AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

2023, Vol. 152, No. 9, 2578–2590
https://doi.org/10.1037/xge0001389

Bigram Semantic Distance as an Index of Continuous Semantic Flow in Natural Language: Theory, Tools, and Applications

Jamie Reilly[1, 2], Ann Marie Finley[1, 2], Celia P. Litovsky[1, 2], and Yoed N. Kenett[3]
[1] Eleanor M. Saffran Center for Cognitive Neuroscience, Temple University
[2] Department of Communication Sciences and Disorders, Temple University
[3] Technion—Israel Institute of Technology, Israel Institute of Technology

Much of our understanding of word meaning has been informed through studies of single words. High-dimensional semantic space models have recently proven instrumental in elucidating connections *between* words. Here we show how bigram semantic distance can yield novel insights into conceptual cohesion and topic flow when computed over continuous language samples. For example, "Cats drink milk" is comprised of an ordered vector of bigrams (cat-drink, drink-milk). Each of these bigrams has a unique semantic distance. These distances in turn may provide a metric of dispersion or the flow of concepts as language unfolds. We offer an R-package ("semdistflow") that transforms any user-specified language transcript into a vector of ordered bigrams, appending two metrics of semantic distance to each pair. We validated these distance metrics on a continuous stream of simulated verbal fluency data assigning predicted switch markers between alternating semantic clusters (animals, musical instruments, fruit). We then generated bigram distance norms on a large sample of text and demonstrated applications of the technique to a classic work of short fiction, *To Build a Fire* (London, 1908). In one application, we showed that bigrams spanning sentence boundaries are punctuated by jumps in the semantic distance. We discuss the promise of this technique for characterizing semantic processing in real-world narratives and for bridging findings at the single word level with macroscale discourse analyses.

Keywords: semantic distance, semantic memory, language

Semantic memory operates both at the microscale level in representing the meanings of individual concepts and at the macroscale level when constructing meaning between concepts (Hills & Kenett, 2022; Kumar, 2021). Much of our understanding of conceptual knowledge has been informed through language-based paradigms involving the production and/or comprehension of single words or meticulously controlled arrays of words. This pattern is evident across a wide range of experimental tasks such as blocked cyclic naming, semantic decision, lexical decision, priming, and picture–word interference (Binney, Ashaie, et al., 2018; Capitani et al., 2003; Cutler, 1981; Farah & McClelland, 1991; Funnell et al., 2006; Grossman et al., 2004; Hillis & Caramazza, 1991; Hodges et al., 1996; Kousta et al., 2011; Lupker, 1979; Pexman et al., 2017; Warrington, 1975; Woollams et al., 2008). Advantages gained in experimental control can, however, come at a cost to ecological validity. People do not communicate using single words. Language is an emergent system whose elements often combine in nonlinear and unpredictable ways to convey meaning at different levels of discourse (Marelli et al., 2017; Price et al., 2015, Westerlund & Pylkkänen, 2014).

Techniques leveraged from natural language processing have recently facilitated more widespread use of connected discourse (e.g., storybooks, podcasts, corpora) in studies of semantic processing (Baldassano et al., 2017; Deniz et al., 2019; Günther et al., 2019;

This article was published Online First April 20, 2023.
Jamie Reilly https://orcid.org/0000-0002-0891-438X
Portions of this work appeared as a preprint on PsyArXiv (https://psyarxiv.com/6fuhv) and were presented at the annual conference of the Academy of Aphasia (2022).

We are grateful to Bonnie Zuckerman for her assistance with programming. We thank the creators of the Lancaster Sensorimotor Norms, AffectVec, and GloVe for developing open-source tools in the best spirit of science. We also thank Professor Mark Davies for his generosity in allowing us to distribute word embeddings trained on the CoCA.

This work was supported by the U.S. National Institute on Deafness and Other Communication Disorders (NIH/NIDCD, DC013063). High performance computing resources supported by U.S. National Science Foundation (162506) and the U.S. Army Research Laboratory (W911NF-16-2-0189). This work was also partially supported by the United States–Israel Binational Science Foundation (2021040).

The authors have no known conflict of interest.

All data and associated codes are available via the Open Science Foundation at https://osf.io/ryhfj/. All tools developed here are open source.

Jamie Reilly served as lead for conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, supervision, writing–original draft, and writing–review and editing. Ann Marie Finley served in a supporting role for methodology and writing–review and editing. Celia P. Litovsky served in a supporting role for methodology and writing–review and editing. Yoed N. Kenett served in a supporting role for conceptualization, methodology, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Jamie Reilly, Department of Communication Sciences and Disorders, Temple University, 1701 N. 13th Street, Philadelphia, PA 19122, United States. Email: reillyj@temple.edu

Hartung et al., 2020; Huth et al., 2016; Jain & Huth, 2018; Johnson et al., 2022; Kumar, 2021; Kumar et al., 2022; Mandera et al., 2015, 2017; Nastase et al., 2020; Popham et al., 2021). A key advantage of such approaches is their capacity to extract distributional statistics (e.g., patterns of co-occurrence) about language by indexing vast numbers of words appearing in real-world corpora (Baldassano et al., 2017; de Heer et al., 2017; Hartung et al., 2020; Huth et al., 2012; Jain & Huth, 2018; Naselaris et al., 2011; Popham et al., 2021; Simony et al., 2016) to construct high-dimensional models of word meaning. Such models have fueled numerous recent advances across a wide range of psychological and linguistic sciences (Beaty et al., 2021; Beaty & Johnson, 2021; Gray et al., 2019; Hills & Kenett, 2022; Johnson et al., 2022; Kenett, 2018, 2019; Kenett et al., 2017; Kumar et al., 2020; Olson et al., 2021), as well as within computational, clinical, and cognitive neuroscience (Anderson et al., 2019; Fernandino et al., 2016, 2022, Fu et al., 2023; Kenett & Faust, 2019).

What Is Semantic Distance?

Semantic distance reflects the similarity (or dissimilarity) between two or more concepts distributed across an *n*-dimensional space, typically derived from analyzing large corpora of texts (Günther et al., 2019). There is no upward limit to the number of potential dimensions that comprise a semantic space. A researcher with an interest in the arousal of curse words, for example, might quantify differences in the arousal ratings of curse words versus neutral words (e.g., Reilly et al., 2020). In this simple example, *arousal* constitutes a one-dimensional semantic space.

Semantic distance is a relative rather than an absolute construct. That is, semantic distance can only be interpreted relative to the unique semantic space used to derive that measure. Cognitive (neuro)scientists are typically interested in constructing semantic spaces that are psychologically and/or neurobiologically plausible (Binder et al., 2016; Crutch et al., 2013; Reilly, Finley, et al., 2021; Sacchett & Humphreys, 1992). An existential challenge for semantic space models is that the true dimensionality of human semantic memory is latent. My semantic network is qualitatively different than yours, and any viable model of semantic memory must have the flexibility to accommodate these differences (Kumar, 2021).

Two broad classes of semantic space models have risen to prominence over the past decade. Experiential models classify concepts along subjectively experienced dimensions (e.g., color, shape), such that words with similar characteristics are more semantically linked. In contrast, word embedding models are predicated upon the idea that in natural language, words that occur together are likely to be semantically related. Both models approach the challenge of specifying dimensionality in fundamentally different ways. In the following sections, we describe both types of models and how they capture meaning through multidimensional vectors.

Experiential Semantic Models

A core assumption underlying all semantic space models is that word meaning can be decomposed into numerous dimensions or features (for early iterations of feature-based approaches see Cree & McRae, 2003; Rosch, 1973). In *experiential* models, raters are typically asked to explicitly rate the salience of many target words across numerous orthogonalized dimensions. For example, a researcher might ask people to rate the salience of artichokes on color, visual imagery, and aggression. These ratings rely on our own subjective experience and as such, have been termed *experiential* (Binder et al., 2016; see also Wingfield & Connell, 2022). Binder et al. (2016) have proposed perhaps the most extensive experiential model to date, characterizing words along 65 sensorimotor, affective, and interoceptive dimensions. In an earlier model known as the Abstract Conceptual Feature space, Crutch et al. characterized English words along 12 dimensions (Crutch et al., 2013; Reilly et al., 2016; Troche et al., 2017). Both the Binder et al. (2016) and Crutch et al. (2013) experiential models tend to produce clusters that mirror classic Linnaean taxonomies. That is, similarities in color, sound, valence, danger, body morphology, and other traits form constellations of intercorrelated features that bound natural categories (see also Cree & McRae, 2003; Garrard et al., 2001; McRae et al., 1999; Rogers et al., 2004). In an experiential semantic space, dogs and wolves have similar vector representations since they highly overlap in shape, color, sound, and other characteristics. However, experiential models do not typically account for contextual or thematic relatedness predicated upon co-occurrence. For this type of semantic relation, we turn to word embedding models, which quantify regularities in the linguistic contexts in which words co-occur (i.e., are embedded).

Word Embeddings, Context, and Co-Occurrence

In contrast to experiential semantic models, embedding models capture semantic similarity based on shared environments. For example, dogs, collars, bones, frisbees, and leashes are all semantically bound via shared contexts. We learn through repeated exposure that the conditional probability of encountering a collar in the context of a dog is high. Common word embedding models such as LSA (Landauer & Dumais, 1997), Word2Vec (Mikolov et al., 2013), Global Vectors for Word Representation (GloVe; Pennington et al., 2014), ELMo (Peters et al., 2017), BERT (Devlin et al., 2018), and GPT-3 (Brown et al., 2020) attempt to capture this co-occurrence in language by modeling abstract vector representations such that words that co-occur more frequently have more similar vector representations.[1] In embedding models, dogs and collars are thus closely linked (with similar vector representations), whereas dogs and platypuses, which do not frequently co-occur, have more dissimilar vector representations. Yet dogs and platypuses share many features (e.g., both are mammals), and thus are likely to share similar *experiential* semantic vector representations.

Word embeddings are typically derived through machine learning algorithms that require no subjective human judgments. Such models are typically trained on vast language corpora (i.e., collections of structured text such as books, newscasts, transcribed podcasts, or Twitter feeds). The outcome of this process is that each discrete token (e.g., word, *n*-gram, phrase) in a training corpus is characterized by a multidimensional array. However, unlike the labeled dimensions that comprise experiential semantic models, the

[1] Many contemporary transformer models (e.g., BERT, GPT-3) use a combination of lexical co-occurrence and deep learning to extract semantic vectors. Thus, references to embedding models as "co-occurrence models" are a misnomer.

dimensions derived from word embeddings are abstract mathematical constructs.

Inter-Word Semantic Distance: A Continuous Bigram Model

Semantic space models have featured prominently in many recent language and neuroimaging studies (Binder et al., 2016; Bonner & Epstein, 2021; Crutch et al., 2013; Marelli et al., 2017; Wingfield & Connell, 2022). However, to our knowledge, these models have not been applied to evaluate the flow of semantic information in continuous language. Consider, for example, the following sentence:

The quick brown fox jumped over the lazy dog.

After omitting closed class words (e.g., *the*), this sentence is composed of five lemmatized bigrams, including (a) quick-brown; (b) brown-fox; (c) fox-jump; (d) jump-lazy; and (e) lazy-dog. When a semantic distance is assigned to each bigram, a vector of distances emerges. This vector is technically a continuous time series reflecting word-by-word conceptual shifts over the course of any language sample.

Many previous studies investigating semantic distance have employed "continuous bag of words" (or CBOW) approaches that either abandon order information or alternatively analyze semantic distance for paradigms where computing continuous bigram distance would be senseless (e.g., word association). However, under a sequential bigram model, order is a critical factor reflecting the dispersion of concepts over time. For example, a person experiencing severe mania or delirium might show markedly high semantic distance perceived by a listener as incoherent. In contrast, low variability in the semantic distance could mark a narrowly focused, boring, or repetitive story. We recently reported evidence of such effects using a "bag of bigrams" approach to spoken narratives produced by people with aphasia relative to age-matched controls (Litovsky et al., 2022). People with aphasia showed significantly lower bigram semantic distances relative to controls, and semantic distances correlated strongly with offline neuropsychological measures of semantic memory functioning.

Our aims in the current study were to describe the development and implementation of a freely available R-package designed to read, clean, tokenize, and append two novel metrics of semantic distance (i.e., experiential vs. embedding) to any continuous language sample. We derived norms for each of these semantic spaces and conducted a validation study demonstrating how jumps in the semantic distance can mark cluster boundaries in a continuous stream of simulated category fluency data (i.e., blocks of musical instruments, fruits, animals). Finally, we demonstrate several novel applications of this continuous bigram model to a renowned work of short fiction, *To Build a Fire* (London, 1908).

Method

Overview

We first derived two novel semantic spaces (experiential and embedding) that provide the foundation for an open-source R-package, titled "semdistflow." This R-package reads, formats, and then transforms any language transcript into a running vector of pairwise semantic distances. We derived distance norms for each semantic space, validated the spaces on a continuous vector of alternating semantic categories, and conducted a series of simulations on a work of short fiction. All scripts and data used here are available for download and use at https://osf.io/ryhfj/.

Derivation of a Feature-Based Semantic Space (SemDist15)

We created a 15-dimension experiential semantic space characterizing English words across a subset of sensorimotor features from the Lancaster Sensorimotor Norms (Lynott et al., 2020) and social–emotional features from the AffectVec word sentiment norms (Raji & da Melo, 2020). The Lancaster norms reflect crowd-sourced salience ratings for 40,000 English words on a 6-point Likert scale. AffectVec reports intensity ratings for 70,000 English words on a 0–1 scale across 239 affective dimensions. We extracted the following sensorimotor dimensions from the Lancaster norms: visual, auditory, gustatory, haptic, interoceptive, olfactory, and hand-arm. We extracted the following social-emotional dimensions from AffectVec: excitement, surprise, happiness, fear, anger, contempt, disgust, and sadness. Since these variables reflect different ranges and measurement scales, we z-transformed each individual dimension relative to its own mean and standard deviation. These procedures yielded a vector of z-scores reflecting the salience of each word. Throughout the remainder of this article, we refer to distance norms generated from this 15-dimension semantic feature space (hereafter SemDist15) as *experiential*.

Derivation of an Embedding Semantic Space

We derived an embedding space using a well-established model, GloVe (Pennington et al., 2014), trained on documents corresponding to the written text[2] within the Corpus of Contemporary American English (CoCA; Davies, 2009). We first split the corpus into a test set ($N = 4{,}273$ documents) and a training set ($N = 212{,}737$ documents) and omitted all words appearing fewer than 5 times. We then trained the model using the text2vec R-package (Selivanov, 2020). At a learning rate of 0.05, the embedding model converged after 22 iterations resulting in a 300-dimension vector space spanning 394,115 unique words. We trimmed this large data file (2.2 Gb) by isolating only those entries with a corresponding lemma listed in the Subtlex-US word frequency norms (Brysbaert & New, 2009), resulting in 60,384 words, each characterized across 300 hyperparameters. We hereafter refer to these embeddings as the Global Vectors of *Written* Contemporary American English (GloWCA). Throughout the remainder of this article, we refer to distance norms generated from GloWCA as *embedding*.

Text Cleaning Algorithm

To apply vector representations to each word in a language sample, we first isolated all content words and reduced them to their respective lemma forms. This cleaning algorithm involved multiple steps (see Table 1) followed by lemmatization of the remaining content words. Readers are invited to inspect all individual commands by

[2] In NLP research, a *document* is typically defined as a discrete language sample. Examples of documents include novels, podcast transcripts, blog entries, or transcriptions of spoken language samples. We focused on English text and excluded all documents corresponding to spoken English transcriptions.

Table 1
Text Stripping, Vectorizing, and Global Formatting

Target	Description of global action (substitution or omission)
Contractions	Replaced/extended contractions (e.g., it's → it is)
Letter case	All text converted to lowercase
Stopwords	Omitted closed class words (e.g., the, a, is) using a custom stopword list ($N = 1{,}104$ words) modified from the System for the Mechanical Analysis and Retrieval of Text stopword list
Nonalphabetic characters	Omitted all punctuation, symbols, emojis, whitespace, and other nonalphabetic characters
Numbers	Omitted all cardinal and ordinal numbers
Morphological derivatives	Lemmatized the text to transform all words into their corresponding dictionary entries

Note. In addition to base R, packages used in the various stages of text preparation included TextStem (V0.1.4; Rinker, 2018a) and Textclean (V0.9.3; Rinker, 2018b).

visiting the R-package vignette for "semdistflow".[3] Table 1 represents some of the primary procedures we implemented in this cleaning algorithm. The final product of these text cleaning and formatting procedures is a vector of sequential bigrams (i.e., each word paired with the next content word in the language sample) to which we append our experiential semantic space (SemDist15) and embedding semantic space (GloWCA) distance values.

Establishing Norms for Semantic Distance

In this analysis, we calculated semantic distances in a large sample of naturally occurring sequential bigrams relative to randomly paired words to establish norms for semantic distance in natural language. We calculated semantic distance via the cosine similarity measure. Cosine is a scalar value that measures the similarity of two vectors' angles. For example, to calculate the experiential semantic distance of "quick" and "brown," we calculated the cosine similarity of the 15-dimensional vectors of "quick" and "brown" represented in SemDist15. We transformed all semantic distances by subtracting the cosine similarity from 1. This procedure constrained each semantic distance to a range between zero and two, such that more dissimilar words were associated with greater distances, and more similar words were associated with smaller distances.

After processing the 4,273 texts in the CoCA test set using our text cleaning algorithm, we extracted 250,000 sequential bigrams to establish norms for semantic distances in naturally occurring language. In this sequential model, we derived estimates of experiential (SemDist15) and embedding (GloWCA) semantic distances. As a comparison (random model), we yoked the initial word of each bigram to another randomly selected word in the test set, yielding nonsequential bigrams.

Validation Using Simulated Category Fluency Data

Semantic fluency tasks typically involve producing as many exemplars of a given category (e.g., animals) as possible over a fixed interval (Christensen & Kenett, 2021). People commonly employ foraging and other search strategies such as clustering and switching when producing a string of exemplars (Binney, Zuckerman, et al., 2018; Troyer, 2000; Troyer et al., 1998; Ovando-Tellez et al., 2022). For example, a category such as animals might evoke a spontaneous clustering and switching strategy involving house pets switching to jungle animals, marine animals, etc. Successful verbal fluency, therefore, requires both effective semantic search within categories, as well as fluid executive functioning as needed to flexibly disengage when a particular category has been exhausted (Ovando-Tellez et al., 2022). Verbal fluency has accordingly emerged as one of the most common metrics of executive and semantic processing employed in clinical neuropsychology.

The overarching structure of semantic category fluency in terms of alternating blocks of semantic clusters may offer a unique opportunity to validate our proposed continuous bigram model. We reasoned that semantic distance within semantic clusters will be low (e.g., dog, cat, hamster) relative to semantic distances for bigrams crossing switch boundaries (e.g., "dog–cat—**hamster | saxophone**—piano–trumpet…"). Thus, large jumps in semantic distance can potentially delineate switches in a continuous stream of category fluency data (for related work on segmenting fluency data using word embedding models see Alacam et al., 2022; Lundin et al., 2022).

We first generated a vector of 7,500 continuous words composed of alternating 10-word blocks of animals, musical instruments, and fruits/vegetables randomly sampled with replacement from fixed lists (see OSF for lists). This vector was, therefore, composed of 750 switches and 6,750 within cluster exemplars, providing a fixed reference for exactly where switches occur (i.e., every 10th word), We applied the distme() function to the unlemmatized ordered word list, generating pairwise semantic distances for every running bigram. We then scaled (z-scored) the resultant distributions and recoded each running word pair as potentially either a within-category cluster (coded as 0) or as a predicted switch between categories (coded as 1) using a threshold of $z > 1$. We examined concordance between the actual distribution of switches (every 10th word) versus the predicted distribution of switches (marked by z-score distance jumps) using a variety of signal detection metrics from the "verification" package (NCAR - Research Applications Laboratory, 2015) of R.

To Build a Fire *(Jack London, 1908)*

We hypothesized that sequential bigrams as naturally occurring within the structured text are more semantically related than random bigrams. Semantic distance should on average be lower for sequential bigrams relative to randomly constructed bigrams (i.e., each word in the story paired with a random word in another database). We tested this hypothesis in the context of *To Build a Fire*, a novella published by the American author, Jack London, in 1908. This famous story ($N = 7{,}125$ words) depicts a man hiking alone through the boreal forest of the Yukon Territory. The man is followed by a native dog (described as a *wolf dog*) unfazed by the deep snow and cold (−75°F). The dog casually follows the man from a distance as his attempts at self-preservation by building a fire become increasingly desperate. The man ultimately succumbs to hypothermia, and the story ends with the dog wandering off indifferently into the forest.

We first imported the original text into R and executed the cleaning algorithm described in the "Text Cleaning Algorithm" section. We computed semantic distances for the sequential bigrams from the original story. We then created a second random bigram vector

[3] R-package at https://github.com/Reilly-ConceptsCognitionLab/semdistflow.

consisting of each lemma of *To Build a Fire* in its original order randomly paired with an entry from the SemDist15 database. We derived experiential and embedding semantic distances for bigrams in both the sequential and random conditions.

Semantic Distance Across Sentence Boundaries in To Build a Fire

Here we tested the hypothesis that sentences constitute micro-topics with higher semantic relatedness (low semantic distance) within a sentence than between adjacent sentences. We tested this prediction by contrasting semantic distances (embedding and experiential) for sequential bigrams within sentences (within condition) relative to bigrams that broke across a sentence boundary (switch condition). We first cleaned the text of *To Build a Fire* using the procedures described in the "Text Cleaning Algorithm" section. We then coded each bigram as either within (nonadjacent to a period) or as a switch trial. Switch bigrams consisted of the final word of one sentence paired with the initial word of the following sentence.

Results

Norms for Bigram Semantic Distance

Tables 2 and 3 summarize descriptive statistics for experiential and embedding semantic distances for sequential bigrams in naturally ordered text (Sequential Model) versus artificially generated bigrams constructed via random word pairings (Random Model). We assigned categorical ranges ("low," "medium," "high") by referencing the interquartile range for each condition. The first and fourth quartiles constitute "low" and "high" distance, whereas the middle quartiles (Q1 to Q3) reflect medium or expected distances.

Figure 1 reflects density plots for both semantic spaces ($N = 250{,}000$ each). The distribution of bigram semantic distances generated by the GloWCA embedding model was tightly clustered and leptokurtic relative to the experiential semantic distance distribution whose variance was over four times higher.

Category Fluency Validation

We simulated a continuous stream of category fluency data consisting of 7,500 words with blocks of alternating 10-word clusters in a fixed order (i.e., animals, musical instruments, fruits/vegetables). This "actual" time series consisted of 6,750 words within clusters and 749 switches (i.e., there is no semantic distance or switch for the final word in the series). We coded all bigrams as either within cluster ($z < 1$) or as constituting a switch point between clusters ($z \geq 1$) using their scaled cosine distance. Table 4 and Figure 2 represent comparisons of the predicted distribution of binary events (switches and clusters) to the actual distribution of events. Overall classification accuracy was similar between the experiential (semdist15, 91.5%) and embedding spaces (GloWCA, 90.9%) with a medium tetrachoric correlation (0.36) between the two predicted time series.

Bigram Distance in *To Build a Fire*: Proof of Concept

Table 5 reflects semantic distances for sequential and random bigrams in *To Build a Fire*. There was a very large effect of text structure (i.e., ordered wording) on embedding-based semantic distances. That is, sequential bigrams in the original text had significantly lower semantic distance relative to random bigram pairings (Welch–Satterthwaite $t(6{,}333.2) = 57.61$, $p < .001$, Cohen's $d = 1.41$ [very large effect]). This effect of text structure was weaker for experiential semantic distances (SemDist15) as evidenced by a small to medium effect for sequential relative to random bigrams (Welch–Satterthwaite $t(6{,}381.2) = 12.57$, $p < .001$, Cohen's $d = 0.32$ [small to medium effect]).

Figure 3 illustrates the distribution of semantic distances across all bigrams of the original story. Note, the random bigram model reflected the pairing of a target word in its original position within *To Build a Fire* with another word randomly selected from the SemDist15 database. The guidelines in Figure 2 reflect boundaries for low, medium, and high distances generated in the norming study (see Tables 2 and 3).

Semantic Distance for Bigrams Across Sentence Boundaries in *To Build a Fire*

Semantic distances differed for sequential bigrams within sentences ($N = 5{,}200$) relative to bigram split across sentence boundaries ($N = 768$) across both the embedding and experiential models. Experiential (SemDist15) distance had a mean cosine distance

Table 2
SemDist15 Experiential Bigram Distances: Norms and Ranges

Statistic	Description of scale	Distance band	Bigram condition: Sequential	Random
M (*SD*)	Raw cosine (−1:1)	n/a	0.13 (0.42)	0.08
	Rescaled reverse scored (0:2)		0.87 (0.42)	0.92 (0.42)
Min to Q1	Raw cosine (−1:1)	Low	0.45 to 1.00	0.39 to 1.00
	Rescaled reverse scored (0:2)		0 to 0.55	0 to 0.61
Q1 to Q2	Raw cosine (−1:1)	Average	0.14 to 0.44	0.08 to 0.38
	Rescaled reverse scored (0:2)		0.56 to 0.86	0.62 to 0.92
Q2 to Q3	Raw cosine (−1:1)		−0.18 to 0.13	−0.76 to 0.07
	Rescaled reverse scored (0:2)		0.87 to 1.18	0.93 to 1.24
Q3 to Max	Raw cosine (−1:1)	High	−0.96 to −0.19	−0.97 to 0.75
	Rescaled reverse scored (0:2)		1.19 to 1.96	1.25 to 1.97

Note. Raw cosine scores reflect the original cosine value on a −1 to 1 scale with a cosine of 1 indicating 0 distance between two vectors. Rescaled reverse scored (0:2) values reflect a transformation of the original cosine values first to a range between 0 and 2 and then reverse scored using the 1-observed. On this transformed scale, 0 indicates no distance (i.e., a word vs. itself), and 2 reflects the highest possible dissimilarity between two words.

Table 3
GloWCA Embedding Bigram Distances: Norms and Ranges

Statistic	Description of scale	Distance band	Bigram condition	
			Sequential	Random
M (*SD*)	Raw cosine (−1:1)	n/a	0.42	0.28
	Rescaled reverse scored (0:2)		0.58 (0.20)	0.72 (0.22)
Min to Q1	Raw cosine (−1:1)	Low	0.56 to 1.00	0.41 to 1.00
	Rescaled reverse scored (0:2)		0 to 0.44	0 to 0.59
Q1 to Q2	Raw cosine (−1:1)	Average	0.42 to 0.55	0.27 to 0.40
	Rescaled reverse scored (0:2)		0.45 to 0.58	0.60 to 0.73
Q2 to Q3	Raw cosine (−1:1)		0.27 to 0.41	0.13 to 0.26
	Rescaled reverse scored (0:2)		0.59 to 0.73	0.74 to 0.87
Q3 to Max	Raw cosine (−1:1)	High	−0.22 to 0.26	−0.59 to 0.12
	Rescaled reverse scored (0:2)		0.74 to 1.22	0.88 to 1.59

Note. Raw cosine scores reflect the original cosine value on a −1 to 1 scale with a cosine of 1 indicating 0 distance between two vectors. Rescaled reverse scored (0:2) values reflect a transformation of the original cosine values first to a range between 0 and 2 and then reverse scored using the 1-observed. On this transformed scale, 0 indicates no distance (i.e., a word vs. itself), and 2 reflects the highest possible dissimilarity between two words.

normalized from 0 to 2 (0 is identical) for within-sentence bigrams (mean cosine distance = 0.86) relative to the between-sentence (switch) condition (mean cosine distance = 0.80), $t(487.89) = 3.09$, $p = .002$, Cohen $d = 0.18$, small effect.[4] Embedding (GloWCA) distances were also slightly higher for between-sentence bigrams (mean cosine distance = 0.85) relative to within-sentence bigrams (mean distance = 0.82), $t(451.48) = 2.7$, $p = .006$, Cohen $d = 0.18$, small effect.

We conducted a replication analysis to determine whether the surprising finding of lower between-sentence than within-sentence semantic distance observed in *To Build a Fire* would hold across a much larger and more varied corpus.[5] Using the same procedures applied to *To Build a Fire*, we analyzed bigram distances across 10 novels (see Table 6) sourced primarily from Project Gutenberg and the Harry Potter R-package (Boehmke, 2022). These texts included a total of approximately 906,421 words and 56,675 sentences. For nine of the 10 sources, embedding-based semantic distance was *higher* for bigrams crossing sentence boundaries relative to within-sentence bigrams. For the remaining text (i.e., *The Portrait of Dorian Gray* by Oscar Wilde), distance was lower between sentence bigrams, although this statistically significant difference constituted a very small effect. These overall results suggest that in most instances readers can expect to experience small jumps in semantic distance across sentence boundaries, as confirmed by a paired *t*-test on total between versus within bigram distances (embedding-based)[6] across all 10 texts, $t(9) = 3.97$, $p = .002$. We interpret these results in the general discussion to follow.

General Discussion

Much remains to be learned about the ways that meaning is conveyed in continuous language, and recent advances in natural language processing have afforded new insights into the ways that words combine at different scales. Here we evaluated a sequential bigram model involving the application of two high-dimensional semantic spaces to any continuous language sample. We developed open-source software for computing sequential bigram distance in a specified language sample of any length.[7] We also derived semantic distance norms that serve as a standard against which other analyses of connected language might be gauged. In the "Category Fluency Validation" section, we conducted a validation study demonstrating how both semantic spaces described in this work could successfully segment a continuous sample of simulated verbal fluency data. Finally, in the "Bigram Distance in to Build a Fire: Proof of Concept and Semantic Distance for Bigrams Across Sentence Boundaries in *To Build a Fire*" sections, we evaluated word-to-word semantic transitions in *To Build a Fire*. We discuss some of the major findings to follow.

Semantic distance is only interpretable in relation to the unique semantic space used to define it. For example, semantic distances between experiential and embedding models are not directly comparable. In addition, cosine values are not typically normalized, reverse-scored, or standardized across semantic spaces. As such, when one observes a cosine similarity value of 0.6, it is almost impossible to determine the magnitude of this semantic distance in the absence of a known standard. In the "Establishing Norms for Semantic Distance" section, we described steps for establishing bigram distance norms for experiential (SemDist15) and embedding (GloWCA) spaces. We queried hundreds of thousands of naturally occurring bigrams in contemporary English text relative to "synthetic" bigrams created by random word pairings. These analyses established bounds for low, medium, and high semantic distance which we then deployed as reference points for the analyses of *To Build a Fire*.

We used the norms established in the "Establishing Norms for Semantic Distance" section to evaluate a proof-of-concept that bigram distance in *To Build a Fire* would be higher for random bigram pairings relative to naturally ordered text, a pattern that

[4] Distances scaled from the original range of −1:1 to 0:2 and reverse scored such that 0 constitutes identical vectors and 2 is the greatest possible distance between any pair of words.

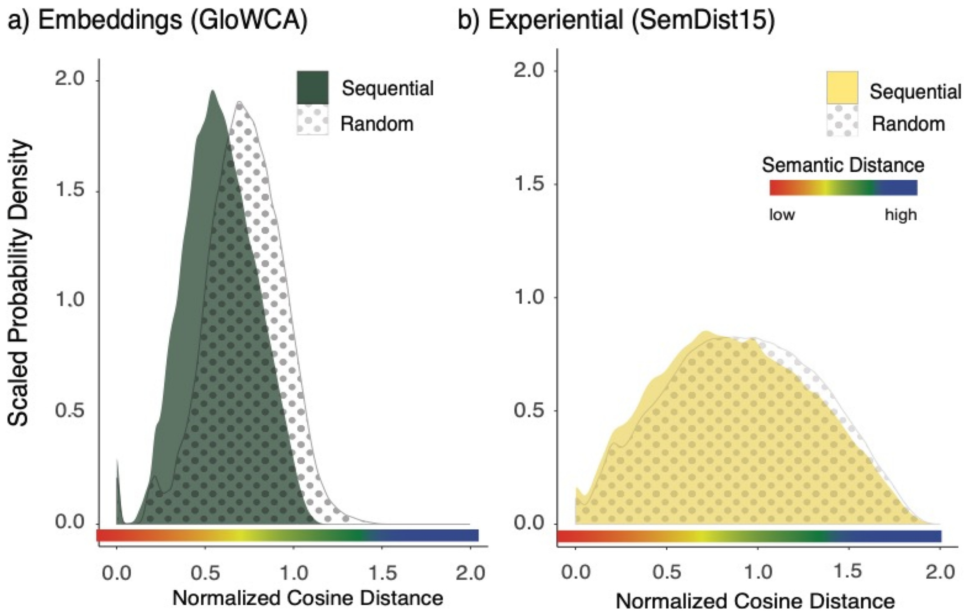
[5] This suggestion was raised by an anonymous reviewer, highlighting the importance of replication and extension with large-scale language models.

[6] There were no reliable or systematic differences in bigram distance between and within sentences via the experiential semantic space (SemDist15).

[7] All associated code is freely available for inspection and use within the "semdistflow" R-package. We encourage researchers to contact us for assistance.

Figure 1

Semantic Distance Density Distributions for Sequential and Random Bigrams



Note. The y-axis represents the scaled probability density for a given x-axis value x_n. Probability density is calculated by subtracting the sample mean from x_n and dividing by the standard deviation. The resulting value is then plugged into the normal probability density function to obtain the probability density for x_n. See the online article for the color version of this figure.

was also evident in the bigram norming sample. As predicted, adjacent words within *To Build a Fire* are more semantically related than randomly sampled bigrams within the same corpus. Relatedness between words in running text was far stronger for embedding (GloWCA) than experiential (SemDist15) distances, suggesting dominance for thematic (e.g., dog-collar) relative to taxonomic (e.g., dog-wolf) semantic relatedness in discourse.

The nature of how thematic and taxonomic semantic systems interact during language comprehension remains one of the most active topics in cognitive science. It has been argued that embedding (i.e., thematic) models are also capable of recovering taxonomic relationships (Grand et al., 2022). However, the reverse pattern appears less likely (i.e., experiential models recovering thematic relationships) since experiential models are more sensitive to perceptual than contextual similarity. Language discourse tends to unfold in terms of thematically related content (i.e., topics) making it more likely to encounter the word "leash" than "wolf" in proximity of "dog." This is not to say that taxonomic relationships are not present or important, particularly for scientific genres such as biology or zoology. However, our data suggest a more integral role for thematic relationships in online language processing.

Table 4

Accuracy of Binary Classification for Simulated Category Fluency

Statistic/Measure	SemDist15	GloWCA
% Accuracy	91.5	90.9
d'	2.02	1.86
Hit rate	93.4%	93.80%
False alarm rate	6.58%	6.20%
Odds ratio	40.66	28.66
Bias	0.96	0.98

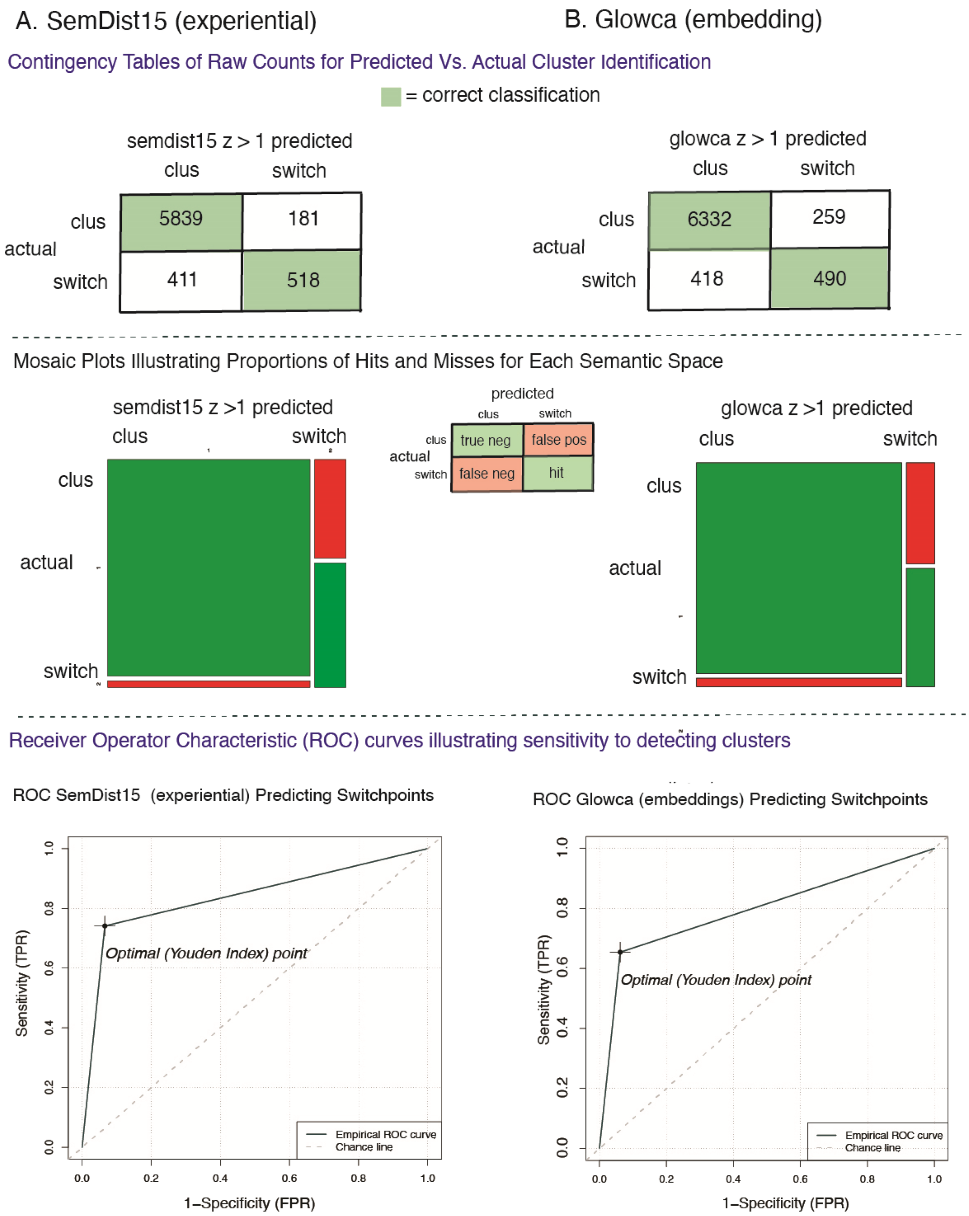
Note. Sensitivity metrics derived from the "verification" and "psych" packages in R.

The distinction of "normal" semantic distance in relation to narrative quality remains unclear. A narrative dominated by low semantic distance could be perceived as repetitive or hyper-focused. In contrast, excessively high semantic distance (i.e., each word highly unrelated to the last word) could be perceived as analogous to "word salad" in terms of cohesion. In our first patient-based extension of this method, Litovsky et al. (2022) sampled hundreds of thousands of bigrams from the narratives of people with aphasia relative to age-matched controls. This bag-of-bigrams approach demonstrated that people with aphasia show reduced bigram semantic distances relative to controls and that compression in semantic distance strongly correlates with semantic ability. In the current project, we extended this approach to model semantic distance across ordered discourse, potentially expanding the power and ecological validity of the measure to treat conceptual drift across words as a time series.

We demonstrated several applications of this continuous bigram approach. First, we conducted a validation study examining whether two metrics of semantic distance could effectively mark cluster boundaries in a simulated stream of verbal fluency data. Both semantic spaces (GloWCA and SemDist15) showed higher than 90% classification accuracy, demonstrating sensitivity to detect semantic fluctuations in continuous language output (see also Zemla et al., 2020). Second, we tested whether bigrams spanning sentence boundaries had higher semantic distances than bigrams within sentences. We predicted that between-sentence bigrams would be marked by a jump in semantic distance relative to within-sentence bigrams. This is premised on the idea that sentences constitute micro-

Figure 2

Sensitivity and Specificity of Binary Classification of Simulated Category Fluency



Note. See the online article for the color version of this figure.

topics organized around thematic semantic content and that transitions between sentences incur associated shifts in meaning.

We initially found that semantic distance was paradoxically lower for bigrams crossing sentence boundaries in *To Build a Fire*. However, a subsequent replication/extension analysis revealed that this finding was likely idiosyncratic. A more extensive corpus analysis across several works of fiction revealed that the final content word of one sentence and the initial content word of the next sentence tend to be punctuated by a jump in semantic distance, consistent with our original hypothesis. The extent to which readers and listeners are sensitive to such jumps to facilitate online sentence parsing remains unclear.

Applications and Future Directions

We envision a variety of applications to conceptualizing language as a continuous time series fluctuating in meaning over time. This includes the following:

(a) Causal modeling of language and physiological relations: Converting word-to-word level changes in meaning to a numeric time series will potentially facilitate causal modeling of how psycholinguistic (e.g., frequency, word length), psychophysiological (e.g., heart rate, pupil surface area), and neurological (e.g., evoked potentials) variables interact

Table 5
Semantic Distances for To Build a Fire

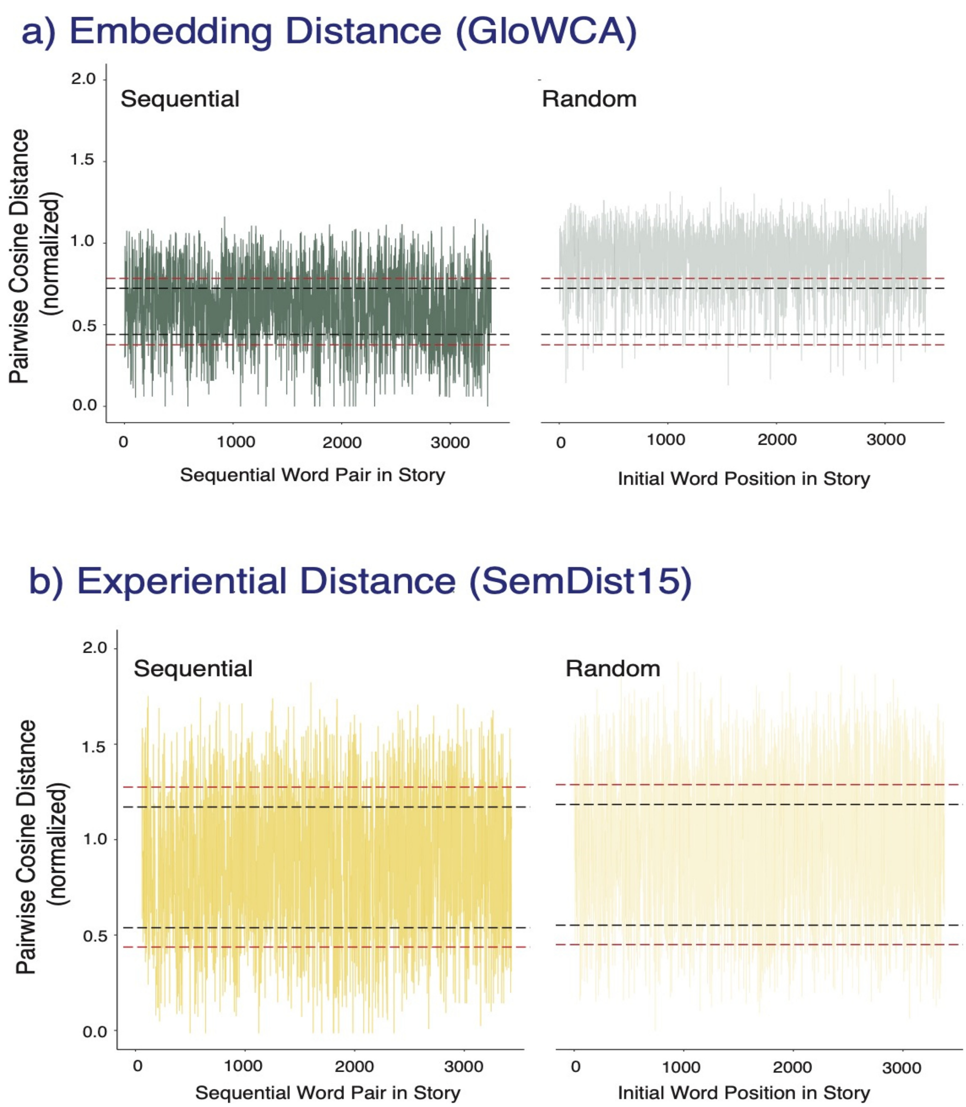
Semantic space	Condition	*N*	Mean distance	*SD*
Semdist15	Sequential	3,134	0.90	0.36
	Randomized	3,255	1.01	0.36
GloWCA	Sequential	3,360	0.62	0.23
	Randomized	3,370	0.91	0.18

Note. N = Number of bigrams, mean distance reflect cosine values transformed to a 0–2 scale where 0 is the highest possible similarity between two words.

during continuous language perception and production. For example, changes in embedding semantic distance in a running narrative could tax cognitive control which in turn perturbs pupil diameter and heart rate. One fruitful extension of this method will involve continuous measurement of neural signals using techniques such as MEG or EEG with an adequate temporal resolution to detect processes underlying predictive coding and the interplay between taxonomic and semantic systems during real-time language comprehension (for recent related work see Brodbeck et al., 2018; Kuperberg et al., 2006; Laszlo & Federmeier, 2009).

(b) Implicit detection of neurological disorders: Patterns of impairment in natural language production (e.g., writing samples, spoken language) have proven sensitive to detecting a wide range of neurological and psychiatric disorders (Fraser et al., 2016; Garrard et al., 2001, 2014). Automated screening using implicit language sampling is emerging as

Figure 3
Semantic Distances in To Build a Fire



Note. The red dashed reference lines reflect z-scores corresponding to ± 1.0 based on the norming procedures for adjacencies (i.e., sequential bigrams) described in the "Norms for Bigram Semantic Distance" section. The black dashed reference lines reflect the boundaries of Q1 and Q3 reflecting the interquartile ranges derived from the norming study in the "Norms for Bigram Semantic Distance" section. See the online article for the color version of this figure.

Table 6
Embedding Distance for Bigrams Within Sentences Versus Crossing Sentence Boundaries

Source	Token counts		Bigram distance		t-Statistic
	Words	Sentences	Within	Between	
Prisoner of Azkaban	104,860	8,936	0.62	0.65	$t(7{,}514.9) = 7.92$, $p < .001$***
Little Women	194,059	9,266	0.61	0.63	$t(8{,}498) = 3.55$, $p < .001$***
Sherlock Holmes	107,372	7,065	0.60	0.61	$t(7{,}335) = 3.01$, $p = .002$**
Portrait of Dorian Gray	82,012	6,687	0.60	0.59	$t(7{,}150) = 3.77$, $p \leq .001$***
Pride and Prejudice	124,719	6,210	0.61	0.62	$t(6{,}569.6) = 2.33$, $p = .02$*
Room with a View	69,931	5,948	0.62	0.63	$t(6{,}192.3) = 2.79$, $p = .01$*
Sorcerer's Stone	77,536	6,474	0.60	0.63	$t(5{,}874.7) = 5.46$, $p < .001$***
Become an Engineer	21,072	1,466	0.64	0.68	$t(1{,}694.6) = 6.92$, $p < .001$***
Honey Bees	91,577	3,182	0.72	0.75	$t(3{,}061.3) = 8.97$ $p < .001$***
Prehistoric Villages	33,283	1,541	0.67	0.72	$t(1{,}521) = 7.05$, $p < .001$***

Note. Token counts derived using the Quanteda package of R (Benoit et al., 2018). Distances reflect 0–2 cosine rescaled and reverse scored (0 is identical). Texts queried: Harry Potter and the Prisoner of Azkaban (J.K. Rowling, 1999); Little Women (Louisa May Alcott, 1868); The Adventures of Sherlock Holmes (Arthur Conan Doyle, 1892); The Portrait of Dorian Gray (Oscar Wilde, 1890); Pride and Prejudice (Jane Austen, 1813); A Room with a View (E.M. Forster, 1908); Harry Potter and the Sorcerer's Stone (J.K. Rowling, 1998); How to Become an Engineer (Frank W. Doughty, 2014); The Honey Bee: Its Natural History, Physiology, and Management (Edward Bevans, 1873); Prehistoric Villages, Castles, and Towers of Southwestern Colorado (Jesse Fewkes, 1919).
*$p < .05$. **$p < .01$. ***$p < .001$.

a powerful tool for the early detection of prodromal dementia (Merkin et al., 2022; Song et al., 2011; Spooner et al., 2020). We know of no algorithm that considers aberrant semantic distance in connected language as a marker of cognitive impairment.

(c) Norming of developmental milestones: Little is known about the trajectory of semantic composition throughout early language development. As children learn to narrate written and oral stories, inter-word semantic distance could prove sensitive to gauging the maturation of semantic knowledge. Specifically, longitudinal changes in semantic distance during storytelling could yield a sensitive marker of combinatorial semantic abilities.

(d) Auto-segmenting verbal fluency data: Our validation study demonstrates the utility of semantic distance in detecting switches between semantic clusters without the necessity for supervised machine learning or human intervention (e.g., manual scoring). This algorithm may prove useful for neuropsychology and other disciplines such as speech-language pathology that rely on verbal fluency as part of their core clinical assessment protocols.

(e) Evaluating how semantic distance is moderated by part-of-speech: One of the key steps needed to refine the proposed model is to improve sensitivity to disambiguate grammatical class (e.g., run as a verb vs. run as a noun). Little is currently known about how bigram semantic distance is moderated by grammatical role and how semantic distance might contribute to thematic role assignment and verb argument structure.

Limitations

Language is a rich symbolic modality comprised of numerous interactive subdomains. In its current form, however, our algorithm can only yield coarse data about how the meaning of one word relates to its neighbor. Our processing pipeline is currently insensitive to grammatical, pragmatic, and/or lexical ambiguity. The program yokes each word to its single entry in one of two lookup databases. The algorithm is agnostic to part of speech, polysemy, or homophony. This shortcoming undoubtedly results in error variance that could potentially be ameliorated by part-of-speech tagging or syntactic parsing. One rate limiting factor for the widespread adoption of such techniques involves the extensive processing resources required to parse large language samples. Such analyses often exhaust the capacity of personal computers, requiring database integration over high-performance clusters. As such, refinement of the semantic algorithms proposed here will require integration of syntactic and pragmatic information to provide a more realistic picture of combinatorial semantic processing.

Conclusions

We have proposed a continuous bigram model and open-source toolkit for analyzing semantic transitions in natural language and have identified numerous applications of the model to address theoretical and clinical questions about combinatorial semantic processing. Much remains to be learned about how to best measure conceptual shifts in language and how such variability either facilitates or compromises human communication. We invite researchers to explore these tools with their own datasets.

References

Alacam, Ö., Schüz, S., Wegrzyn, M., Kißler, J., & Zarrieß, S. (2022). Exploring semantic spaces for detecting clustering and switching in verbal fluency. In *Proceedings of the 29th international conference on computational linguistics* (pp. 178–191). https://aclanthology.org/2022.coling-1.16

Anderson, A. J., Binder, J. R., Fernandino, L., Humphries, C. J., Conant, L. L., Raizada, R. D. S., Lin, F., & Lalor, E. C. (2019). An integrated neural decoder of linguistic and experiential meaning. *The Journal of Neuroscience*, *39*(45), 8969–8987. https://doi.org/10.1523/JNEUROSCI.2575-18.2019

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, *95*(3), 709–721. https://doi.org/10.1016/j.neuron.2017.06.041

Beaty, R. E., & Johnson, D. R. (2021). Automating creativity assessment with SemDis: An open platform for computing semantic distance. *Behavior Research Methods*, *53*(2), 757–780. https://doi.org/10.3758/s13428-020-01453-w

Beaty, R. E., Zeitlen, D. C., Baker, B. S., & Kenett, Y. N. (2021). Forward flow and creative thought: Assessing associative cognition and its role in divergent thinking. *Thinking Skills and Creativity*, *41*, Article 100859. https://doi.org/10.1016/j.tsc.2021.100859

Benoit, K., Watanabe, K., Wang, H., Nulty, P., Obeng, A., Müller, S., & Matsuo, A. (2018). Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software*, *3*(30), 774. https://doi.org/10.21105/joss.00774

Binder, J. R., Conant, L. L., Humphries, C. J., Fernandino, L., Simons, S. B., Aguilar, M., & Desai, R. H. (2016). Toward a brain-based componential semantic representation. *Cognitive Neuropsychology*, *33*(3–4), 130–174. https://doi.org/10.1080/02643294.2016.1147426

Binney, R. J., Ashaie, S. A., Zuckerman, B. M., Hung, J., & Reilly, J. (2018). Frontotemporal stimulation modulates semantically guided visual search during confrontation naming: A combined tDCS and eye tracking investigation. *Brain and Language*, *180–182*, 14–23. https://doi.org/10.1016/j.bandl.2018.04.004

Binney, R. J., Zuckerman, B. M., Waller, H. N., Hung, J., Ashaie, S. A., & Reilly, J. (2018). Cathodal tDCS of the bilateral anterior temporal lobes facilitates semantically-driven verbal fluency. *Neuropsychologia*, *111*, 62–71. https://doi.org/10.1016/j.neuropsychologia.2018.01.009

Boehmke, B. (2022). *Harrypotter*. https://github.com/bradleyboehmke/harrypotter

Bonner, M. F., & Epstein, R. A. (2021). Object representations in the human brain reflect the co-occurrence statistics of vision and language. *Nature Communications*, *12*(1), Article 4081. https://doi.org/10.1038/s41467-021-24368-2

Brodbeck, C., Hong, L. E., & Simon, J. Z. (2018). Rapid transformation from auditory to linguistic representations of continuous speech. *Current Biology*, *28*(24), 3976–3983. https://doi.org/10.1016/j.cub.2018.10.042

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems* 33 (pp. 1877–1901). https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Capitani, E., Laiacona, M., Mahon, B. Z., & Caramazza, A. (2003). What are the facts of semantic category-specific deficits? A critical review of the clinical evidence. *Cognitive Neuropsychology*, *20*(3–6), 213–261. https://doi.org/10.1080/02643290244000266

Christensen, A. P., & Kenett, Y. N. (2021). Semantic network analysis (SemNA): A tutorial on preprocessing, constructing, and analyzing semantic networks. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000463

Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, *132*(2), 163–201. https://doi.org/10.1037/0096-3445.132.2.163

Crutch, S. J., Troche, J., Reilly, J., & Ridgway, G. R. (2013). Abstract conceptual feature ratings: The role of emotion, magnitude, and other cognitive domains in the organization of abstract conceptual knowledge. *Frontiers in Human Neuroscience*, *7*, Article 186. https://doi.org/10.3389/fnhum.2013.00186

Cutler, A. (1981). Making up materials is a confounded nuisance: Or will we be able to run any psycholinguistic experiments at all in 1990? *Cognition*, *10*(1–3), 65–70. https://doi.org/10.1016/0010-0277(81)90026-3

Davies, M. (2009). The 385+ million-word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, *14*(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

de Heer, W. A., Huth, A. G., Griffiths, T. L., Gallant, J. L., & Theunissen, F. E. (2017). The hierarchical cortical organization of human speech processing. *The Journal of Neuroscience*, *37*(27), 6539–6557. https://doi.org/10.1523/JNEUROSCI.3267-16.2017

Deniz, F., Nunez-Elizalde, A. O., Huth, A. G., & Gallant, J. L. (2019). The representation of semantic information across human cerebral cortex during listening versus reading is invariant to stimulus modality. *The Journal of Neuroscience*, *39*(39), 7722–7736. https://doi.org/10.1523/JNEUROSCI.0675-19.2019

Devlin, J. T., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *Bert: Pre-training of deep bidirectional transformers for language understanding*. ArXiv Preprint arXiv: 1810.04805.

Farah, M. J., & McClelland, J. L. (1991). A computational model of semantic memory impairment: Modality specificity and emergent category specificity. *Journal of Experimental Psychology: General*, *120*(4), 339–357. https://doi.org/10.1037/0096-3445.120.4.339

Fernandino, L., Binder, J. R., Desai, R. H., Pendl, S. L., Humphries, C. J., Gross, W. L., Conant, L. L., & Seidenberg, M. S. (2016). Concept representation reflects multimodal abstraction: A framework for embodied semantics. *Cerebral Cortex*, *26*(5), 2018–2034. https://doi.org/10.1093/cercor/bhv020

Fernandino, L., Tong, J.-Q., Conant, L. L., Humphries, C. J., & Binder, J. R. (2022). Decoding the information structure underlying the neural representation of concepts. *Proceedings of the National Academy of Sciences*, *119*(6), Article e2108091119. https://doi.org/10.1073/pnas.2108091119

Fraser, K. C., Meltzer, J. A., & Rudzicz, F. (2016). Linguistic features identify Alzheimer's disease in narrative speech. *Journal of Alzheimer's Disease*, *49*(2), 407–422. https://doi.org/10.3233/JAD-150520

Fu, Z., Wang, X., Wang, X., Yang, H., Wang, J., Wei, T., Liao, X., Liu, Z., Chen, H., & Bi, Y. (2023). Different computational relations in language are captured by distinct brain systems. *Cerebral Cortex*, *33*(4), 997–1013. https://doi.org/10.1093/cercor/bhac117

Funnell, E., Hughes, D., & Woodcock, J. (2006). Age of acquisition for naming and knowing: A new hypothesis. *Quarterly Journal of Experimental Psychology*, *59*(2), 268–295. https://doi.org/10.1080/02724980443000674

Garrard, P., Lambon Ralph, M. A., Hodges, J. R., & Patterson, K. (2001). Prototypicality, distinctiveness, and intercorrelation: Analyses of the semantic attributes of living and nonliving concepts. *Cognitive Neuropsychology*, *18*(2), 125–174. https://doi.org/10.1080/02643290125857

Garrard, P., Rentoumi, V., Gesierich, B., Miller, B., & Gorno-Tempini, M. L. (2014). Machine learning approaches to diagnosis and laterality effects in semantic dementia discourse. *Cortex*, *55*, 122–129. https://doi.org/10.1016/j.cortex.2013.05.008

Grand, G., Blank, I. A., Pereira, F., & Fedorenko, E. (2022). Semantic projection recovers rich human knowledge of multiple object features from word embeddings. *Nature Human Behaviour*, *6*(7), 975–987. https://doi.org/10.1038/s41562-022-01316-8

Gray, K., Anderson, S., Chen, E., Kelly, J. M., Christian, M., Patrick, J., Huang, L., Kenett, Y. N., & Lewis, K. (2019). Forward flow: A new measure to quantify free thought and predict creativity. *American Psychologist*, *74*(5), 539–554. https://doi.org/10.1037/amp0000391

Grossman, M., McMillan, C., Moore, P., Ding, L., Glosser, G., Work, M., & Gee, J. (2004). What's in a name: Voxel-based morphometric analyses of MRI and naming difficulty in Alzheimer's disease, frontotemporal dementia and corticobasal degeneration. *Brain*, *127*(3), 628–649. https://doi.org/10.1093/brain/awh075

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Hartung, F., Kenett, Y. N., Cardillo, E. R., Humphries, S., Klooster, N., & Chatterjee, A. (2020). Context matters: Novel metaphors in supportive and non-supportive contexts. *NeuroImage*, *212*, Article 116645. https://doi.org/10.1016/j.neuroimage.2020.116645

Hillis, A. E., & Caramazza, A. (1991). Category-specific naming and comprehension impairment: A double dissociation. *Brain*, *114*(5), 2081–2094. https://doi.org/10.1093/brain/114.5.2081

Hills, T. T., & Kenett, Y. N. (2022). Is the mind a network? Maps, vehicles, and skyhooks in cognitive network Science *Topics in Cognitive Science*, *14*(1), 189–208. https://doi.org/10.1111/tops.12570

Hodges, J. R., Patterson, K., Graham, N., & Dawson, K. (1996). Naming and knowing in dementia of Alzheimer's type. *Brain and Language*, *54*(2), 302–325. https://doi.org/10.1006/brln.1996.0077

Huth, A. G., De Heer, W. A., Griffiths, T. L., Theunissen, F. E., & Gallant, J. L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, *532*(7600), 453–458. https://doi.org/10.1038/nature17637

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224. https://doi.org/10.1016/j.neuron.2012.10.014

Jain, S., & Huth, A. G. (2018). *Incorporating context into language encoding models for fMRI*. BioRxiv. https://doi.org/10.1101/327601

Johnson, D. R., Kaufman, J. C., Baker, B. S., Patterson, J. D., Barbot, B., Green, A. E., van Hell, J., Kennedy, E., Sullivan, G. F., Taylor, C. L., Ward, T., & Beaty, R. E. (2022). Divergent semantic integration (DSI): Extracting creativity from narratives with distributional semantic modeling. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01986-2

Kenett, Y. N. (2018). Investigating creativity from a semantic network perspective. In Z. Kapoula, E. Volle, J. Renoult, & M. Andreatta (Eds.), *Exploring transdisciplinarity in art and science* (pp. 49–75). Springer.

Kenett, Y. N. (2019). What can quantitative measures of semantic distance tell us about creativity? *Current Opinion in Behavioral Sciences*, *27*, 11–16. https://doi.org/10.1016/j.cobeha.2018.08.010

Kenett, Y. N., & Faust, M. (2019). Clinical cognitive networks: A graph theory approach. In M. S. Vitevitch (Ed.), *Network science in cognitive science* (pp. 136–165). Routledge.

Kenett, Y. N., Levi, E., Anaki, D., & Faust, M. (2017). The semantic distance task: Quantifying semantic distance with semantic network path length. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1470–1489. https://doi.org/10.1037/xlm0000391

Kousta, S.-T. T., Vigliocco, G., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, *140*(1), 14–34. https://doi.org/10.1037/a0021446

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin and Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80. https://doi.org/10.3758/s13423-020-01792-x

Kumar, A. A., Balota, D. A., & Steyvers, M. (2020). Distant connectivity and multiple-step priming in large-scale semantic networks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(12), 2261–2276. https://doi.org/10.1037/xlm0000793

Kumar, A. A., Steyvers, M., & Balota, D. A. (2022). A critical review of network-based and distributional approaches to semantic memory structure and processes. *Topics in Cognitive Science*, *14*(1), 54–77. https://doi.org/10.1111/tops.12548

Kuperberg, G. R., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic, and thematic relationships in sentences. *Language and Cognitive Processes*, *21*(5), 489–530. https://doi.org/10.1080/01690960500094279

Landauer, T. K., & Dumais, S. T. (1997). 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

Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326–338. https://doi.org/10.1016/j.jml.2009.06.004

Litovsky, C., Finley, A., Zuckerman, B., Sayers, M., Schoenhard, J., Kennett, Y., & Reilly, J. (2022). *Semantic flow and its relation to controlled lexical-semantic retrieval deficits in aphasia*.

London, J. (1908). *To build a fire*. The Century.

Lundin, N. B., Brown, J. W., Johns, B. T., Jones, M. N., Purcell, J. R., Hetrick, W. P., O'Donnell, B. F., & Todd, P. M. (2022). *Neural switch processes guide semantic and phonetic foraging in human memory*. PsyArXiv. https://doi.org/10.31234/osf.io/857he

Lupker, S. J. (1979). The semantic nature of response competition in the picture-word interference task. *Memory and Cognition*, *7*(6), 485–495.

Lynott, D., Connell, L., Brysbaert, M., Brand, J., & Carney, J. (2020). The Lancaster sensorimotor norms: Multidimensional measures of perceptual and action strength for 40,000 English words. *Behavior Research Methods*, *52*(3), 1271–1291. https://doi.org/10.3758/s13428-019-01316-z

Mandera, P., Keuleers, E., & Brysbaert, M. (2015). How useful are corpus-based methods for extrapolating psycholinguistic variables? *Quarterly Journal of Experimental Psychology*, *68*(8), 1623–1642. https://doi.org/10.1080/17470218.2014.988735

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Marelli, M., Gagné, C. L., & Spalding, T. L. (2017). Compounding as abstract operation in Semantic Space: A data-driven, large-scale model for relational effects in the processing of novel compounds. *Cognition*, *166*, 207–224. https://doi.org/10.1016/j.cognition.2017.05.026

McRae, K., Cree, G. S., Westmacott, R., & Sa, V. R. D. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology / Revue Canadienne de Psychologie Expérimentale*, *53*(4), 360–373. https://doi.org/10.1037/h0087323

Merkin, A., Krishnamurthi, R., & Medvedev, O. N. (2022). Machine learning, artificial intelligence and the prediction of dementia. *Current Opinion in Psychiatry*, *35*(2), 123–129. https://doi.org/10.1097/YCO.0000000000000768

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *NIPS'13: Proceedings of the 26th international conference on neural information processing systems* (pp. 3111–3119).

Naselaris, T., Kay, K. N., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *Neuroimage*, *56*(2), 400–410. https://doi.org/10.1016/j.neuroimage.2010.07.073

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, Article 117254. https://doi.org/10.1016/j.neuroimage.2020.117254

NCAR - Research Applications Laboratory. (2015). *Verification: Weather forecast verification utilities (1.42)*. https://cran.r-project.org/web/packages/verification/index.html.

Olson, J. A., Nahas, J., Chmoulevitch, D., Cropper, S. J., & Webb, M. E. (2021). Naming unrelated words predicts creativity. *Proceedings of the National Academy of Sciences*, *118*(25), Article e2022340118. https://doi.org/10.1073/pnas.2022340118

Ovando-Tellez, M., Benedek, M., Kenett, Y. N., Hills, T. T., Bouanane, S., Bernard, M., Belo, Y., Belanger, B., Bieth, T., & Volle, E. (2022). An investigation of the cognitive and neural correlates of semantic memory search related to creative ability. *Communications Biology*, *5*(1), Article 604. https://doi.org/10.1038/s42003-022-03547-x

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics.

Peters, M. E., Ammar, W., Bhagavatula, C., & Power, R. (2017). *Semi-supervised sequence tagging with bidirectional language models*. ArXiv Preprint arXiv: 1705.00108

Pexman, P. M., Heard, A., Lloyd, E., & Yap, M. J. (2017). The Calgary semantic decision project: Concrete/abstract decision data for 10,000 English words. *Behavior Research Methods*, *49*(2), 407–417. https://doi.org/10.3758/s13428-016-0720-6

Popham, S. F., Huth, A. G., Bilenko, N. Y., Deniz, F., Gao, J. S., Nunez-Elizalde, A. O., & Gallant, J. L. (2021). Visual and linguistic semantic representations are aligned at the border of human visual cortex. *Nature Neuroscience*, *24*(11), 1628–1636. https://doi.org/10.1038/s41593-021-00921-6

Price, A. R., Bonner, M. F., Peelle, J. E., & Grossman, M. (2015). Converging evidence for the neuroanatomic basis of combinatorial semantics in the angular gyrus. *The Journal of Neuroscience*, *35*(7), 3276–3284. https://doi.org/10.1523/JNEUROSCI.3446-14.2015

Raji, S., & da Melo, G. (2020). What sparks joy: The AffectVec emotion database. In *Proceedings of the web conference* (2991–2997). ACM.

Reilly, J., Finley, A. M., Kelly, A., Zuckerman, B., & Flurie, M. (2021). Olfactory language and semantic processing in anosmia: A neuropsychological case control study. *Neurocase*, *27*(1), 86–96. https://doi.org/10.1080/13554794.2020.1871491

Reilly, J., Kelly, A., Zuckerman, B. M., Twigg, P. P., Wells, M., Jobson, K. R., & Flurie, M. (2020). Building the perfect curse word: A psycholinguistic investigation of the form and meaning of taboo words. *Psychonomic Bulletin and Review*, *27*(1), 139–148. https://doi.org/10.3758/s13423-019-01685-8

Reilly, J., Peelle, J. E., Garcia, A., & Crutch, S. J. (2016). Linking somatic and symbolic representation in semantic memory: The dynamic multilevel reactivation framework. *Psychonomic Bulletin & Review*, *23*(4), 1002–1014. https://doi.org/10.3758/s13423-015-0824-5

Rinker, T. W. (2018a). *Textstem: Tools for stemming and lemmatizing (Version 0.1.4)* [Computer software]. https://cran.r-project.org/web/packages/textstem/index.html

Rinker, T. W. (2018b). *Textclean (Version 0.9.3)* [Computer software]. https://CRAN.R-project.org/package=textclean

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*(1), 205–235. https://doi.org/10.1037/0033-295X.111.1.205

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350. https://doi.org/10.1016/0010-0285(73)90017-0

Sacchett, C., & Humphreys, G. W. (1992). Calling a squirrel a squirrel but a canoe a wigwam: A category-specific deficit for artefactual objects and body parts. *Cognitive Neuropsychology*, *9*(1), 73–86. https://doi.org/10.1080/02643299208252053

Selivanov, D. (2020). *text2vec: Modern text mining framework for R (Version 0.6)* [Computer software]. https://cran.r-project.org/web/packages/text2vec/index.html

Simony, E., Honey, C. J., Chen, J., Lositsky, O., Yeshurun, Y., Wiesel, A., & Hasson, U. (2016). Dynamic reconfiguration of the default mode network during narrative comprehension. *Nature Communications*, *7*(1), Article 12141. https://doi.org/10.1038/ncomms12141

Song, X., Mitnitski, A., & Rockwood, K. (2011). Nontraditional risk factors combine to predict Alzheimer disease and dementia. *Neurology*, *77*(3), 227–234. https://doi.org/10.1212/WNL.0b013e318225c6bc

Spooner, A., Chen, E., Sowmya, A., Sachdev, P., Kochan, N. A., Trollor, J., & Brodaty, H. (2020). A comparison of machine learning methods for survival analysis of high-dimensional clinical data for dementia prediction. *Scientific Reports*, *10*, Article 20410. https://doi.org/10.1038/s41598-020-77220-w

Troche, J., Crutch, S. J., & Reilly, J. (2017). Defining a conceptual topography of word concreteness: Clustering properties of emotion, sensation, and magnitude among 750 English words. *Frontiers in Psychology*, *8*, Article 1787. https://doi.org/10.3389/fpsyg.2017.01787

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychoogy*, *22*(3), 370–378. https://doi.org/10.1076/1380-3395(200006)22:3;1-V;FT370

Troyer, A. K., Moscovitch, M., Winocur, G., Alexander, M. P., & Stuss, D. (1998). Clustering and switching on verbal fluency: The effects of focal frontal- and temporal-lobe lesions. *Neuropsychologia*, *36*(6), 499–504. https://doi.org/10.1016/S0028-3932(97)00152-8

Warrington, E. K. (1975). The selective impairment of semantic memory. *Quarterly Journal of Experimental Psychology*, *27*(4), 635–657. https://doi.org/10.1080/14640747508400525

Westerlund, M., & Pylkkänen, L. (2014). The role of the left anterior temporal lobe in semantic composition vs. semantic memory. *Neuropsychologia*, *57*, 59–70. https://doi.org/10.1016/j.neuropsychologia.2014.03.001

Wingfield, C., & Connell, L. (2022). Sensorimotor distance: A grounded measure of semantic similarity for 800 million concept pairs. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01965-7

Woollams, A. M., Cooper-Pye, E., Hodges, J. R., & Patterson, K. (2008). Anomia: A doubly typical signature of semantic dementia. *Neuropsychologia*, *46*(10), 2503–2514. https://doi.org/10.1016/j.neuropsychologia.2008.04.005

Zemla, J. C., Cao, K., Mueller, K. D., & Austerweil, J. L. (2020). SNAFU: The semantic network and fluency utility. *Behavior Research Methods*, *52*(4), 1681–1699. https://doi.org/10.3758/s13428-019-01343-w

Received June 30, 2022
Revision received January 9, 2023
Accepted January 20, 2023 ■