What is semantic diversity and why does it facilitate visual word recognition?

Benedetta Cevoli[1], Chris Watkins[2] and Kathleen Rastle[1]

[1]Psychology Department, Royal Holloway, University of London

[2]Computer Science Department, Royal Holloway, University of London

Word Count: 4,141

RUNNING HEAD: WHAT IS SEMANTIC DIVERSITY?

Correspondence Address

Benedetta Cevoli

Department of Psychology

Royal Holloway University of London

Egham, TW20 0EX, UK

Email: Benedetta.Cevoli.2018@live.rhul.ac.uk

Abstract

Previous research has speculated that semantic diversity and lexical ambiguity may be closely related constructs. This research sought to test this claim. We computed multidimensional contextual representations of words using Latent Semantic Analysis and from these derived semantic diversity values for 28,555 words. Using existing data resources, we demonstrated that greater semantic diversity is associated with more rapid word recognition, particularly for low-frequency words, but found no evidence that effects of lexical ambiguity on word recognition could be ascribed to semantic diversity. Further analysis of the LSA-based contextual representations revealed that they do not capture the distinct meanings of ambiguous words. Instead, these contextual representations appear to capture general information about the topics and types of written material in which words occur. These analyses suggest that the semantic diversity metric facilitates word recognition (particularly for low frequency words) because high diversity words are likely to have been encountered no matter what one has read, whereas many participants may not have encountered lower diversity words simply because the topics and types of written material in which they occur are more restricted.

Key words: Semantic Diversity, Word Frequency, Lexical Ambiguity, Latent Semantic Analysis

Becoming a skilled reader involves the accumulation of experience with individual words. This experience is thought to be encoded in lexical representations and to contribute to word recognition. Most often, we think of lexical experience in terms of word frequency (i.e. the number of times that a word is encountered). It is well known that word frequency is a powerful determinant of word recognition time, with high frequency words recognized more rapidly than low frequency words (e.g. Forster & Chambers, 1973; see Brysbaert, Mandera, & Keuleers, 2018; Murray & Forster, 2004 for reviews).

The conceptualisation of lexical experience in terms of word frequency reflects a theoretical commitment about the nature of learning; specifically, that learning is strengthened through repetition. However, recent research has suggested that the accumulation of lexical experience is more nuanced than a simple count of one's encounters with individual words. Instead, this research suggested that *change* may be important for learning (Jones, Dye, & Johns, 2017), and that the variety of semantic and syntactic contexts in which words are encountered may therefore provide a superior conceptualisation of lexical experience (Nation, 2017).

One means of capturing contextual variation in lexical experience is through a construct known as semantic diversity (Hoffman, Lambon Ralph, & Rogers, 2013). The semantic diversity metric is calculated using Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), and reflects the average semantic similarity across all of the contexts in which a word occurs. Words high in semantic diversity occur in contexts that have lower similarity to one another than words low in semantic diversity. Previous research shows that this measure of semantic diversity facilitates word recognition in both adults (Hoffman & Woollams, 2015; Johns, Sheppard, Jones, & Taler, 2016) and children (Hsiao & Nation, 2018; Pagán, Bird, Hsiao, & Nation, 2019) beyond the effect of word frequency.

Most words in English (as for other languages) have multiple interpretations (Rodd, Gaskell, & Marslen-Wilson, 2002). Words that map onto two or more unrelated meanings (e.g. bark) are *homonyms*, while words characterized by multiple related senses (e.g. run) are *polysemes* (Rodd et al., 2002). Research has suggested that polysemous words are recognized faster and more accurately than unambiguous controls, while homonymous

words are recognized more slowly and less accurately than unambiguous controls (e.g., Armstrong & Plaut, 2016; Klepousniotou, Titone, & Romero, 2008; Rodd et al., 2002)

Much of the lexical ambiguity literature conceptualises words as falling into discrete categories (e.g. polysemous, unambiguous) based on the structure of dictionary entries (Klein & Murphy, 2001; Rodd et al., 2002) or subjective ratings (Hino, Lupker, & Pexman, 2002; Pexman, Hino, & Lupker, 2004). In contrast, semantic diversity has been offered as an alternative, computationally-derived measure of ambiguity that varies in a continuous manner (Hoffman et al., 2013) and which may be derived objectively from large text corpora. Indeed, if variation in the contextual usage of a word reflects variation in semantic meaning, then semantic diversity and lexical ambiguity might appear to be measuring the same thing. Consequently, researchers have recently speculated that "*the processing advantage for polysemous words in lexical decision might be related to the fact that polysemous words tend to be more semantically diverse*" (Hsiao & Nation, 2018, p. 115). However, to our knowledge, there has not yet been a direct investigation of the relationship between these constructs.

Though semantic diversity and lexical ambiguity might appear to be closely related, modelling dynamically changing meaning of words in context is challenging. According to Hoffman et al. (2013)'s methodology, the context of a word is the 1,000-word section of text in which it occurs, and the contextual representation of each word is modelled by the entire section of text containing the word. For example, if one section of corpus contains the sentence: "*The elephant played the moonlight sonata on the piano*", then the words *elephant*, *played*, *moonlight*, *sonata*, and *piano* will all have the same vector representation in this context. If Hoffman et al.'s methodology cannot distinguish the nuances of meaning that separate the words in this sentence, then it would be surprising indeed if it could capture the nuances of meaning that separate different related usages of polysemous words such as *run*, or different unrelated instances of ambiguous words such as *calf*.

Our work seeks to (a) release materials to compute LSA context vectors and semantic diversity, (b) investigate the impact of semantic diversity on word recognition in megastudies, and (c) test the relationship between semantic diversity and lexical ambiguity by determining whether semantic diversity accounts for behavioural effects of different types of lexical ambiguity. Following previous work (e.g. Hoffman et al., 2013; Hsiao & Nation, 2018), we derived multidimensional contextual representations of words using LSA,

and from these, computed semantic diversity. We then established the semantic diversity advantage on word recognition using data from the English Lexicon Project (ELP; Balota et al., 2007) and British Lexical Project (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012) megastudy databases. Finally, we turned to an investigation of whether semantic diversity is able to account for the effects of lexical ambiguity in two high-quality published studies for which materials were available (Armstrong & Plaut, 2016; Rodd et al., 2002). Our analyses suggest that semantic diversity could not account for the results of these published studies, and thus we conclude by investigating the nature of information captured through the semantic diversity metric.

Method

Distributional semantics models propose that a word's meaning may be derived from the contexts in which it occurs. Words within these models are represented as multi-dimensional vectors, and the distance or angle between vectors provides a measure of their similarity (e.g. Firth, 1957; Landauer & Dumais, 1997; Mikolov, Chen, Corrado, & Dean, 2013). Thus, these models provide a plausible means of characterising the distribution of a word's meaning in a continuous manner.

Previous implementations of the semantic diversity metric have used vectors derived from LSA operating on the British National Corpus (Hoffman et al., 2013; Hsiao & Nation, 2018; The British National Corpus, 2007). Although semantic diversity values have been made available previously (Hoffman et al., 2013), we are not aware of any open-access code that would allow psycholinguists to calculate a word's semantic diversity across different languages and corpora. Thus, we sought to replicate the procedures described by Hoffman et al. (2013) and Hsiao and Nation (2018) for computing a word's semantic diversity (see Figure 1 for illustration of the procedure). The code implementing these processing steps is available at https://osf.io/7hxvu/?view_only=2cfaf744136e444da2c9429db5359be2.

Our implementation of semantic diversity used the British National Corpus, a collection of 4,049 samples of written and spoken British English from a wide range of sources, from newspapers to popular fiction, and comprising 100 million words (British National Corpus Consortium, 2007). We selected only the 3,141 written documents and divided these into 1,000-word contexts. The final chunks from each document were excluded

because they may have included less than 1,000 words. We removed all non-alphabetic characters (e.g. digits, punctuation), as well as one-letter words and function words. Finally, we excluded any words that appeared less than 50 times in the entire corpus and in less than 40 contexts. These pre-processing steps resulted in 44,477 contexts and 28,555 words, which were used to build a co-occurrence matrix.

The co-occurrence matrix represents contexts in rows and words in columns, and thus reveals the distribution of particular words across different contexts and the clustering of different words in particular contexts. We applied a log entropy weighting to normalise this co-occurrence matrix, and reduced its dimensionality with singular value decomposition (Berry, Dumais, & O'Brien, 1995). Semantic diversity for a particular word is computed by measuring the pairwise cosine similarity between each of the word's 300-dimensional context vectors, averaging these values, and then applying a log-transform and sign reversal.

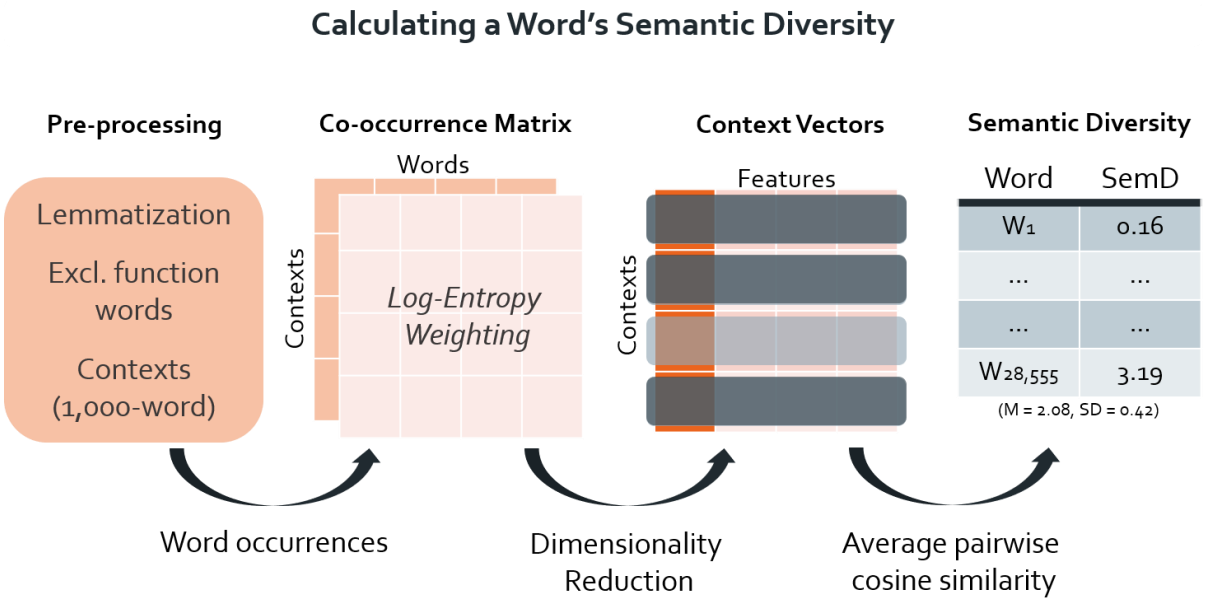


Figure 1. Illustration of Semantic Diversity Procedure.

Our implementation followed Hsiao and Nation (2018) in using a lemmatized version of the British National Corpus and in excluding function words, while Hoffman et al. (2013) used an inflected version of the same corpus and included function words. Previous literature suggests that these procedural differences should not substantially change the nature of the semantic diversity metric (Hsiao & Nation, 2018) or the performance of semantic vector models (Bullinaria & Levy, 2012). However, in order to make sure that our implementation replicated Hoffman et al. (2013), we also computed semantic diversity using the Hoffman et al. (2013) version of the corpus. The resulting semantic diversity values (using lemmatized and inflected versions of the corpus) correlate highly with each other ($r = 0.93$). Nonetheless, both estimates showed a lower correlation than expected with the semantic diversity measures provided by Hoffman et al. (2013). Specifically, the measures that we computed with the inflected corpus had a correlation of $r = 0.67$ with the ones provided by Hoffman et al. (2013), while the measures computed with the lemmatized corpus had a correlation of $r = 0.73$.

These correlations are too low given that (a) we replicated the pre-processing procedure described by Hoffman et al. (2013) exactly; and (b) we used the same corpus as Hoffman et al. (2013) for one of the semantic diversity implementations. Although we cannot provide a definitive explanation for the discrepancy since the code used by Hoffman et al. (2013) is not available, we suspect they did not weight their co-occurrence matrix by the singular values. Indeed, if we change our implementation to compute semantic space coordinates using unweighted singular vectors, the correlation between our measures and those of Hoffman et al. (2013) increases substantially ($r=0.98$ for inflected corpus; $r=0.93$ for lemmatized corpus; see Figure 2). Since the output of LSA is generally weighted by the singular values, for the remainder of analyses reported in this article, we retained our original semantic diversity measures, computed with the lemmatized corpus and weighted accordingly.

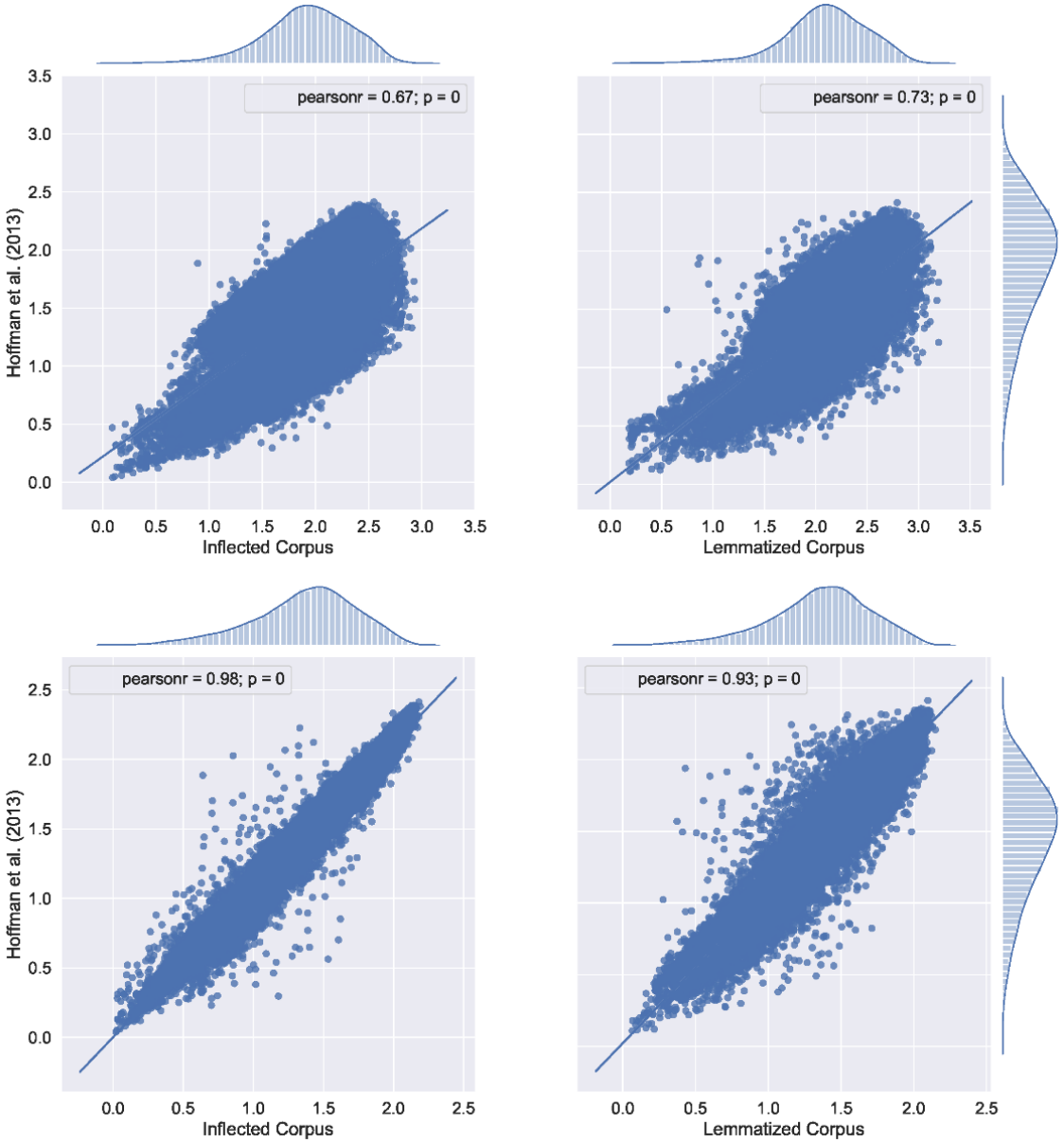


Figure 2. Scatter plots of resulting semantic diversity measures on x-axes and the norms reported by Hoffman et al. (2013) and the y-axes. On the left, values obtained following the pre-processing procedure described in the methods (lemmatized corpus, exclusion of stop words etc.), while, on the right, values obtained following Hoffman et al. (2013) pre-processing procedure. On the top row are the measures obtained with the classical output of LSA (weighting by the singular values) while on the bottom row are the measures obtained without considering the singular values.

Results

To validate our newly computed measures, we first assessed the impact of semantic diversity on lexical decision and reading aloud latencies within the English Lexicon Project (ELP; Balota et al., 2007) and the British Lexicon Project (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012). The ELP consists of trial-level lexical decision and reading aloud data for

40,481 words collected from 444 participants, while the BLP consists of trial-level lexical decision data for 28,730 words from 78 participants. Semantic diversity measures for 28,555 words were computed following the corpus analysis described in the previous section, while word frequency estimates were retrieved from the Subtlex-UK corpus (Van Heuven, Mandera, Keuleers, & Brysbaert, 2014).

We used linear mixed effects models to examine the effect of semantic diversity and its interaction with word frequency on lexical decision and reading aloud data. Analyses were conducted using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2018). Models included semantic diversity, word frequency and their interaction as fixed effects, while participant and item were included as random intercepts. We also added word length and trial number as fixed factors. To reduce autocorrelation effects from previous trials (Baayen & Milin, 2010), models included fixed effects of previous trial accuracy and latency. Only correct word trials were included in reaction time (RT) analyses, and data points with absolute standardized residuals exceeding 2.5 standard deviations were removed (based on log-transformed RTs; Baayen & Milin, 2010). For visualisation purposes, model estimates were obtained through the package Effects (Fox & Hong, 2009) and transformed RT data were transformed back to raw RTs for ease of interpretation. P-values were estimated using the Satterthwaite approximation for degrees of freedom (lmerTest; Kuznetsova, Brockhoff, & Christensen, 2017).

We observed significant facilitatory effects of both semantic diversity and frequency on reaction time and accuracy data (see **Error! Reference source not found.** for summary of results), as well as significant interactions between frequency and semantic diversity, indicating that the effect of semantic diversity is greater for low frequency words than for high frequency words (see Figure 3).

Database	Task	Data	Predictors	Estimate	Std. Error	t value	*p*-value
BLP	LD	Accuracy	Freq	4.48	0.02	83.77	**< 0.001**
			SemD	1.07	0.02	4.31	**< 0.001**
			Freq*SemD	0.93	0.02	-4.85	**< 0.001**
		Reaction Times	Freq	-0.12	< 0.01	-96.72	**< 0.001**
			SemD	< 0.01	< 0.01	-2.74	**< 0.01**
			Freq*SemD	< 0.01	< 0.01	3.94	**< 0.01**
ELP	LD	Accuracy	Freq	2.72	< 0.01	9650	**< 0.001**
			SemD	1.13	< 0.01	1525	**< 0.001**
			Freq*SemD	0.97	< 0.01	-864	**< 0.01**
		Reaction Times	Freq	-0.07	< 0.01	-96.13	**< .001**
			SemD	- 0.01	< 0.01	-10.36	**< .001**
			Freq*SemD	< 0.01	< 0.01	2.08	**< 0.05**
ELP	Naming	Accuracy	Freq	0.74	0.01	-52.73	**< 0.001**
			SemD	0.02	0.01	-5.11	**< 0.001**
			Freq*SemD	- 0.03	0.01	3.12	**< 0.05**
		Reaction Times	Freq	< 0.01	< 0.01	-66.15	**< 0.001**
			SemD	- 0.05	< 0.01	-5.11	**< 0.001**
			Freq*SemD	< 0.01	< 0.01	3.12	**< 0.01**

Table 1. Summary of results.

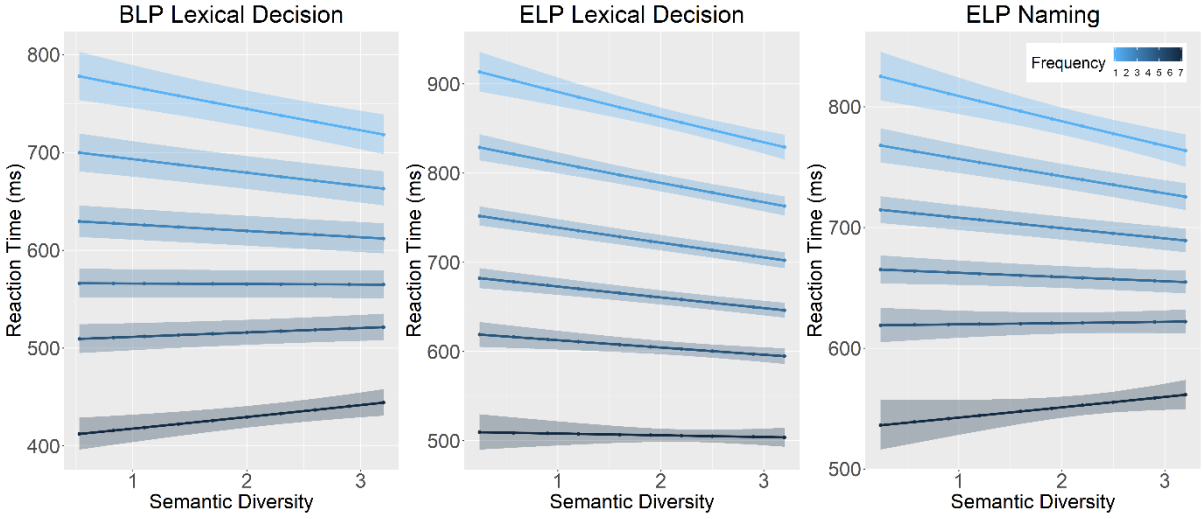


Figure 3. Model estimates of the effect of semantic diversity by frequency on reaction time data as a function of database.

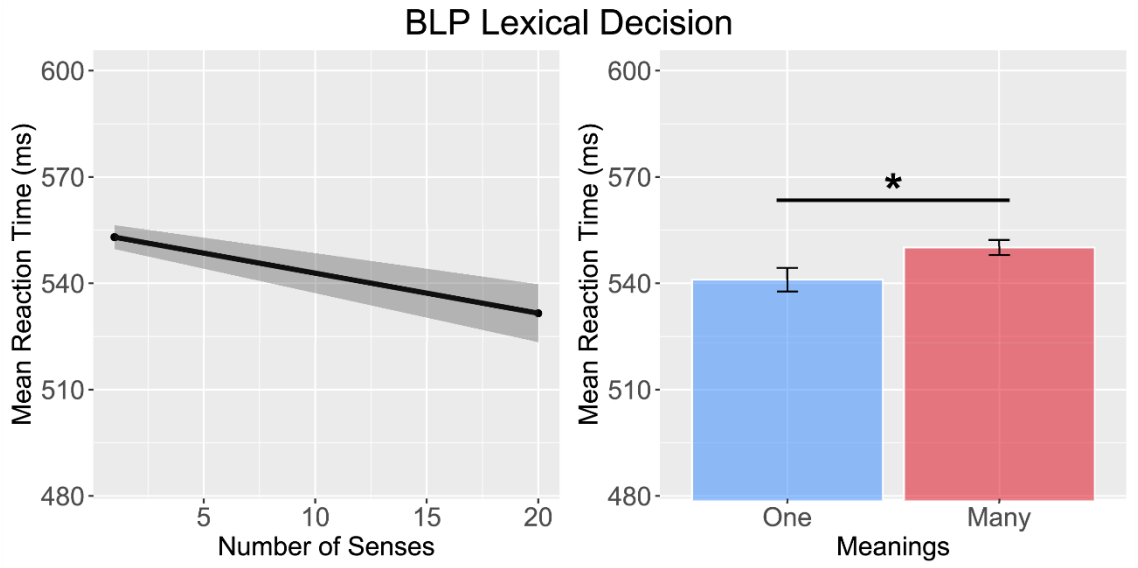
We next turned to investigate the relationship between lexical ambiguity and semantic diversity. We selected two prominent studies reporting differences in processing polysemous and homonymous word compared to unambiguous controls in visual lexical decision (Armstrong & Plaut, 2016; Rodd et al., 2002), and sought to replicate these using

response time and accuracy measures from the BLP and ELP lexical decision data, and then using our newly computed semantic diversity measures.

Simulation 1 - Rodd et al. (2002)

Stimuli were selected from two experiments of Rodd et al. (2002), one of the first visual lexical decision studies reporting contrasting effects of polysemy and homonymy on lexical decision performance. Based on the structure of dictionary entries, Rodd et al. (2002) used the number of a word's meanings and senses as proxies of homonymy and polysemy, respectively.

In their first experiment, Rodd et al. (2002) used a regression design to investigate the impact of multiple meanings and multiple senses on word recognition. They observed that word recognition was slowed when words were characterised by multiple meanings but speeded when words were characterised by multiple senses. This combination of results is also observed in the BLP (number of meanings, $\beta = 0.02, SE = 0.01, t = 1.97, p < 0.05$; number of senses, $\beta = -0.01, SE < 0.01, t = -2.23, p < 0.05$) and in the ELP (number of meanings, $\beta = 0.03, SE = 0.01, t = 2.5, p < 0.05$; number of senses, $\beta = -0.01, SE < 0.01, t = -1.96, p < 0.05$; see Figure 4). However, our analyses revealed that these effects could not be ascribed to semantic diversity. Semantic diversity values did not differ for these items on number of meanings ($\beta = 0.05, SE = 0.04, t = 1.06, p = 0.29$) or number of senses ($\beta = 0.02, SE = 0.02, t = 0.95, p = 0.34$; see Figure 5).



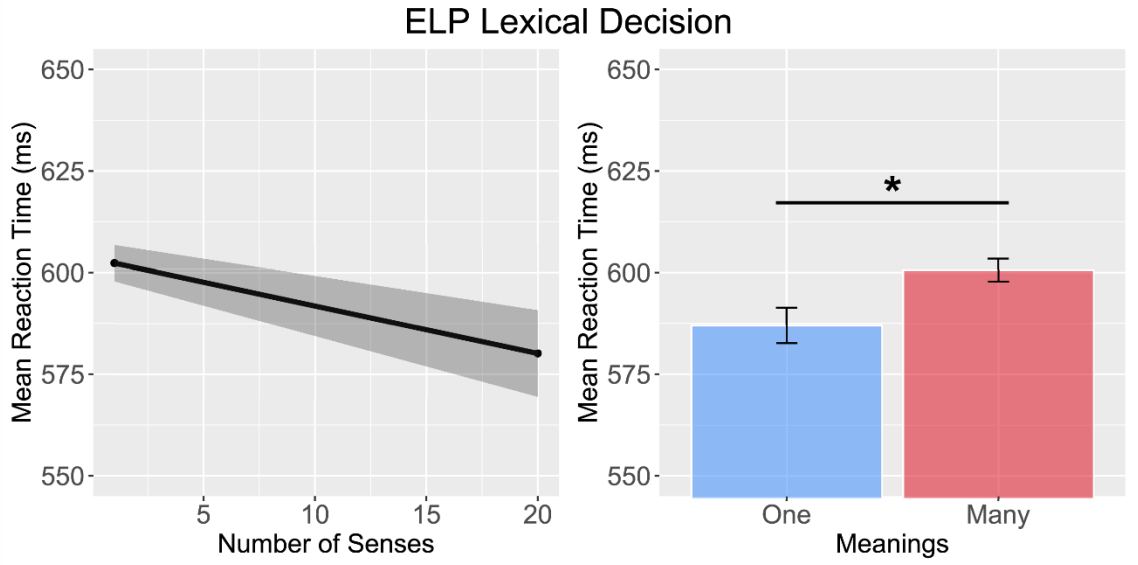


Figure 4. Results of the simulation analysis of Experiment 1 of Rodd et al. (2002) on reaction time data of BLP and ELP. Both datasets show that increasing number of senses speeds performance while increasing number of meanings slows performance.

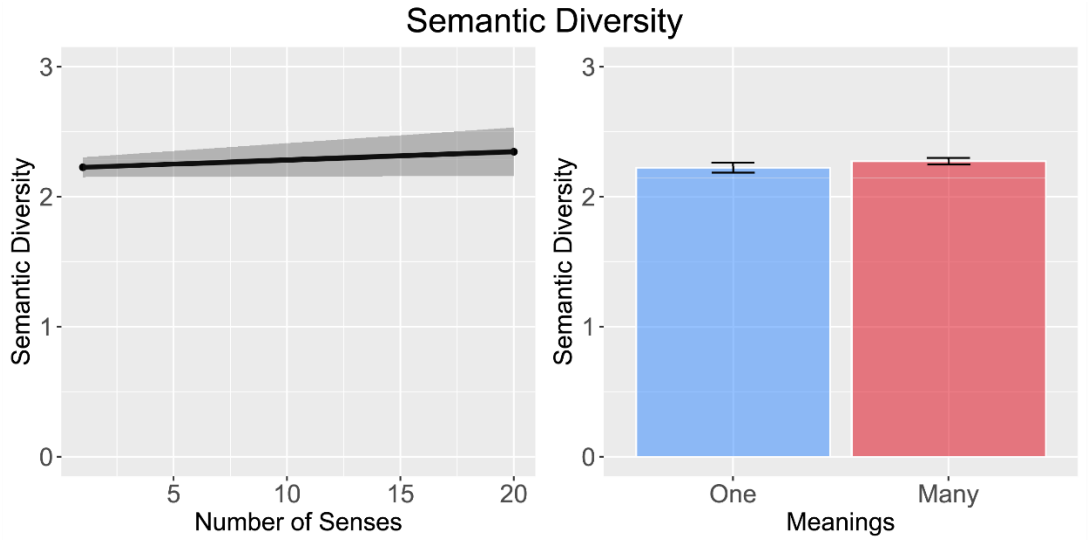


Figure 5. Results of the simulation analysis of Experiment 1 of Rodd et al. (2002) showing no difference in semantic diversity for words with many or few senses and meanings.

In a second experiment, Rodd et al. (2002) used a factorial design manipulating number of senses and number of meanings and reported only a significant effect of the former. Following the statistical analysis pipeline reported in Rodd et al. (2002), lexical decision data from the BLP and ELP revealed a significant main effect of the number of senses on response time (BLP: $F_1(1,4726) = 17.40, p < 0.001$; $F_2(1,121) = 12.55, p < 0.001$; $\Delta RT = 15\ ms$; ELP: $F_1(1,3763) = 7.23, p < 0.01$; $F_2(1,121) = 6.33, p < 0.05$; $\Delta RT =$

16 ms; see Figure 6). There was no effect of number of meanings in the BLP ($F_1(1,4735) = 3.5, p = 0.06$; $F_2(1,121) = 3.2, p = 0.07$; $\Delta RT = 8\,ms$) or the ELP ($F_1(1,3763) = 2.99, p = 0.08$; $F_2(1,121) = 2.22, p = 0.13$; $\Delta RT = 8\,ms$). Once again, while Rodd et al.'s (2002) data were perfectly captured in the BLP and ELP, the pattern reported could not be ascribed to semantic diversity. Semantic diversity values did not differ for number of senses ($F(1,121) < 0.001, p = 0.99$) or number of meanings ($F(1,121) = 0.38, p = 0.53$) for these items (see Figure 7).

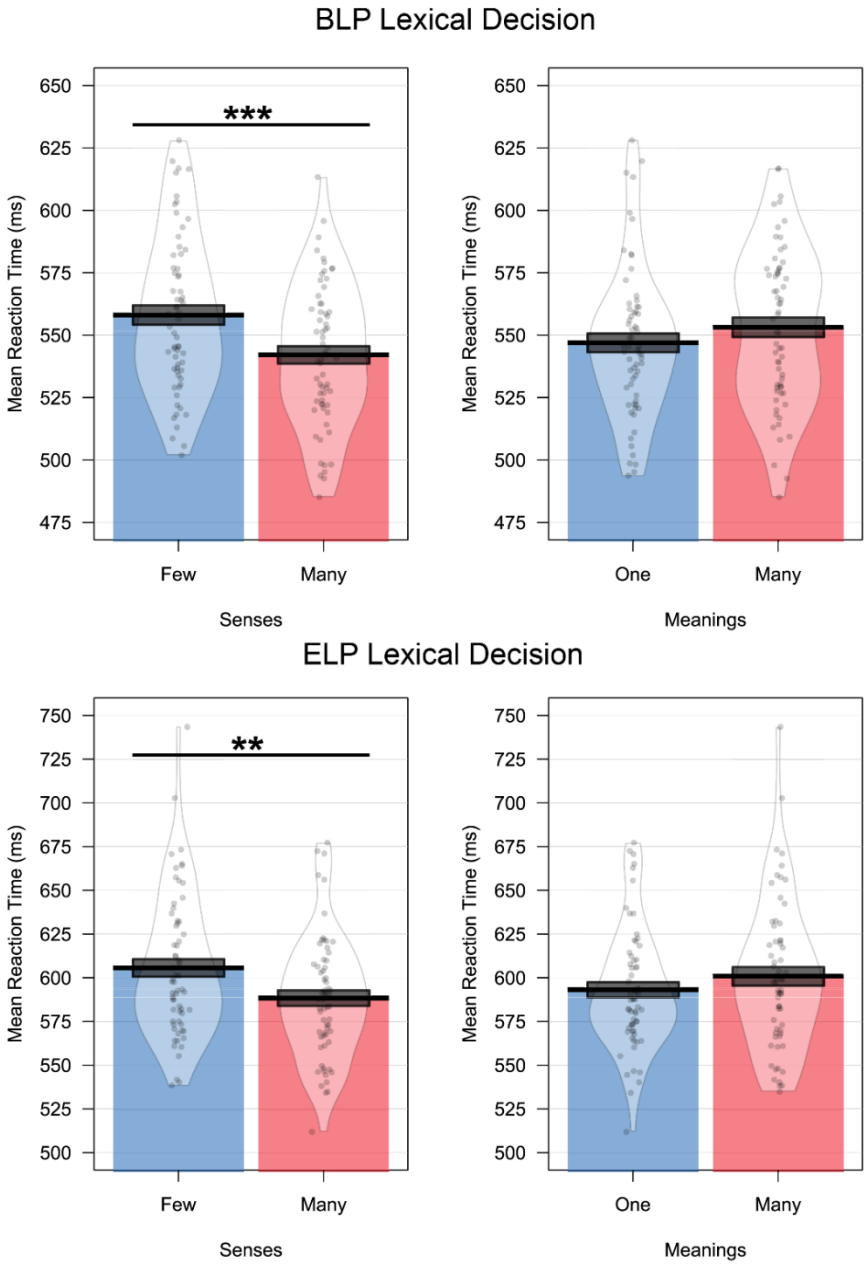


Figure 6. Results of the simulation analysis of Experiment 2 of Rodd et al. (2002) on response time data from the BLP and ELP. Data show that an increased number of senses speeds lexical decision latency, but that there is no effect of the number of meanings.

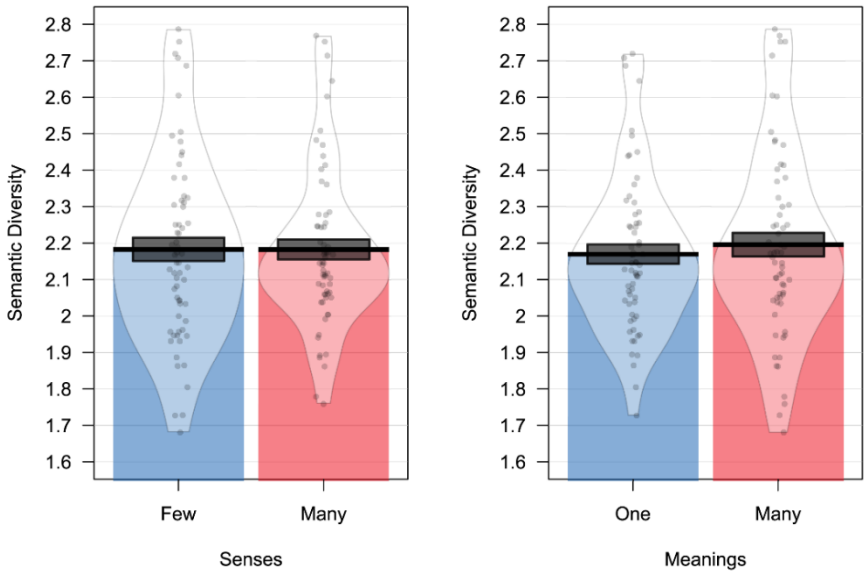


Figure 7. Results of the simulation analysis of Experiment 2 of Rodd et al. (2002) showing no difference in semantic diversity values for words with many or few senses or meanings.

Simulation 2 - Armstrong and Plaut (2016)

Armstrong and Plaut (2016) investigated whether the polysemy advantage and homonymy disadvantage found in visual lexical decision are modulated by task difficulty and stimulus contrast. Task difficulty was manipulated by varying the wordlikeness of nonwords in the lexical decision task. Armstrong and Plaut (2016) observed that the polysemy advantage reduced while the homonymy disadvantage increased as task difficulty increased. The authors argued that these findings may help us to understand why in standard lexical decision tasks (which usually corresponds to the lower task difficulty condition), the homonymy disadvantage is weak or completely absent, while the polysemy advantage is consistently reported (see Armstrong & Plaut, 2016 for discussion and Eddington & Tokowicz, 2015 for a review of ambiguity literature). Thus, a similar pattern of results is expected to be found in lexical decision data of the BLP and ELP.

Results from our analysis of the BLP and ELP also showed faster recognition of polysemous words relative to unambiguous controls (BLP: $b = 0.02, SE < 0.01, t = -5.91, p < 0.01$; ELP: $b = -0.02, SE = 0.01, t = -270, p < 0.01$). No significant difference was observed between homonymous and unambiguous words (BLP: $\beta < 0.01, SE < 0.01, t = 1.33, p = 0.18$; ELP: $b = -0.11, SE = 0.28, t = -0.39, p = 0.69$;see Figure 8). However, again

none of these effects are observed in the semantic diversity measures. There was no significant difference in semantic diversity between unambiguous and polysemous words ($\beta < 0.01, SE = 0.01, t = -0.37, p = 0.7$); and while there was a near-significant difference in semantic diversity between unambiguous and homonymous words ($\beta = -0.02, SE < 0.01, t = -1.93, p = 0.05$), it was in the opposite to predicted direction (see Figure 9).

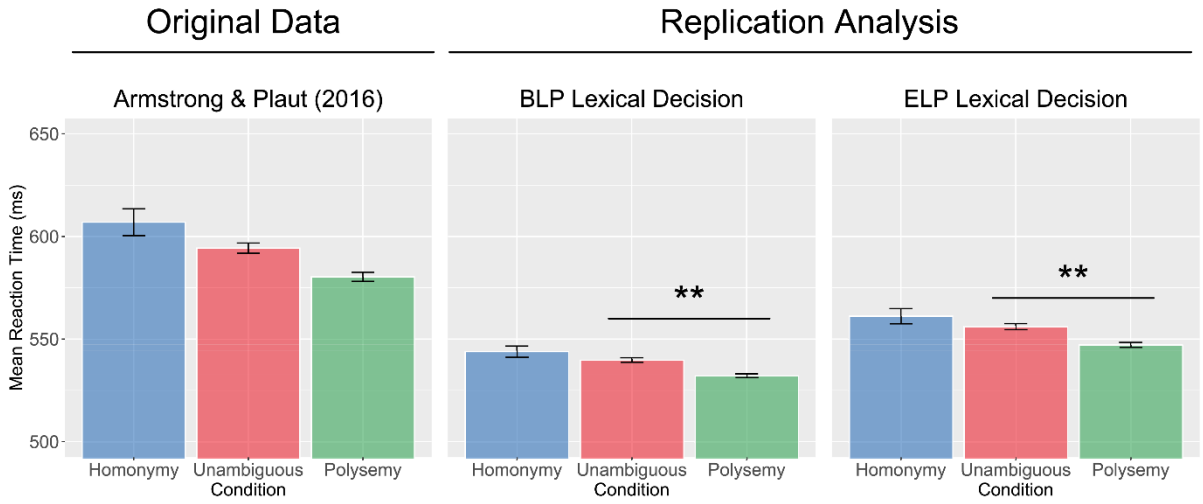


Figure 8. Descriptive bar plots of response time data (left) by type of ambiguity (pooled between all experimental conditions) as reported by Armstrong & Plaut (2016) and bar plots of replication analysis of BLP and ELP (middle and right, respectively) showing a polysemy advantage but no homonymy disadvantage.

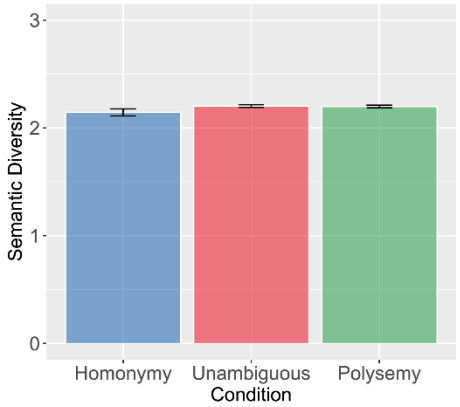


Figure 9. Results of the simulation analysis of Armstrong & Plaut (2016) on semantic diversity measures showing no difference across ambiguity type.

Though the effects of polysemy and homonymy reported by Rodd et al. (2002) and Armstrong and Plaut (2016) were also observed in the BLP and ELP, there was no evidence that these effects could be ascribed to semantic diversity. This result is inconsistent with the claim that semantic diversity provides a continuous measure of the multiple senses and meanings with which words are used in different contexts (Hsiao & Nation, 2018; Hoffman

et al., 2013). In the discussion, we consider more fully what semantic diversity is and why it facilitates visual word recognition.

Discussion

Previous research has proposed that semantic diversity and lexical ambiguity are closely related. However, our analyses suggest that LSA-based measures of semantic diversity do not seem to capture differences between homonymous, polysemous and unambiguous words. These results may suggest that these different forms of words are not characterised by differences in contextual variation, although this seems unlikely (e.g. that *bank* would not be characterised by greater contextual variation than *perjury*). The other possibility is that the LSA-based measure of semantic diversity does not capture this contextual variation. Yet, if this is the case then it is unclear what semantic diversity is capturing or why it facilitates word recognition.

One potential explanation is that, as a measure of central tendency, semantic diversity does not reflect the *distribution* of a word's contexts, and consequently is unable to differentiate between ambiguous and unambiguous words. That is, it may be that the context vectors of ambiguous words such as *bank* show greater variation than those of unambiguous words, but that the averaging process in the calculation of semantic diversity masks this variation. However, it is also possible that the context vectors themselves are insensitive to the contextual meanings of words. LSA has been extensively used as a topic model aiming at organizing and summarizing large collections of written text by automatically identifying abstract topics (text classification purposes and recommender systems; Evangelopoulos, Zhang, & Prybutok, 2012; Landauer et al., 2007). However, much less is known about the extent to which LSA captures the contextual nature of semantic content. Therefore, the nature of information represented within these context vectors requires exploration.

To investigate these possibilities, we selected three examples of highly ambiguous words from Rodd et al. (2002), *calf*, *mole*, and *pupil*, and manually labelled a random 50% of the contexts in which each word occurred within the corpus used to derive our context vectors. For the word *calf*, for example, we decided whether each occurrence related to an

animal, a body part, or some other meaning. We then visualised the labelled context vectors using the t-Distributed Stochastic Neighbour Embedding (t-SNE) technique for dimensionality reduction (Van Der Maaten & Hinton, 2008). By visualising the contexts in this manner, we sought to determine whether (a) the context vectors do indeed capture contextual variation but that the averaging within the semantic diversity metric fails to reflect this; or (b) the context vectors are insensitive to this semantic variation.

It is immediately apparent from Figure 10 that the LSA-context vectors of the word *calf* are not represented in distinct clusters (as would be expected due to its unrelated meanings), but are instead spread widely across the semantic space. The same pattern holds for the distinct meanings of *mole* and *pupil*. To quantitatively assess whether there is evidence that the context vectors are representing distinct semantic clusters in the multidimensional LSA space, we computed a Calinski-Harabasz score for each sample word. This score reflects variance between and within clusters; higher scores indicate superior goodness of fit with defined clusters (Caliński & Harabasz, 1974). Relatively low scores were found for all three examples (3.28, 2.08 and 4.43 for *calf*, *mole*, and *pupil*, respectively). These are similar to the scores derived when the same 50% of contexts were assigned the three possible labels randomly (0.92, 1.13, and 0.97, respectively). These data suggest that LSA-based context vectors are not sensitive to the contextual meanings of ambiguous words, and thus the failure to capture lexical ambiguity effects with the semantic diversity metric lies in the modelling approach itself rather than the computation of the semantic diversity metric. This outcome also raises the important questions of what information is captured by LSA context vectors, and why this metric appears to facilitate word recognition.

Figure 10. t-SNE plots of the context vectors in which of the word calf occurs.

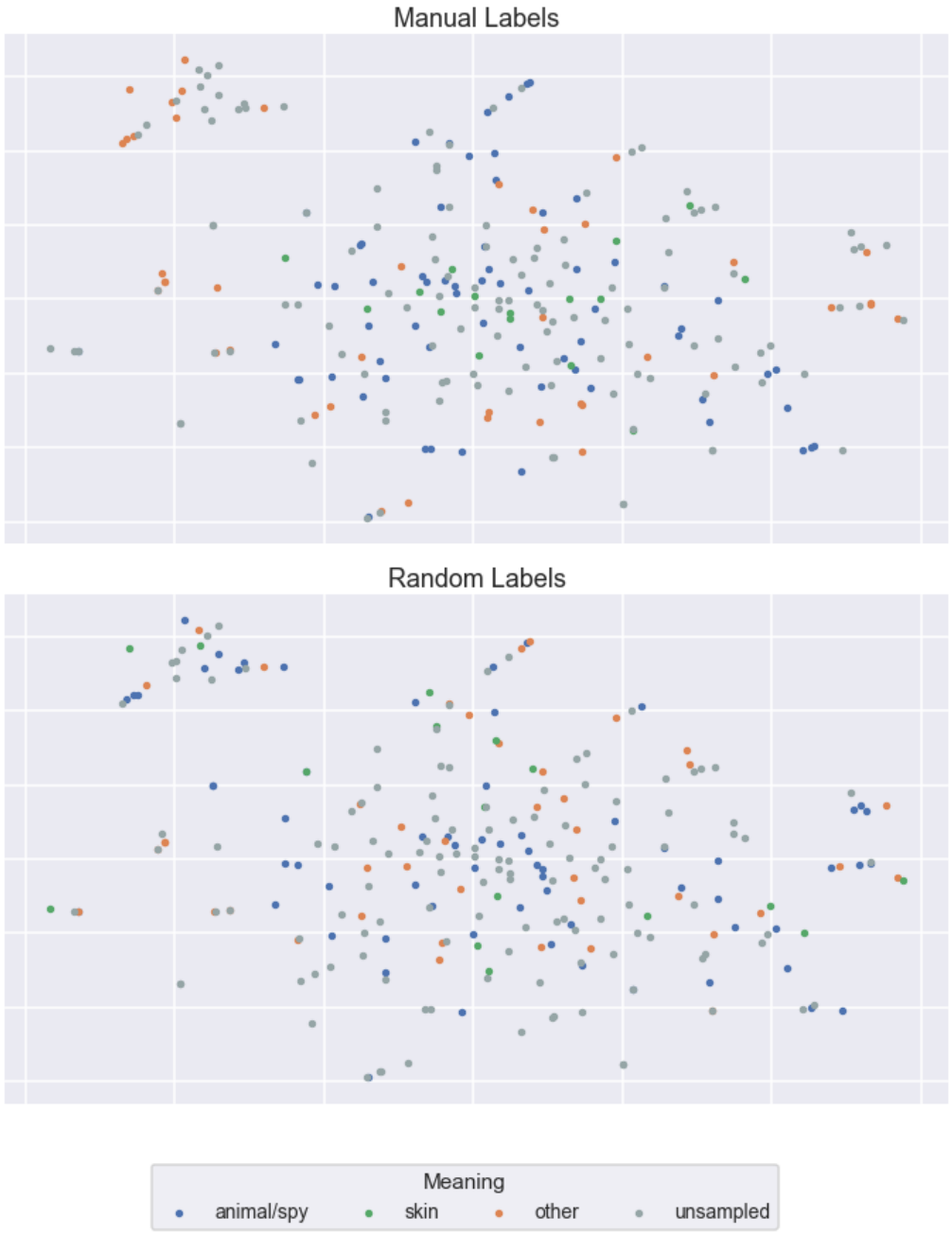


Figure 11. t-SNE plots of the context vectors in which of the word mole occurs.

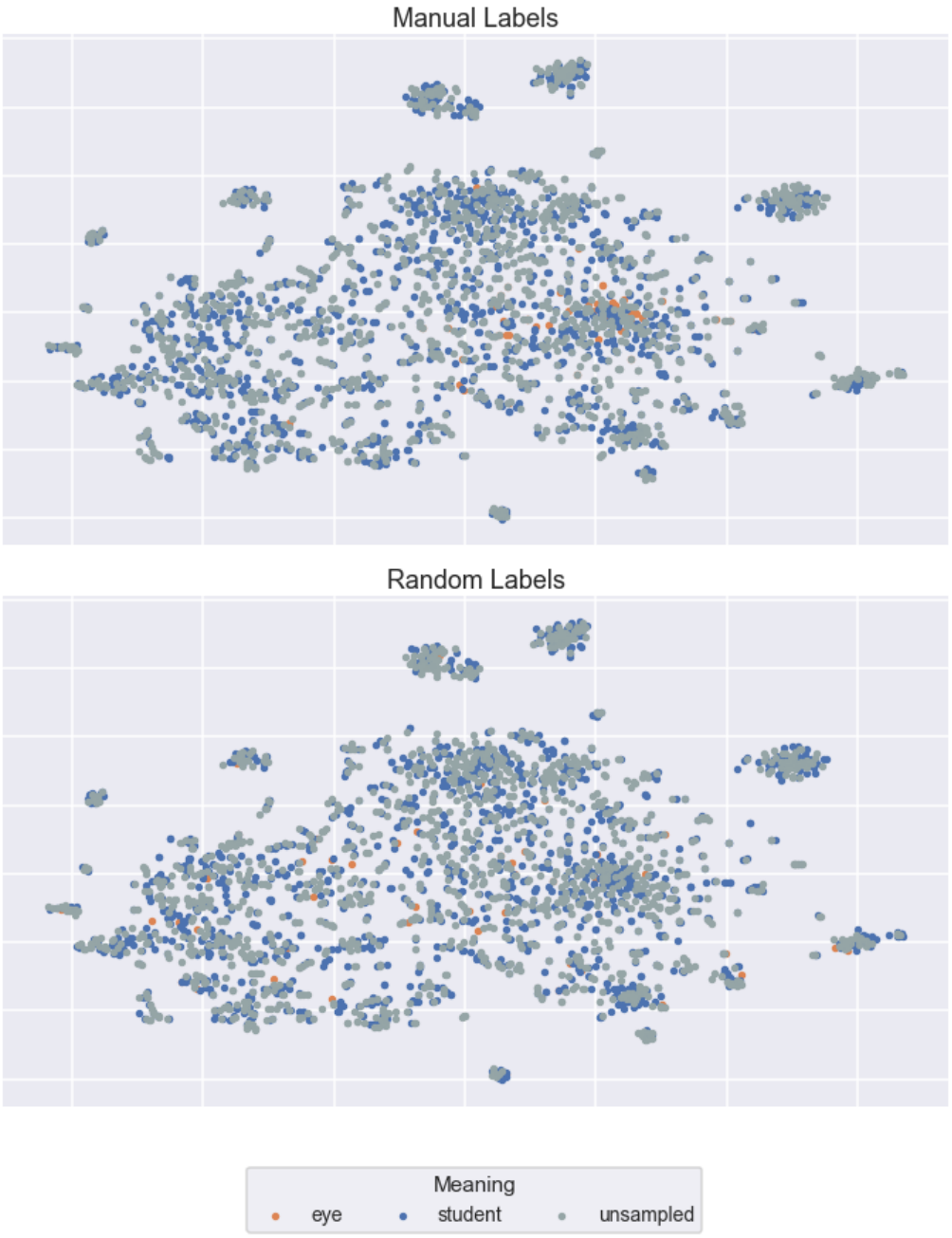


Figure 12. t-SNE plots of the context vectors in which of the word pupil occurs.

To understand more fully what LSA-context vectors represent, we labelled every context within our corpus using metadata pertaining to the general domain of the contexts (e.g. natural science, world affairs) and the type of written material in which the contexts appear (e.g. fiction, newspaper). These data are visualised in Figures Figure ***13*** and Figure ***14***. It is apparent that contexts cluster well along these dimensions and this is confirmed by the Calinski-Harabasz scores comparing clustering based on general domain (320.74) relative to random allocation (M = 1.00, SD = 0.04 for 1,000 iterations), and clustering based on type of

written material (301.12) relative to random allocation (*M* = 1.00, *SD* = 0.05 for 1,000 iterations). These data suggest that the LSA context vectors are capturing general properties about how words occur in a corpus, but not capturing information about the nature of word meaning.

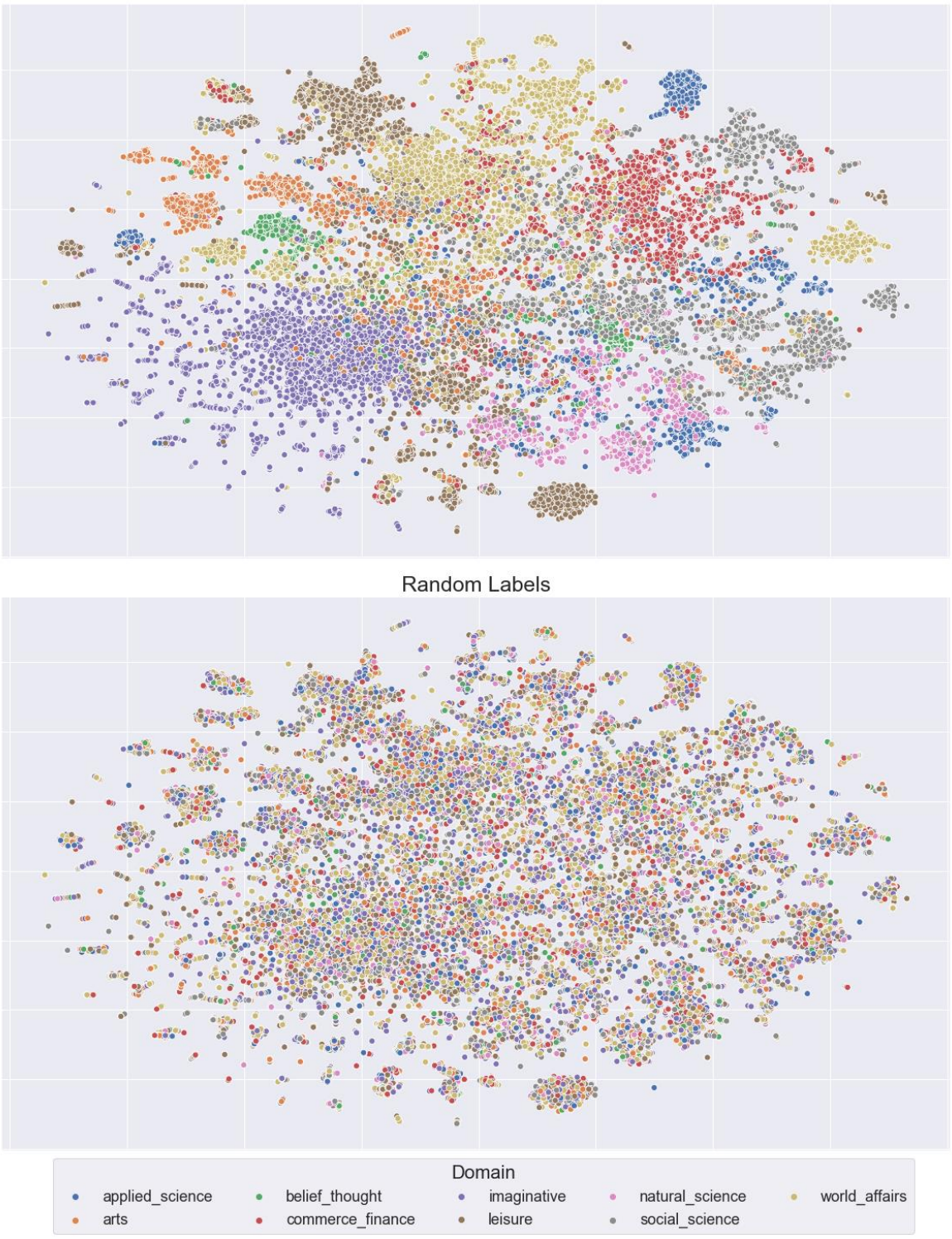


Figure 13. t-SNE plots of the whole corpus labelled by domain on the top (variance ratio: 320.74), while on the bottom are the same labels randomly assigned for comparison (M = 1.00, SD = 0.04 for 1,000 iterations).

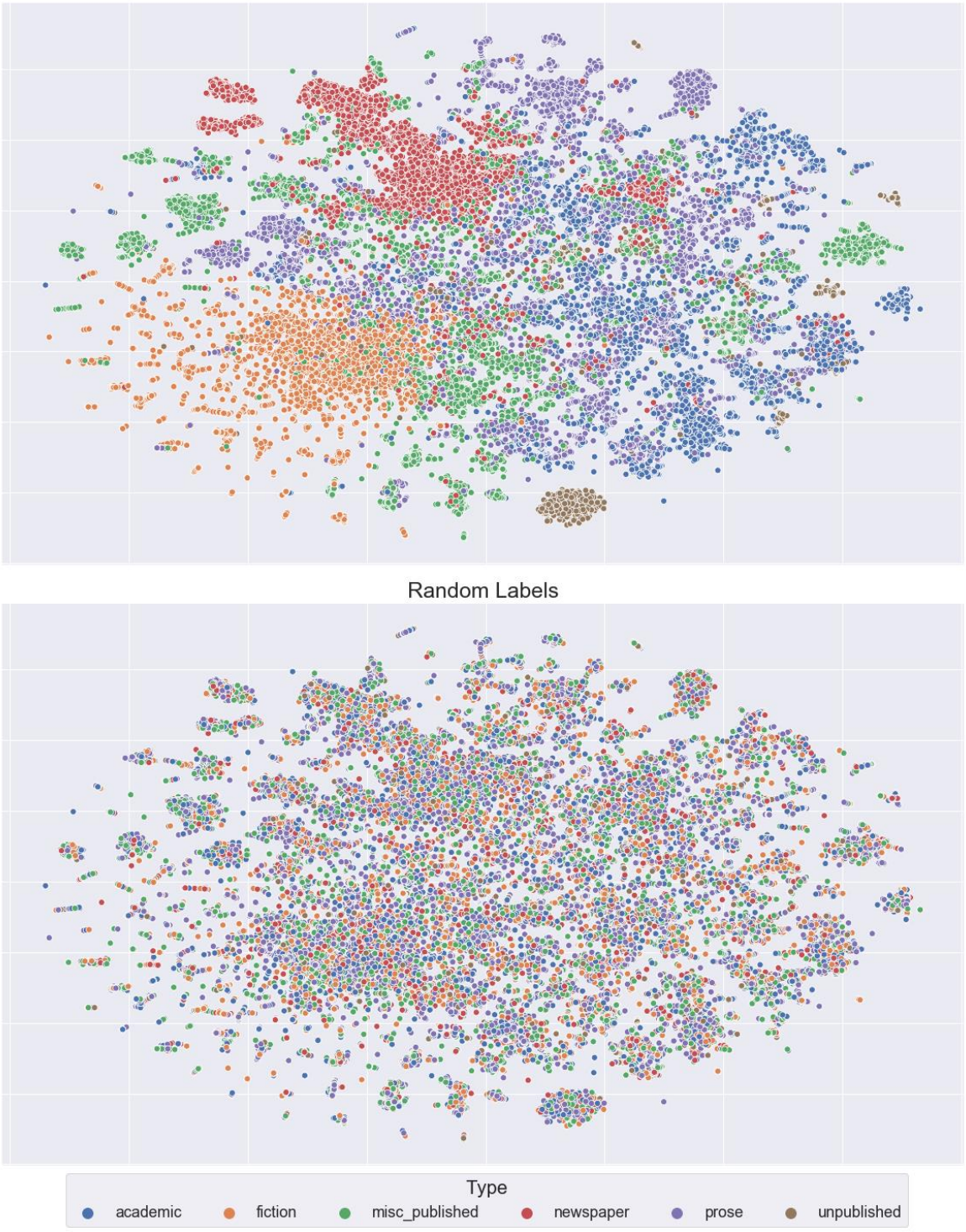


Figure 14. t-SNE plots of the whole corpus labelled by type of written material on the top (variance ratio: 301.12), while on the bottom are the same labels randomly assigned for comparison (M = 1.00, SD = 0.05 for 1,000 iterations).

These analyses lead us to suggest that semantic diversity is a measure of a word's spread across topics and types of contexts, rather than a measure of the diversity of a word's

contextual usage. The metric defined by Hoffman et al. (2013) is insensitive to the diversity of a word's meanings but instead it is capturing general information about the range of reading situations in which a word might be encountered. Words that are high in semantic diversity are well-distributed across topics and types of contexts, while words that are low in semantic diversity are specific to particular contexts. Thus, we propose this metric should instead be referred to as *textual diversity*.

This proposal has important theoretical implications for understanding the beneficial effect of semantic diversity on word recognition. We suggest that textual diversity is related to the probability that a word will be encountered, and it is for this reason that it facilitates word recognition (particularly for low frequency words). Words that are high in textual diversity (e.g. diverge) are spread across topics and types of material and will therefore be encountered irrespective of what is read. In contrast, words that are low in textual diversity (e.g. crampon) arise only in specific topics or types of material, and therefore some readers may almost never encounter them if they do not read about certain specialised topics. For these readers, these words may be very rare or never encountered, so that performance in word recognition tasks suffers disproportionately on average. The impact of textual diversity may be less relevant for high-frequency words since these are likely to be encountered irrespective of the topic or type of material in which they occur.

To summarise, we sought to investigate the relationship between semantic diversity and lexical ambiguity. We implemented new LSA-based context vectors from which we derived the semantic diversity metric, and we demonstrated that this metric is associated with the speed of word recognition and reading aloud. Despite previous speculation that semantic diversity and lexical ambiguity are closely associated, we found no evidence that semantic diversity could explain the effects of lexical ambiguity on word recognition. Further analysis of the LSA-based context vectors used to derive the semantic diversity metric revealed that they do not capture information about the different contextual meanings of individual words, and instead appear to encode more general information about the manner in which words occur within a corpus. Thus, we proposed the term textual diversity as a better fit for describing the semantic diversity metric defined by Hoffman et al. (2013). These findings have important theoretical implications for understanding why the semantic diversity metric facilitates word recognition.

The field of natural language processing has seen exceptionally rapid development in the last twenty years, providing a variety of state-of-art techniques that might be more suitable for modelling the distribution of the semantic contents of individual words (Young, Hazarika, Poria, & Cambria, 2017). Future work using more up-to-date models has the potential to capture contextual variation across different words, and ultimately, to help us to understand more deeply the nature of lexical experience.

Acknowledgements

We thank Paul Hoffman and Yaling Hsiao for their support in calculating the semantic diversity metric. We also thank Blair Armstrong for his support in accessing data from Armstrong and Plaut (2016) and for his insights pertaining to these research themes.

Open Practices Statement

Data, materials and code for computing contextual representations and the semantic diversity metric are available at (https://osf.io/7hxvu/?view_only=2cfaf744136e444da2c9429db5359be2). None of this work was preregistered.

References

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966. https://doi.org/10.1080/23273798.2016.1171366

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12. https://doi.org/10.21500/20112084.807

Balota, D. A., Yap, M. J., Hutchison, K. A., Cortese, M. J., Kessler, B., Loftis, B., … Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*(3), 445–459. https://doi.org/10.3758/BF03193014

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/doi: 10.18637/jss.v067.i01

Berry, M. W., Dumais, S. T., & O'Brien, G. W. (1995). Using linear algebra for intelligent information retrieval. *SIAM Review*, *37*(4), 573–595. https://doi.org/10.1137/1037127

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The Word Frequency Effect in Word Processing: An Updated Review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Bullinaria, J. A., & Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD. *Behavior Research Methods*, *44*(3), 890–907. https://doi.org/10.3758/s13428-011-0183-8

Caliński, T., & Harabasz, J. (1974). A Dendrite Method Foe Cluster Analysis. *Communications in Statistics*, *3*(1), 1–27. https://doi.org/10.1080/03610927408827101

Evangelopoulos, N., Zhang, X., & Prybutok, V. R. (2012). Latent semantic analysis: Five methodological recommendations. *European Journal of Information Systems*, *21*(1), 70–86. https://doi.org/10.1057/ejis.2010.61

Firth, J. (1957). A Synopsis of Linguistic Theory 1930-1955. In *Studies in Linguistic Analysis*. Oxford: Philological Society.

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Fox, J., & Hong, J. (2009). Effect Displays in R for Multinomial and Proportional-Odds Logit Models: Extensions to the effects Package. *Journal of Statistical Software*, *32*(1), 1–24. Retrieved from http://www.jstatsoft.org/v32/i01/

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*(4), 686–713. https://doi.org/10.1037/0278-7393.28.4.686

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730. https://doi.org/10.3758/s13428-012-0278-x

Hoffman, P., Rogers, T. T., & Ralph, M. A. L. (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*(9), 2432–2446. https://doi.org/10.1162/jocn.2011.21614

Hoffman, P., & Woollams, A. M. (2015). Opposing effects of semantic diversity in lexical and semantic relatedness decisions. *Journal of Experimental Psychology. Human Perception and Performance*, *41*(2), 385–402. https://doi.org/10.1037/a0038995

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114–126. https://doi.org/10.1016/J.JML.2018.08.005

Johns, B. T., Sheppard, C. L., Jones, M. N., & Taler, V. (2016). The Role of Semantic Diversity in Word Recognition across Aging and Bilingualism. *Frontiers in Psychology*, *7*, 703. https://doi.org/10.3389/fpsyg.2016.00703

Jones, M. N., Dye, M., & Johns, B. T. (2017). Context as an Organizing Principle of the Lexicon. *Psychology of Learning and Motivation*, *67*, 239–283. https://doi.org/10.1016/bs.plm.2017.03.008

Keuleers, E., Lacey, P., Rastle, K., & Brysbaert, M. (2012). The British Lexicon Project: Lexical decision data for 28,730 monosyllabic and disyllabic English words. *Behavior Research Methods*, *44*(1), 287–304. https://doi.org/10.3758/s13428-011-0118-4

Klein, D. E., & Murphy, G. L. (2001). The Representation of Polysemous Words. *Journal of Memory and Language*, *45*, 259–282. https://doi.org/10.1006/jmla.2001.2779

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making Sense of Word Senses: The Comprehension of Polysemy Depends on Sense Overlap. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1534–1543. https://doi.org/10.1037/a0013012

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models . *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

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*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of Latent Semantic Analysis*. *Handbook of Latent Semantic Analysis*. Lawrence Erlbaum Associates Publishers.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. Retrieved from http://arxiv.org/abs/1301.3781

Murray, W. S., & Forster, K. I. (2004). Serial mechanisms in lexical access: The rank hypothesis. *Psychological Review*, *111*(3), 721–756. https://doi.org/10.1037/0033-295X.111.3.721

Nation, K. (2017). Nurturing a lexical legacy: reading experience is critical for the development of word reading skill. *Npj Science of Learning*, *2*(1), 3. https://doi.org/10.1038/s41539-017-0004-7

Pagán, A., Bird, M., Hsiao, Y., & Nation, K. (2019). Both Semantic Diversity and Frequency Influence Children's Sentence Reading. *Scientific Studies of Reading*, 1–9. https://doi.org/10.1080/10888438.2019.1670664

Pexman, P. M., Hino, Y., & Lupker, S. J. (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning Memory and Cognition*, *30*(6), 1252–1270. https://doi.org/10.1037/0278-7393.30.6.1252

R Core Team. (2018). R: A Language and Environment for Statistical Computing. Vienna, Austria.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language*, *46*(2), 245–266.

https://doi.org/10.1006/jmla.2001.2810

The British National Corpus. (2007). Version 3 (BNC XML Edition). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium.

Van Der Maaten, L., & Hinton, G. (2008). *Visualizing Data using t-SNE*. *Journal of Machine Learning Research* (Vol. 9).

Van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2017). Recent Trends in Deep Learning Based Natural Language Processing. Retrieved from http://arxiv.org/abs/1708.02709