abstract. None of the other measures has this property, suggesting that SemD may be more successful in accounting for variation in semantic processing than are the other measures. We tested this hypothesis in the next section.

### Ambiguity effects in the semantic judgments of healthy participants and of semantically impaired stroke patients

Ambiguity typically has a negative effect on semantic processing in healthy individuals, since interference between multiple, competing meanings must be resolved (Hino et al., 2002; Yap et al., 2011). Similarly, neuropsychological patients
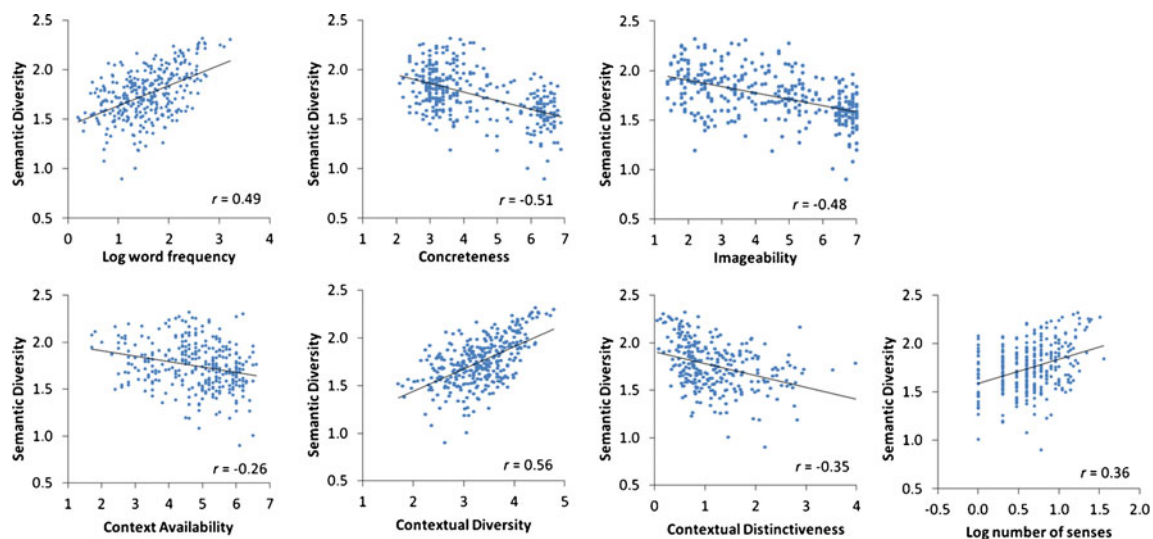


**Fig. 2** Relationship between SemD and other lexical–semantic variables

**Table 3** Varimax-rotated factor loadings from principal components analysis

| Variable | Factor 1 | Factor 2 |
|---|---|---|
| Semantic diversity | **.58** | **−.61** |
| Log word frequency | **.93** | −.12 |
| Concreteness | −.11 | **.93** |
| Imageability | −.22 | **.82** |
| Context availability | .24 | **.77** |
| Contextual diversity | **.93** | −.21 |
| Contextual distinctiveness | **−.83** | −.15 |
| Log number of senses | **.64** | −.05 |

Factor loadings with absolute values greater than .3 are shown in boldface.

with semantic deficits following stroke have particular problems comprehending ambiguous words (Noonan, Jefferies, Corbett, & Lambon Ralph, 2010). In this section, we tested whether SemD could successfully capture the semantic-processing disadvantage for ambiguous words in the semantic processing of two groups of participants:

(a) patients diagnosed with semantic aphasia (SA; Head, 1926; Jefferies & Lambon Ralph, 2006)—a form of aphasia in which verbal and nonverbal comprehension impairments are observed following left-hemisphere stroke. This was a reanalysis of data reported previously by Hoffman, Rogers, et al. (2011).

(b) novel data collected from a group of healthy individuals matched in age and educational level to the semantic aphasia patients.

All of the participants were asked to perform a synonym judgment task (e.g., which word is closest in meaning to *advantage*: *benefit*, *condition*, or *tendency*?). This task required participants to access the appropriate aspect of a word's meaning to match it to its synonym, meaning that words with highly variable meanings would likely be at a disadvantage. In earlier work, we established that patients with semantic aphasia have difficulty selecting which aspects of their semantic knowledge are relevant in a particular task or context ("semantic control"; Jefferies & Lambon Ralph, 2006; Noonan et al., 2010; Corbett, Jefferies, & Lambon Ralph, 2011). Thus, we predicted that patients with semantic aphasia would have difficulty comprehending highly ambiguous words because these words maximize selection demands. Highly ambiguous words are associated with a wide variety of semantic information, much of which is not relevant in a specific situation (Metzler, 2001; Rodd, Davis, & Johnsrude, 2005; Bedny, Hulbert, & Thompson-Schill, 2007; Hoffman, Jefferies, et al., 2010; Noonan et al., 2010). We predicted that healthy participants would also show negative effects of increased ambiguity, but that these effects would be evident primarily in terms of increased RTs rather than errors.

Method

*Participants* A group of 13 semantic aphasia patients were tested (described in more detail by Hoffman, Rogers, et al., 2011). They had a mean age of 66 (range = 36–81). All had suffered left-hemisphere strokes that resulted in chronic verbal and nonverbal comprehension deficits. A group of 26 healthy older adults also took part (mean age = 62, range = 36–74), recruited from research volunteer databases in Manchester and Cambridge, U.K.[3]

*Task* The synonym judgment task is described in detail elsewhere (Jefferies et al., 2009; Hoffman, Rogers, et al., 2011). Briefly, each trial consisted of a probe word and three choices. Participants were asked which of the three choices was related in meaning to the probe, in a total of 96 trials. For patients, the words for each trial were presented on a sheet of paper and read aloud by the experimenter. The patient responded by pointing. For healthy participants, trials were presented on a computer screen using E-prime software, and participants indicated their response with a buttonpress as quickly as possible.

*Data analysis* We began by removing three trials for which we could not obtain a SemD value for the probe (because the probes appeared in fewer than 40 contexts in our corpus) and one trial (with the probe *suffix*) that had an unusually high error rate in healthy participants. For the remaining 92 trials, RTs from the healthy participants were screened by removing any outlying values that were over two standard deviations from a participant's mean. We obtained values for SemD, log frequency, contextual diversity, and number of senses for each trial, as described earlier. Since the properties of both the probe and its synonym could influence performance, for each trial we averaged the values across the probe and synonym to give values for the trial as a whole. We also obtained imageability values for the probes from the MRC database (Coltheart, 1981). Context availability and contextual distinctiveness values were not available for enough trials to be included in the analysis.

The effects of SemD were investigated in two ways. First, we performed a median split on the 92 trials to create sets of high- and low-SemD trials. Performance across the two conditions was compared in each group. For healthy

---

[3] The data from healthy participants were briefly referred to earlier in this article, when their responses on the test were used to test the semantic representations of the LSA model.

individuals, we investigated error rates and RTs. RT data were not collected for the patients, so only error rates were analyzed. Second, to test for unique effects of SemD, independent of other confounding variables, we performed a series of regression analyses in which the lexical–semantic variables were used to predict performance on a by-trials basis. This also allowed us to directly compare the predictive power of the different ambiguity measures. In these analyses, we focused on error rates in the patients and RTs in the healthy participants. Errors in the healthy group were not analyzed due to ceiling effects.

Results

Performance in the high- and low-SemD conditions is shown in Fig. 3. Semantic aphasia patients made more errors on high-SemD trials, as we predicted [$t(12) = 8.27$, $p < .001$]. Healthy participants exhibited much lower error rates overall, but also made more errors on the high-SemD trials [$t(25) = 5.03$, $p < .001$]. Furthermore, their RTs were significantly slower on high-SemD trials [$t(25) = 8.44$, $p < .001$]. These findings support the hypothesis that the greater semantic variability of high-SemD words makes them more difficult to comprehend. High- and low-SemD trials did, however, differ on a number of other variables, consistent with the pattern of intercorrelations described in the previous section. High-SemD trials were significantly higher in frequency, contextual diversity, and number of senses than were low-SemD trials, and significantly lower in imageability (see Table 4). To test for unique effects of SemD, we performed a series of regression analyses on the results by trials.

In the first analysis, frequency and imageability were entered as predictors in the first step and SemD in the second step. Correlations between predictors are presented in Table 5. Results for the patients' error rates and healthy participants' RTs are shown in Table 6. The patterns of results were similar in both cases. The inclusion of SemD
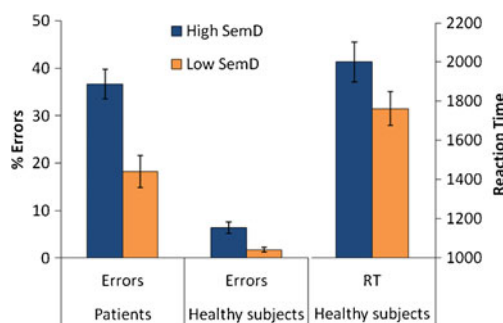
**Table 4** Properties of high- and low-SemD trials

| Variable | High SemD | Low SemD |
|---|---|---|
| Semantic diversity | 1.93 | 1.43[*] |
| Log word frequency | 1.78 | 1.04[*] |
| Imageability | 3.92 | 5.11[*] |
| Contextual diversity | 3.54 | 2.69[*] |
| Log number of senses | 0.71 | 0.40[*] |

[*] The high- and low-SemD conditions differ at $p < .001$.

significantly improved the fits of both regression models [patients, $\Delta R^2 = 9.4$ %, $F(1, 88) = 11.9$, $p = .001$; healthy participants, $\Delta R^2 = 4.8$ %, $F(1, 88) = 7.21$, $p = .009$]. Moreover, in both cases the addition of SemD into the models uncovered a latent effect of word frequency that was not present in the first step of the model or in the raw correlations. This result was reported previously in semantic aphasia patients by Hoffman, Rogers, et al. (2011). Here, we demonstrate that the same pattern is observed in the RTs of healthy individuals. This analysis demonstrates the potential value of taking SemD into account when investigating semantic judgments. The final results indicate that high SemD has a negative effect on performance, and high word frequency has a positive effect. However, the strong correlation between these two variables means that the effect of frequency can only be observed when SemD is also taken into account.

In a series of further regressions, we directly compared the predictive power of the three ambiguity measures SemD, contextual diversity, and number of senses. In each case, we entered two of the ambiguity measures in the first step, along with frequency and imageability, and observed the improvement in the model when the final ambiguity measure was added. These results give an indication of the independent predictive power of each measure beyond the others. Figure 4 shows the results. The addition of SemD resulted in a large increase in the $R^2$s of the models, beyond that afforded by the other variables. In contrast, the inclusion of contextual diversity and number of senses did not significantly improve the fits of the models. This suggests that the SemD measure captures unique variance in semantic judgments that is not accounted for either by a simple count of the number of contexts in which the word appears (contextual diversity) or by the traditional approach of considering the number of dictionary senses that a word has.

General discussion

Ambiguity is an important feature of lexical–semantic processing and has traditionally been quantified on the basis of subjective estimates of the number of discrete meanings or senses associated with a word. In this study, we described an



**Fig. 3** Effects of SemD on synonym judgments in healthy participants and semantic aphasia patients. Bars indicate one standard error of the mean

**Table 5** Correlation matrix for synonym judgment analysis

| Variable | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|
| 1. Patients' errors | .56*** | .41*** | .08 | −.47*** | .11 | .14 |
| 2. Healthy participants' RT | – | .42*** | .08 | −.61*** | .10 | .02 |
| 3. Semantic diversity | | – | .70*** | −.45*** | .73*** | .57*** |
| 4. Log word frequency | | | – | −.11 | .99*** | .67*** |
| 5. Imageability | | | | – | −.15 | −.15 |
| 6. Contextual diversity | | | | | – | .68*** |
| 7. Log number of senses | | | | | | – |

*p < .05. **p < .01. ***p < .001. N = 92.

alternative, computationally derived measure of ambiguity based on the assumption that the meanings of words vary continuously as a function of their context. On this view, a word's meaning changes whenever it appears in a different context—subtly if the change in context is slight, but more substantially if the context is radically different. Using this principle, we estimated the degree of ambiguity associated with a word by measuring the similarity of the different contexts in which it could appear, using latent semantic analysis to determine how similar two contexts were. This quantity, termed *semantic diversity* or *SemD*, was found to be correlated with the number of distinct senses that a word had as well as with other measures of the contextual variability of words. Considerable variation in SemD values was also found for words with only one sense, which would typically all be classed as unambiguous. SemD was significantly correlated with word frequency, consistent with the idea that words that are used more frequently are likely to be associated with more contextual variation. SemD was also correlated with imageability/concreteness—a relationship that was not present for number of senses or for other contextual variation measures. However, although SemD shares covariance with both word frequency and imageability, we also demonstrated that this measure accounts for substantial unique variance in semantic judgment performance, both for healthy individuals and

aphasic patients with comprehension deficits. Perhaps most notably, latent effects of word frequency emerged in the semantic task once SemD was taken into account, suggesting that inclusion of SemD in predictive models can impact the observed effects of established psycholinguistic variables.

In this study, we have focused on the effects of ambiguity in semantic processing—that is, in tasks for which the meaning of a word must be considered explicitly in order to generate the appropriate response. Small effects of ambiguity have also been reliably observed in lexical decision and single-word reading, though for these tasks more ambiguous words are processed *more* efficiently (Borowsky & Masson, 1996; Rodd et al., 2002; Rodd, 2004; Woollams, 2005; Yap et al., 2011). In these tasks, activation of semantic information is thought to support task performance, but it is not necessary to access a particular sense or interpretation of the word. The additional semantic information associated with ambiguous words may therefore be beneficial in these cases. The role of SemD in word recognition remains a target for future work, but we predict that the more diverse semantic information associated with high-SemD words would result in more efficient recognition. Indeed, Jones, Johns, and Recchia (2012), using an approach similar to the one described here, recently found that being associated with a broad range of disparate contexts had a facilitative effect on word recognition.

**Table 6** Results of the regression analyses on synonym judgment accuracy

| Predictor | Patients' Errors | | | | Healthy Participants' RTs | | | |
|---|---|---|---|---|---|---|---|---|
| | $R^2$ | B | Standard Error | β | $R^2$ | B | Standard Error | β |
| *Step 1* | .22*** | | | | .37*** | | | |
| Log frequency | | .008 | .03 | .03 | | 7.57 | 44.2 | .01 |
| Imageability | | −.064 | .01 | −.46*** | | −150 | 20.9 | −.61*** |
| *Step 2* | .31*** | | | | .42*** | | | |
| Log frequency | | −.088 | .04 | −.30* | | −115 | 62.5 | −.22(*) |
| Imageability | | −.038 | .01 | −.27* | | −117 | 23.7 | −.47*** |
| Semantic diversity | | .340 | .10 | .50*** | | 435 | 163 | .36** |

*p < .05. **p < .01. ***p < .001. (*)p = .069.
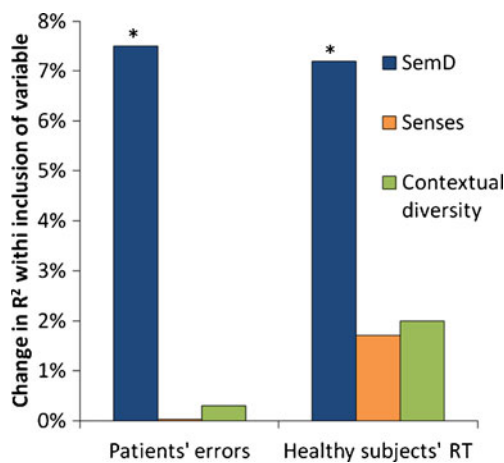
* indicates significant increase in $R^2$; $p < 0.05$.

**Fig. 4** Improvement of regression models with the inclusion of each variable. *Significant increase in $R^2$, $p < .05$

Our results also have implications for theories of semantic representation more generally. One longstanding view suggests that each meaning of an ambiguous word is represented as a separate lexical or semantic node in a network, with these nodes competing for activation when the word is processed (Rubenstein, Garfield, & Millikan, 1970; Morton, 1979; Jastrzembski, 1981). Such models, developed primarily to account for homonyms with unrelated meanings (though, for a similar account of polysemy, see Klein & Murphy, 2001), assume that each word is associated with a fixed number of discrete meanings, each with its own node in the network. This view is clearly related to the traditional idea that ambiguity is best measured by counting the discrete number of senses that a word is thought to have. It is less clear, however, how the localist representation of alternative meanings might be reconciled with the view, supported by the present study, that ambiguity of meaning varies on a continuum.

An alternative approach views representations of word meaning, not as discrete nodes that compete for activation, but as graded and distributed patterns of activation that capture conceptual similarity structure (e.g., Harm & Seidenberg, 2004; Rogers, Lambon Ralph, Garrard, Bozeat, McClelland, Hodges & Patterson, 2004; Rogers & McClelland, 2004; Lambon Ralph, Sage, Jones, & Mayberry, 2010). In such models, similar meanings are represented by similar patterns of activation over the same set of processing units, and similarity of meaning can vary in a continuous fashion. Words with highly stable (i.e., unambiguous) meanings are associated with semantic representations that vary minimally across different uses; words with highly ambiguous meanings are associated with patterns that vary greatly; and the various senses of polysemous words are associated with patterns that are related but somewhat distinct. This approach allows for

many varying and graded degrees of ambiguity, as assumed by our measure of SemD.

Rodd et al. (2004) demonstrated the promise of this approach by modeling the recognition of homonyms, polysemous words, and unambiguous words in a Hopfield network. On each presentation, a given word was associated with a particular pattern of semantic activation. Unambiguous words were always associated with the same pattern, whereas homonyms (like "bank") were associated with very distinct patterns on different presentations. Importantly, polysemous words were, on different presentations, associated with a core semantic pattern distorted by some degree of noise, with each distortion corresponding to one of the word's different senses. The result was that the various senses of a word had distinct but overlapping patterns of activation.

Following training, Rodd et al. (2004) measured low long it took for the network to settle into a steady state following presentation of a given word as input. Homonyms—those that had been associated with very different semantic representations—took the longest time to settle; unambiguous words settled somewhat faster; but polysemous words were the fastest to settle overall. Interestingly, this is just the pattern observed in the response times of human beings performing such lexico-semantic tasks as word recognition: Homonyms with two unrelated meanings take longer to recognize than do unambiguous words, but polysemous words with many related meanings are recognized faster than unambiguous words (Rodd et al., 2002). Rodd et al.'s (2004) model provides an explanation as to why this should be: The association of a single orthographic representation with a variety of overlapping semantic representations led the model to form a single large attractor basin for the word, allowing it to settle more quickly (and, hence, to recognize the item more rapidly) for polysemous words as compared to unambiguous words. In contrast, homonyms with two nonoverlapping semantic representations were associated with two completely separate attractor basins. Settling thus involved a kind of "competition" between the two attractors that took comparatively longer to resolve.

Rodd et al. (2004) did not explicitly state why polysemous words should be associated with more noisy, variable patterns of semantic activation. We propose that the noise in their simulations can be understood as a model analogue of the way that prior context modulates the semantic activation generated upon presentation of a word. Polysemous words appear in a somewhat diverse set of contexts, producing a degree of variability in the semantic activation generated each time that the word is processed. Words that consistently appear in very similar contexts are associated with similar patterns of activation whenever they occur. This context-dependent modulation of semantic activation has been implemented in connectionist models of sentence processing (McClelland et al., 1989; Elman, 1990; St. John, 1992). In these models, words are

processed sequentially, but the distributed representation activated upon the presentation of each word is influenced not just by the identity of the word, but also by the representations of the immediately preceding words. The consequence is that the pattern of activation elicited for each word depends on its prior context, allowing for context-dependent variation in meaning. Such models are able to build up "gestalt" representations of the meanings of entire sentences by combining the representations of individual words in this way (McClelland et al., 1989). The implication for the present work is that the representation of an individual word is different each time that it appears in a different context, explaining why high-SemD words are likely to have variable patterns of semantic activation.

This consideration of the Rodd et al. (2004) model also illustrates some limitations of the present work. Specifically, we have considered the degree to which the contexts in which a given word appears vary in their meanings, but we have not considered the extent to which the different contexts *cluster* into quite distinct meanings. For instance, homonyms like *bark* may occur in two "kinds" of contexts—one set that mainly pertains to dogs, and another that mainly pertains to trees. The full range of contexts in which *bark* appears may be quite diverse across these two sets, but much less diverse within them. In contrast, a polysemous word like *chance* may occur in a broad variety of different contexts that do not tend to naturally fall into clusters. Our measure, SemD, will register both kinds of words as highly semantically diverse, but a model like that described by Rodd et al. (2004) will treat these two situations quite differently, forming two nonoverlapping basins of attraction for homonyms like *bark* and a single, very broad basin for polysemous words like *chance*.

This consideration opens up two important avenues for future inquiry. First, it will be useful to marry the ideas in the present work to explicit, implemented computational models of semantic processing in order to better understand how behavior is expected to be influenced by SemD and related measures. Second, future work should focus on measuring not only the variability in meanings of the various contexts associated with a given word, but also on their apparent substructure—that is, the degree to which the different contexts group into distinct clusters in representational space. Such a measure may allow us to understand better the apparent processing differences that exist between homonyms and polysemous words, and also to reconcile competing views of the representational structure underlying semantic ambiguity—on the one hand, the view that ambiguity involves competition between a small number of discrete representations of distinct meanings and is only present for a subset of words, and on the other, the view that ambiguity is a graded phenomenon and that all words are associated with *some* degree of context-dependent variability in meaning.

## References

Adelman, J. S., Brown, G. D. A., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science, 17,* 814–823.

Altarriba, J., Bauer, L. M., & Benvenuto, C. (1999). Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words. *Behavior Research Methods, Instruments, & Computers, 31,* 578–602.

Bedny, M., Hulbert, J. C., & Thompson-Schill, S. L. (2007). Understanding words in context: The role of Broca's area in word comprehension. *Brain Research, 1146,* 101–114.

Borowsky, R., & Masson, M. E. J. (1996). Semantic ambiguity effects in word identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 63–85. doi:10.1037/0278-7393.22.1.63

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisities on understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

British National Corpus Consortium. (2007). *British National Corpus version 3* (BNC XMLth ed.). Oxford, U.K.: Oxford University Computing Services.

Coltheart, M. (1981). The MRC Psycholinguistic Database. *Quarterly Journal of Experimental Psychology, 33A,* 497–505. doi:10.1080/14640748108400805

Corbett, F., Jefferies, E., & Lambon Ralph, M. A. (2011). Deregulated semantic cognition follows prefrontal and temporoparietal damage: Evidence from the impact of task constraint on non-verbal object use. *Journal of Cognitive Neuroscience, 23,* 1125–1135.

Cruse, D. A. (1986). *Lexical semantics*. Cambridge, U.K.: Cambridge University Press.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–211.

Galbraith, R. C., & Underwood, B. J. (1973). Perceived frequency of concrete and abstract words. *Memory & Cognition, 1,* 56–60.

Giles, J. T., Wo, L., & Berry, M. W. (2004). GTP (general text parser) software for text mining. In H. Bozdogan (Ed.), *Statistical data mining and knowledge* (pp. 455–471). Boca Raton, FL: CRC Press.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review, 114,* 211–244. doi:10.1037/0033-295X.114.2.211

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Cooperative division of labor between visual and phonological processes. *Psychological Review, 111,* 662–720.

Head, H. (1926). *Aphasia and kindred disorders*. Cambridge, U.K.: Cambridge University Press.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology. Human Perception and Performance, 22,* 1331–1356. doi:10.1037/0096-1523.22.6.1331

Hino, Y., Lupker, S. J., & Pexman, P. M. (2002). Ambiguity and synonymy effects in lexical decision, naming, and semantic categorization tasks: Interactions between orthography, phonology, and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 686–713.

Hoffman, P., Jefferies, E., & Lambon Ralph, M. A. (2010). Ventrolateral prefrontal cortex plays an executive regulation role in comprehension of abstract words: Convergent neuropsychological and rTMS evidence. *Journal of Neuroscience, 46,* 15450–15456.

Hoffman, P., Jefferies, E., & Lambon Ralph, M. A. (2011). Remembering "zeal" but not "thing": Reverse frequency effects as a

consequence of deregulated semantic processing. *Neuropsychologia, 49,* 580–584.

Hoffman, P., Rogers, T. T., & Lambon Ralph, M. A. (2011). Semantic diversity accounts for the "missing" word frequency effect in stroke aphasia: Insights using a novel method to quantify contextual variability in meaning. *Journal of Cognitive Neuroscience, 23,* 2432–2446.

Jastrzembski, J. E. (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology, 13,* 278–305.

Jefferies, E., & Lambon Ralph, M. A. (2006). Semantic impairment in stroke aphasia vs. semantic dementia: A case-series comparison. *Brain, 129,* 2132–2147.

Jefferies, E., Patterson, K., Jones, R. W., & Lambon Ralph, M. A. (2009). Comprehension of concrete and abstract words in semantic dementia. *Neuropsychology, 23,* 492–499. doi:10.1037/a0015452

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology, 66,* 115–124. doi:10.1037/a0026727

Kellas, G., Ferraro, F. R., & Simpson, G. B. (1988). Lexical ambiguity and the timecourse of attentional allocation in word recognition. *Journal of Experimental Psychology. Human Perception and Performance, 14,* 601–609.

Kieras, D. (1978). Beyond pictures and words: Alternative information-processing models for imagery effect in verbal memory. *Psychological Bulletin, 85,* 532–554. doi:10.1037/0033-2909.85.3.532

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language, 45,* 259–282.

Lambon Ralph, M. A., Sage, K., Jones, R., & Mayberry, E. (2010). Coherent concepts are computed in the anterior temporal lobes. *Proceedings of the National Academy of Sciences, 107,* 2717–2722.

Landauer, T. K. (2001). Single representations of multiple meanings in latent semantic analysis. In D. S. Gorfein (Ed.), *On the consequences of meaning selection: Perspectives on resolving lexical ambiguity* (pp. 217–232). Washington, DC: American Psychological Association.

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. doi:10.1037/0033-295X.104.2.211

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers, 28,* 203–208. doi:10.3758/BF03204766

McClelland, J. L., St. John, M. F., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language & Cognitive Processes, 4,* 287–335.

McDonald, S. A., & Shillcock, R. C. (2001). Rethinking the word frequency effect: The neglected role of distributional information in lexical processing. *Language and Speech, 44,* 295–323.

Metzler, C. (2001). Effects of left frontal lesions on the selection of context-appropriate meanings. *Neuropsychology, 15,* 315–328.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38,* 39–41.

Morton, J. (1979). Word recognition. In J. Morton & J. C. Marshall (Eds.), *Psycholinguistics 2: Structure and processes* (pp. 108–156). Cambridge, MA: MIT Press.

Noonan, K. A., Jefferies, E., Corbett, F., & Lambon Ralph, M. A. (2010). Elucidating the nature of deregulated semantic cognition in semantic aphasia: Evidence for the roles of the prefrontal and temporoparietal cortices. *Journal of Cognitive Neuroscience, 22,* 1597–1613.

Piercey, C. D., & Joordens, S. (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition, 28,* 657–666.

Rodd, J. M. (2004). The effect of semantic ambiguity on reading aloud: A twist in the tale. *Psychonomic Bulletin & Review, 11,* 440–445. doi:10.3758/BF03196592

Rodd, J. M., Davis, M. H., & Johnsrude, I. S. (2005). The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex, 15,* 1261–1269.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language, 46,* 245–266. doi:10.1006/jmla.2001.2810

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28,* 89–104. doi:10.1207/s15516709cog2801_4

Rogers, T. T., Lambon Ralph, M. A., Garrard, P., Bozeat, S., McClelland, J. L., Hodges, J. R., & Patterson, K. (2004). Structure and deterioration of semantic memory: A neuropsychological and computational investigation. *Psychological Review, 111,* 205–235. doi:10.1037/0033-295X.111.1.205

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

Rubenstein, H., Garfield, L., & Millikan, J. A. (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior, 9,* 487–494. doi:10.1016/S0022-5371(70)80091-3

Saffran, E. M., Bogyo, L., Schwartz, M. F., & Marin, O. S. M. (1980). Does deep dyslexia reflect right hemisphere reading? In M. Coltheart, J. C. Marshall, & K. E. Patterson (Eds.), *Deep dyslexia* (pp. 361–406). London, U.K.: Routledge & Kegan Paul.

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language, 27,* 499–520.

Schwanenflugel, P. J., & Shoben, E. J. (1983). Differential context effects in the comprehension of abstract and concrete verbal materials. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 82–102. doi:10.1037/0278-7393.9.1.82

St. John, M. F. (1992). The story gestalt: A model of knowledge-intensive processes in text comprehension. *Cognitive Science, 16,* 271–306. doi:10.1207/s15516709cog1602_5

Woollams, A. M. (2005). Imageability and ambiguity effects in speeded naming: Convergence and divergence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 878–890.

Yap, M. J., Tan, S. E., Pexman, P. M., & Hargreaves, I. S. (2011). Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review, 18,* 742–750.

Zeno, S., Ivens, S., Millard, R., & Duvvuri, R. (1995). *The educator's word frequency guide.* Brewer, NY: Touchstone.