COGNITIVE SCIENCE
A Multidisciplinary Journal

Check for updates

Cognitive Science 48 (2024) e13413

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13413

Determining the Relativity of Word Meanings Through the Construction of Individualized Models of Semantic Memory

Brendan T. Johns

Department of Psychology, McGill University

Received 6 September 2022; received in revised form 11 November 2023; accepted 27 January 2024

Abstract

Distributional models of lexical semantics are capable of acquiring sophisticated representations of word meanings. The main theoretical insight provided by these models is that they demonstrate the systematic connection between the knowledge that people acquire and the experience that they have with the natural language environment. However, linguistic experience is inherently variable and differs radically across people due to demographic and cultural variables. Recently, distributional models have been used to examine how word meanings vary across languages and it was found that there is considerable variability in the meanings of words across languages for most semantic categories. The goal of this article is to examine how variable word meanings are across individual language users within a single language. This was accomplished by assembling 500 individual user corpora attained from the online forum Reddit. Each user corpus ranged between 3.8 and 32.3 million words each, and a count-based distributional framework was used to extract word meanings for each user. These representations were then used to estimate the semantic alignment of word meanings across individual language users. It was found that there are significant levels of relativity in word meanings across individuals, and these differences are partially explained by other psycholinguistic factors, such as concreteness, semantic diversity, and social aspects of language usage. These results point to word meanings being fundamentally relative and contextually fluid, with this relativeness being related to the individualized nature of linguistic experience.

Keywords: Lexical semantics; Distributional modeling; Cognitive modeling; Machine learning; Big data

Correspondence should be sent to Brendan T. Johns, Department of Psychology, McGill University, 2001 McGill College Avenue, Montreal, QC H3A 1G1, Canada. E-mail: brendan.johns@mcgill.ca

Linguistic experience, and the experiences that humans have in general, differs radically across people, with the main driver of this variability being the culture that one is raised in (Kramsch, 2014). Demographic factors influence the nature of experience within a culture, such as socioeconomic status, political alignment, age, and education, among others, as well as personal factors, such as one's choice of entertainment (e.g., the genre of fiction books that one enjoys reading or the television shows that one watches). If you were to select two individuals at random, they would likely differ considerably across these variables, which entails that they would have corresponding differences in the types of language that they would have experienced.

Understanding the impact of experience on human behavior is an integral component of cognitive theory. For instance, a classic goal in cognitive modeling is to integrate environmental structure into theories of cognition (e.g., Estes, 1955; Simon, 1969; see Johns, Jamieson, & Jones, 2023 for a recent discussion of these seminal works). However, classic theoretical attempts have struggled with how to integrate human experience into models of cognition at scale, mainly due to the technological limitations of the time. For example, to account for word meanings in cognitive models of episodic memory, most models have relied upon randomly generated representations that do not map onto the semantic structure of words in lexical memory (Johns & Jones, 2010), mainly due to no realistic alternative being possible. However, recently the cognitive mechanisms underlying the acquisition of word meanings through experiential learning have been explored with a new class of computational cognitive model, entitled distributional models of semantics (alternatively, they are also referred to as word embedding models). There have been multiple models of this type developed (e.g., Griffiths, Steyvers, & Tenenbaum, 2007; Jones & Mewhort, 2007; Landauer & Dumais, 1997; Mikolov, Sutskever, Chen, Corrado, & Dean, 2013), all with different theoretical orientations and goals. However, all distributional models are based around the central premise that a word's meaning can be derived through the learning of word co-occurrence statistics. In this article, when discussing word meanings this will refer to the representations that distributional models derive, with an acknowledgment that these models do not encapsulate all aspects of the word meanings that humans have.

Humans experience millions of episodic experiences with language every year (Brysbaert, Stevens, Mandera, & Keuleers, 2016). This level of experience enables people to acquire sophisticated representations of the meaning of words, a process replicated and explored by distributional models. The theoretical goals of distributional modeling, and experientially dependent theories of cognition more generally, is to explain the variability in human behavior with this variability in human experience (see, e.g., the weaker links hypothesis for an experiential account of bilingualism; Gollan, Montoya, Cera, & Sandoval, 2008).

As stated, distributional models are experientially dependent, meaning that they require training materials to learn from. These training materials are typically in the form of linguistic corpora and, in cases where they are used to explain behavioral data, are meant to be a proxy to the types of language that a typical human being may experience during their lifetime. These corpora are used by the models to extract word co-occurrence statistics in order to form vector representations of the meaning of words. A variety of different corpus types have been used across studies, including textbooks (Landauer & Dumais, 1997), newspaper articles

(Aujla, 2021; Davies, 2009), television and movie subtitles (Brysbaert & New, 2009), fiction books (Johns & Jamieson, 2018; Johns, Dye, & Jones, 2020), social media (Herdağdelen & Marelli, 2017; Johns, 2019), and online encyclopedias (Shaoul & Westbury 2010), among many others.

The experiential nature of distributional models is their main theoretical advantage, as they allow for a determination of how much semantic knowledge is derived from the experience that humans have with their environment (Landauer & Dumais, 1997) and the type of cognitive mechanisms that are required to produce behavior when a cognitive model is grounded with a realistic representation of word meanings (Johns, Jamieson, & Jones, 2020, 2023). One specific insight that these models have provided, and one that was unavailable to generations of previous researchers due to technological limitations, is that considerable amounts of knowledge can be derived from linguistic experience at scale. Given the considerable success that these models have had in explaining a variety of behavioral phenomenon (for recent reviews, see Bhatia & Aka, 2022; Günther, Rinaldi, & Marelli, 2019; Kumar, 2020), these models suggest that the main driver of the human ability to learn word meanings is based upon the experience that humans have with the natural language environment.

If culture influences human experience, and human experience in turn dictates the type of language that one processes during daily life, it suggests that there could be considerable variability in the word meanings that different languages have. This issue of meaning variability has recently been examined by Thompson, Roberts, and Lupyan (2020) (see also Wang & Bi, 2021 and Martí, Wu, Piantadosi, & Kidd, 2023 for related research) using distributional modeling. The goal of this study was to understand how word meanings vary across languages, using a measure they proposed entitled semantic alignment, derived from a distributional model trained with corpora of different languages. In particular, Thompson et al. examined the variability in the meaning of a set of 1010 words across 41 languages, using a common distributional model type entitled neural embedding models (Mikolov et al., 2013). They found that there was considerable variability in the meanings of words across languages for most semantic categories, and that the deviations were predicted by cultural aspects of language usage (e.g., languages spoken in closer geographical range were more aligned than languages that had a greater geographical distance). The result of Thompson et al. demonstrates the promise that distributional modeling offers in answering fundamental questions about the impact of culture on language processing.

The theoretical aims of Thompson et al. were to examine universalists versus relative perspectives on lexical semantics. The universalist position posits that human semantic knowledge is based mainly on a core set of innate concepts that are universal across cultures, with words from different languages mapping onto these core concepts (see Gleitman & Fisher, 2005; Li & Gleitman, 2002; Pinker, 1994). That is, the universalist approach assumes that much of the meaning of words is formed outside of actual experience and is based upon innate knowledge. In contrast, the relative approach to semantics proposes that the experience that one has with the world drives the meaning that people acquire about words (see Evans & Levinson, 2009; Lupyan, 2016; Lupyan & Dale, 2016). This perspective entails that since different cultures afford different experiences for the people embedded in them, the languages of those cultures have different meanings associated with the same words. Thus, there should be varied word meanings across cultures. Given that Thompson et al. found limited semantic alignment (with the exception of a few select categories such as numbers) across languages, this finding suggests that there is a fair amount of relativity in word meanings, and thus semantic concepts in general, across cultures.

However, as detailed previously, there are many factors that could impact the linguistic exposure of an individual within a single language, given the varied nature of human experience. The goal of this article is to examine variability in word meanings across individuals and determine its connection to other psycholinguistic variables, in order to gain insight into the various factors that impact word meaning relativity and an initial understanding of the topography of word meaning relativity across individual users of a language. Although the goals of Thompson et al. were considerably different than the current article, as the objective here is to determine the within language meaning variability rather than across language meaning variability, this previous study serves as a theoretical and methodological inspiration for the work contained here.

Theoretically, the current article aims to assess the universality of word meanings within a single language (namely, English). Specifically, the question this article asks is whether the individualized experience that humans receive during their lifespan results in a significant level of variability of word meanings across individual language users. Additionally, the article seeks to determine the impact that communicative discourse has on the flexibility and fluidity of word meanings across individuals. Determining the answer to this question will lead to a greater level of understanding of the universality of word meanings, since if it is found that there are low amounts of semantic alignment across individuals, it would indicate that language relativity holds even within a language, not just across languages.

Although it is difficult to gain a quantitative understanding of the variability in linguistic experience, the rise of corpus-based cognitive models has provided some tools to analyze this question with. As a first pass at understanding how individualized experience influences lexical processing, Johns and Jamieson (2018) examined individual variability in language usage at a large scale using distributional semantic modeling. This was done by assembling a set of approximately 2000 fiction books organized by author and genre. They then assessed the similarity of books depending on if they were written by the same author, in the same genre, or across authors and genres. It was found that books written in the same genre had a small increase in similarity compared to books written in different genres, suggesting semantic consistency within a genre. However, the biggest difference emerged at the author level, where the books written by an individual author were much more similar to each other than books written by different authors, regardless of genre. This result was replicated by Johns et al. (2020) using a much larger set of over 25,000 books. This finding suggests that individual language usage is quite distinct across users of language, with each author having a unique signature of language usage contained in their writings.

To examine these issues more specifically, Johns, Jones, and Mewhort (2019) devised an optimization algorithm to improve the performance of distributional models by manipulating the underlying experience that a model receives. The method, entitled Experiential Optimization (EO), was implemented by assembling a large collection of different types of language materials, and determining the optimal combination of lexical materials when fitting an experientially dependent model to different types of lexical behavior. The underlying motivation for the development of EO was to determine if the different experiences that people have with a language are reflected in their lexical behavior. To accomplish this, the different text sources were split into smaller sections. Then, a hill-climbing algorithm was used to iteratively select the sections that allowed for a model's representation to be best optimized to a given set of data. For example, one could take a set of free association data and find the sets of texts that allows for a distributional model to best account for these data. The end result of EO is a set of language experiences that maximize the performance of a model to a set of data.

The results of Johns et al. (2019) were multifold. As with standard parameter optimization techniques used in cognitive modeling (Shiffrin, 2010; Shiffrin, Lee, Kim, & Wagenmakers, 2008), it was found that optimizing distributional models (or process models that utilized a representation from a distributional model) with EO vastly increased multiple model's fits to various types of data across lexical semantics, lexical organization, sentence processing, and episodic recognition, suggesting that being able to control linguistic experience provides a powerful tool in accounting for human behavior. Additionally, it was found that EO could infer demographic information about the participants of the data that was analyzed, as it was found that when using EO to optimize to lexical decision to data collected from younger and older adults, the optimization method preferentially selected young adult fiction books when fitting to young adult lexical decision data but preferred more advanced fiction books when fitting to older adult's lexical decision data. This suggests that the types of language experience that an individual has experienced are embedded within their lexical behavior.

To further validate this method, Johns and Jamieson (2019) collected a set of approximately 26,000 books organized by the date and country of birth of the various authors and also assembled multiple lexical decision and word familiarity datasets that were collected across time and from both the United States and the UK. When using EO with these books as the individual language sections, it was found that the method had a very strong preference toward the time- and place-appropriate language sections for the different datasets (i.e., the method selected UK authors to explain data collected in the UK, and vice versa for data collected in the United States). Similar results were found by Taler, Johns, and Jones (2020) when using demographically organized corpora to examine a mega dataset of verbal fluency performance across the aging spectrum, with older adults being better explained by a distributional model trained on older books (and vice-versa for younger participants).

Overall, these results strongly suggest that the specific types of linguistic experience that people have deeply influence their language processing system, and that linguistic experience is latent within their patterns of lexical behaviors. From a cognitive perspective, these results suggest that the experience that different individuals have with language results in different representations being formed within memory, leading to inherent variability in the meanings that people have for words. Thus, rather than lexical semantic memory being consistent across individuals of a population, this perspective entails that there could be considerable differences in the meanings of words that different speakers of a language have.

The issue of within language relativity is a central component of modern theories of psycholinguistics, including usage-based (Tomasello, 2003), adaptive (Beckner et al., 2009; Christiansen & Chater, 2008), and social/ecological perspectives (Tiv et al., 2022). All of these theories propose that individual experience with language drives the language comprehension and production systems. This entails that there should be variability in the word meanings acquired across people due to the nature of individualized linguistic experience. Thus, understanding the dynamics of within language relativity will shed light on the impact that individual experience has on language processing in general and provide further pathways for theoretical development.

In order to assess within language meaning variability, it is obviously necessary to have large samples of multiple individual's usage of language. The need for large individual corpora comes from distributional models requiring adequate numbers of samples of a word's usage to construct stable representations of word meanings. There are a number of possible candidates for constructing these corpora—for example, the fiction books assembled and analyzed by Johns and Jamieson (2019) and Johns et al. (2020). However, even though some individual authors had many books in the collection, most did not reach a large enough size to drive a distributional model. Instead, here individual user corpora will be formed from high-level commenters on the internet forum Reddit. This language source has recently been used to examine communicative and socially based effects in lexical organization (Chang, Jones, & Johns, 2023; Johns, 2021a, 2022; Johns & Jones, 2022; Taler, Johns, & Jones, 2020; Senaldi, Titone, & Johns, 2022) and lexical semantics (Johns, 2021b, 2023) using distributional semantic modeling.

To generate individual user corpora, a website entitled pushshift.io was utilized. This website aggregates Reddit posts using the publicly available Reddit API[1] in order to enable researchers, mostly from the computational social sciences, to conduct research on this important data source (Baumgartner, Zannettou, Keegan, Squire, & Blackburn, 2020). All posts from the beginning of the website (June 2005) to June 2021 were assembled, and the 500 most prolific commenters, who had public usernames, on the website were found by counting the number of comments that they made during their time on the forum (see the Methods section for descriptions of quality control). This resulted in 500 user corpora that ranged from approximately 3.8 to 32.3 million words, with an average corpus size of approximately 7.3 million words. The top panel of Fig. 1 contains a histogram of the distribution of corpus size for the different user corpora. These are considerably smaller than the typical corpora used to train modern distributional models, which are typically in the hundreds of millions or billions or words. However, the average user corpus analyzed here is roughly equivalent to the classic TASA corpus that Landauer and Dumais (1997) used, which contained approximately 7 million words. However, even though the user corpora do not reach modern corpus sizes, they do contain considerable amounts of language for each language user, and the reduction in corpus size is a necessary sacrifice to produce a new avenue of research for distributional modeling.

To get an initial understanding into the variability of language usage contained in the user corpora, a preliminary simulation was conducted examining the similarity of word frequency distributions across the different users. Following the work of Johns and Jamieson (2018, 2019), each corpus was reduced to a vector of word frequency values of the 100,000 most frequent words. Each value in the vector was transformed with a natural logarithm, in order to remove the contribution of very high-frequency words. The similarity between all of the user's word frequency vectors was then taken, resulting in 124,750 across user comparisons.

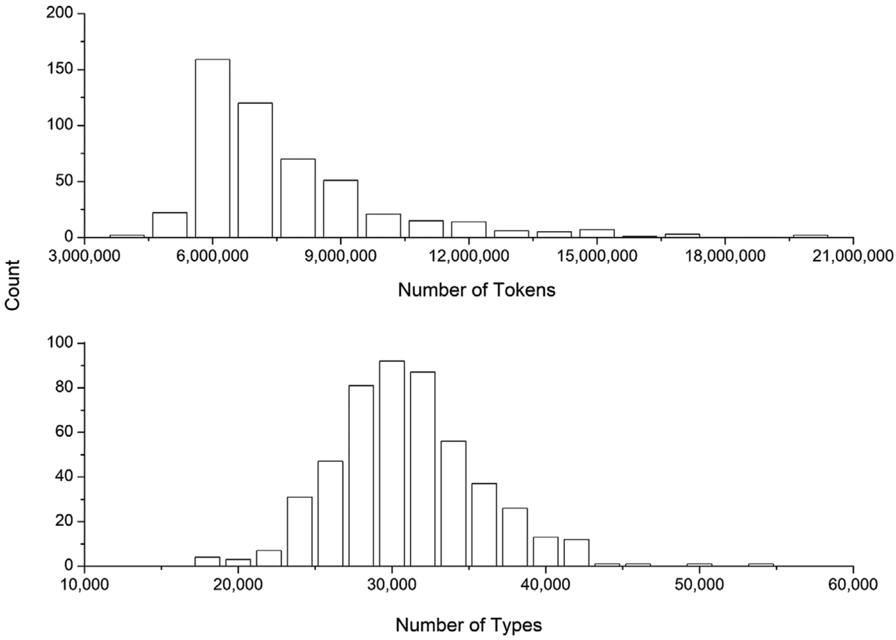

Fig. 1. Histograms of the number of tokens (top panel) and the number of types (bottom panel) that each user corpus contains.

Similarity was calculated with a vector cosine (see Eq. 6), as is standard in distributional modeling. These values were compared against similarity values obtained by splitting each user corpus in half and taking the similarity between the corresponding word frequency vectors from each half. This resulted in 500 within user comparisons. The histograms for the across user and within user comparisons are contained in Fig. 2. This figure shows a very strong trend—each user's overall use of language is very distinct from others, such that there is very little overlap between the distributions. Indeed, only 6% of the datapoints overlap in the two distributions. The result of this initial analysis demonstrates that individuals differ substantially in their overall language usage at a large-scale. This has theoretical consequences for distributional models as the representations that these models construct are generally not flexible (i.e., not context-dependent) in the information that they contain. The question then becomes how does this variability extend to individual word meanings and what are the psycholinguistic/environmental factors that influence this variability?

Besides providing an opportunity to examine the language usage of individuals at a large scale, the Reddit user corpora offer a number of additional advantages. One major advantage, as previously explored by Johns, 2021a, 2021b) is that much of the language contained the forum is communicative in nature (as they involve multiple users communicating with each other), rather than descriptive language such as what is contained in book or encyclopedia corpora. Furthermore, for each comment that a user produced on the site, a number of additional metadata can be extracted for that comment, such as the discourse (Subreddit; discussions

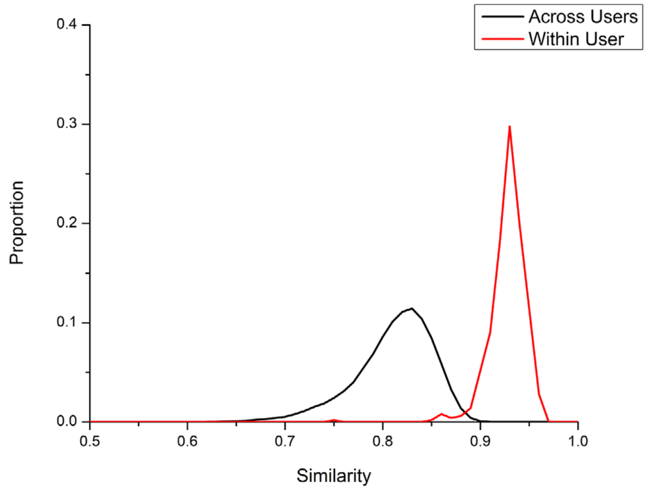

Fig. 2. A histogram of the similarity values of word frequency vectors for across user comparisons versus the similarity values obtained by splitting each user corpus in half (the within user comparisons). This figure shows that each individual has a very distinct usage of language.

around a certain topic) and date that the comment was made in. This metadata will be used in the following analyses to explore different aspects of semantic alignment, such as how the type of discourse that individuals communicate in impacts two user's semantic alignment and how aligned an individual's own usage of language is across time.

The theoretical aim of this article is to use the above-specified user corpora to understand the semantic alignment of words across individual users of language, in order to determine the relative nature of word meanings within a language. The results will be used to further examine issues of universality and relativeness in lexical semantic memory by examining the semantic alignment of word meanings that different individuals have. This will be accomplished by using a count-based distributional approach previously used by Johns, Mewhort, and Jones (2019) and Johns, 2021b, 2023), which incorporates training routines from neural embedding models, trained on the individual user corpora described above. This model will be used to determine how varied the meanings of words are across individuals and how this variability maps onto established measures in psycholinguistics, using two different measures of semantic alignment. Additionally, metadata about user's commenting patterns will be used to examine social and communicative pressures on the relativity of word meanings.

An additional goal of this article is to provide a high-level examination of how meaning changes within single language users across discourse types, in order to inform future cognitively inspired distributional model development. The majority of current distributional models are centralized models (see Jamieson, Johns, Vokey, & Jones, 2022 for a recent discussion on centralized vs. decentralized computational models of cognition) that derive a single representation of a word's meaning from a large corpus. These representations are typically context-independent, meaning that the representations the models form do not change based around other aspects of the overall language environment (such as the discourse one is in). Thus, simulations in this article will examine how variable language usage is within and across individuals when they communicate in the same discourse versus across discourses. If is found that there are high levels of semantic alignment in the word meanings that individuals use across discourses, this would signal that utilizing centralized, context-independent representations is theoretically sound. However, if there are limited levels of semantic alignment, this would signal that future distributional models have to include mechanisms that allow the formation of context-dependent representations. Thus, the results of this article will also provide important criteria for the development of more cognitively plausible large-scale models of language processing.

1. Methods

There will be five aspects of the modeling work described in this section: (1) the construction of individual user corpora from Reddit; (2) prior research on using distributional models to calculate semantic alignment; (3) the distributional modeling framework utilized here; (4) the two semantic alignment measures that are to be evaluated; and (5) implementation details about the computing of the alignment measures.

1.1. Construction of individual user corpora

The construction of the user corpora followed the previous work of Johns (2021a) in building user corpora from users of Reddit who have publicly available usernames on Reddit. For Johns (2021a), Reddit data stopped being collated in September 2019. Here, the data were collected through June 2021, which added a considerable amount of text considering the current popularity of the forum. However, given that in this article only the 500 most prolific users found will be analyzed, the update provided only a moderate increase in the amount of text used here. As stated, the Reddit data were downloaded from pushshift.io, which makes every comment made on Reddit publicly available in JSON database files for each month analyzed. Each comment in the JSON file contains the text of the comments, as well as associated metadata (e.g., the user who produced the comment, time it was produced, subreddit it was produced in, etc…). Due to the size of these files, a dedicated hardware setup was required to construct the user corpora.

User corpora were constructed for every user who had produced over 3000 comments, which resulted in 520,348 separate corpora. From these, the 1000 users who had produced the greatest number of words were selected. Then, the number of word types in each of the 1000 corpora were calculated. The word list used to calculate types consisted of the combined words contained in the word prevalence data of Brysbaert, Mandera, McCormick, and Keuleers (2019), the English lexicon project (ELP; Balota et al., 2007), and the British lexicon project (BLP; Keuleers, Lacey, Rastle, & Brysbaert, 2012). This resulted in a total of 80,815 words. Any corpus that had fewer than 10,000-word types was rejected, in order to ensure that the user had a varied vocabulary. Additionally, each corpus was inspected by hand in order to make sure that no corpora were formed by bots or automated accounts of some kind. Of the remaining corpora, the 500 corpora with the greatest number of words were selected.

One drawback to the construction and usage of the individualized corpora described here (as opposed to say the usage of fiction books such as what was used by Johns & Jamieson, 2018, 2019) is that it is difficult to ensure that the corpora contain the writings of a single individual. This is a general drawback to the usage of social media resources, as they are composed of comments by users who are typically anonymous. However, the analysis of single commenters on a social media website allows for samples of individual language that is not attainable elsewhere (most authors of fiction are not prolific enough to approach to size of the corpora used here). This is the main reason why 500 corpora were employed, to ensure that there were enough corpora to ensure that the individualized nature of language usage was being measured. Additionally, as is displayed in Fig. 2, each user corpus seemed to have a relatively unique signal of language usage, which is quite similar to what was found for fiction authors in Johns and Jamieson (2018) and Johns et al. (2020).

The top panel of Fig. 1 contains a histogram of the number of tokens that each user corpus contains, while the bottom panel contains number of types contained in each user corpus. As previously stated, each corpus contains an average of approximately 7.3 million words (with a minimum of 3.8 million words and a maximum of 32.3 million words) for a total of approximately 3.6 billion words across all 500 users, a sizeable sample of language. The user with the lowest number of types was 16,996 and the maximum was 52,032, with an average of 30,008.42 across all users. This in essence represents the user's productive vocabulary size when communicating on the website. As the bottom panel of Fig. 1 shows, there is a fair amount of variability in vocabulary size across users. There is only a moderate correlation between the number of tokens and number of types used across the corpora, with a Pearson correlation of $r = .41$, $p < .001$, signaling that the number of types that a user produced is not completely connected to the total amount of words that they produced on the site. Rather, the number of types that a user employs is likely a reflection of the discourses that they are communicating in, such as those users who communicate in subreddits that require technical knowledge (e.g., some of the users are prolific commenters about computer hardware and computer programming).

In order to gain a better understanding of the vocabulary overlap across the users selected for this analysis, the percent overlap in the types produced (i.e., the overlap in productive vocabulary) across users was computed. This was done by calculating, for each possible user pair, the number of word types that a user pair had in common and dividing that value by the total number of types used across both users. If two users had exactly the same vocabulary, by this metric they would have a 100% overlap, while if they used no words in common, they would have an overlap of 0%. There were 124,750 user pair comparisons done. The result of this analysis is contained in Fig. 3. This figure shows that there is considerable variability in the vocabulary of different users, with the average word type overlap being 51.97%, with a minimum value of 28.87% and a maximum value of 71.19%. This finding suggests that many of the users selected for this analysis have considerable differences in the words that they use while communicating on the site, likely reflective of the topics that they communicate about (the main reason they are on Reddit). Thus, given that the goal of this article is to examine the

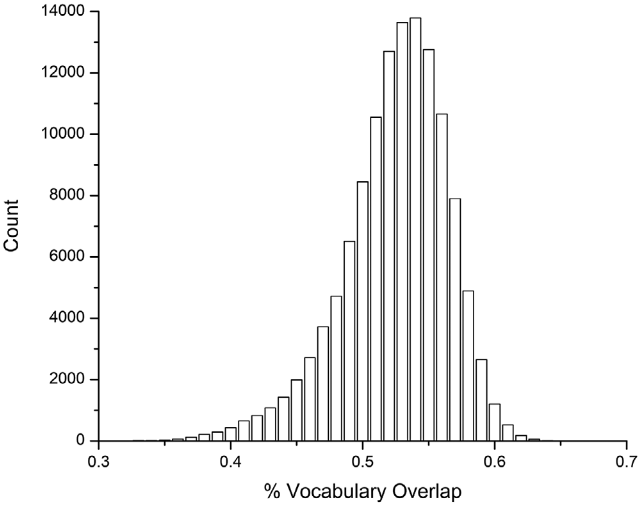

Fig. 3. A histogram of the percent vocabulary overlap for each user pair across all user corpora. Overlap was calculated by assessing the number of words in common that two users had in common and dividing it by the total number of word types the two users produced. This was done for each possible user pair, totaling 124,750 comparisons. The figure shows that there is considerable variability in the vocabularies of the users examined in this article.

alignment of word meanings across individual language users, this simulation suggests that there are substantial differences in the vocabularies of the users selected to examine within language semantic alignment. This is important as it ensures that there is a varied sample of individuals contained in the different user corpora.

1.2. Prior research on using distributional models to calculate semantic alignment

Thompson et al. (2020) used a distributional model type entitled neural embedding models (Mikolov et al., 2013; see Kumar, 2020 for a discussion and review of this model type in the psychological sciences). These models use a predictive neural network to learn word co-occurrences from large corpora. Although there are multiple training procedures for this model type, Thompson et al. (2020) used a skip-gram implementation, where processing a word in a sentence prompts the network to make a prediction about the likely words that should surround that word in that sentential context. An error signal is then used to refine the model's predictions across training. Neural embedding models have been shown to provide an excellent fit to a variety of different behavioral data types (Mandera, Keuleers, & Brysbaert, 2017).

Thompson et al. (2020) trained neural embedding models on 41 different corpora, each corresponding to the different languages that they were analyzing. The semantic alignment of words was then assessed by determining the similarity of the semantic neighbors that a word had across languages (see next section for an in-depth discussion of calculating semantic alignment). In this article, this method of calculating semantic alignment will be referred to as indirect semantic alignment, as it relies on a second-order examination of a word's representation. That is, indirect semantic alignment does not directly compare the representation of a word across corpora, but instead relies upon extracting and comparing the semantic space of words across corpora using a metric based upon the closeness of a word's semantic neighbors.

The use of indirect semantic alignment by Thompson et al. (2020) was dictated by the research question they were attempting to answer. In their analysis, they only had a subset of words that had translations across all languages. Thus, most of the training words used were not consistent across corpora, necessitating the usage of a higher-order examination of the inferred representations. When examining a single language, such as what is being done here, this is not necessary and lexical representations should be able to be directly compared to each other.

However, neural embedding models have a number of components that make it difficult to do direct comparisons across different corpora. One major issue is that prior to learning the network is initiated with random weights within the network, and representations drawn from these random weights are difficult to compare. Additionally, neural embedding models utilize a number of probabilistic functions, such as subsampling (the skipping of high-frequency words) and negative sampling (the generation of unrelated to hone the model's representation), which are dependent on a precomputation of a corpus's word frequency distribution, and thus will operate differently across the different user corpora (Mikolov et al., 2013). Finally, neural embedding models have been shown to have poor performance on smaller corpora relative to other approaches (Levy, Goldberg, & Dagan, 2015), which is typically not an issue in most applications of distributional models but is a challenge for the current work due to the small size of the various user corpora. This poor performance on smaller corpora may be related to the dependence of the probabilistic functions based on word frequency, as the word frequency distributions of small corpora may not be reliable enough to generate the appropriate functions.

Given these limitations, neural embedding models are not an ideal model type to examine the word meanings that are derived from small corpora. Instead, a count-based model that accumulates direct co-occurrence values within a Word × Word matrix (referred to here as the WW model), and modified with transformations based on the training methodologies from neural embedding models, will be utilized here. This method, first described by Johns et al. (2019) and further evaluated by Johns (2021b), has been shown to provide roughly equivalent levels of performance to neural embedding models, while being considerably simpler in implementation. Furthermore, Johns (2023) recently demonstrated that this model can extract sensible distributional representations from small corpora and provide a good accounting of lexical semantic behavioral data. The specifications and advantages of this distributional modeling approach are described below.

1.3. The WW model

The WW model is a count-based that directly accumulates word-word co-occurrences across contexts. The prototypical model of this type is pointwise mutual information (PMI), an information-theoretic measure of the likelihood of two words co-occurring together. PMI has been shown to provide an excellent fit to a variety of lexical behaviors (Bullinaria & Levy, 2007, 2012; Levy & Goldberg, 2014; Levy et al., 2015; Recchia & Jones, 2009) and rivals the performance of neural embedding despite being considerably simpler.

The representation employed by the model used here is a matrix, with each row in the matrix containing a word's representation and each column being a feature of some type. Johns et al. (2019) and Johns (2021b) explored different feature types with different theoretical aims, but here the feature type used will be other words. Thus, the representation will be the number of times that a target word co-occurs with all other words in a sentential context. The resulting matrix has a dimensionality of $W \times W$, where W is the number of words in the model's vocabulary. Each entry in the matrix is the number of times two words occurred together.

The goal of Johns et al. (2019) was to determine whether the success of neural embedding models was due to the use of a prediction mechanism, as has been claimed (e.g., Mandera et al., 2017), or due to the surrounding training methodologies that the model employs. Specifically, Johns et al. (2019) examined the impact of negative sampling, which has been shown to have a sizeable effect on neural embedding model's performance (e.g., Goldberg & Levy, 2014; Levy et al., 2015). In negative sampling, a set number of unrelated words (controlled with a free parameter) are generated. These words are randomly selected based on preprocessed word frequency from the training corpus, with the sampling procedure differing across model implementations. The network is then made to suppress (or become unpredictive) of these sampled words, with the idea being that the network should only predict the words that a word co-occurs with, but not predict words that they do not co-occur with.

Johns et al. (2019) tested whether the use of negative sampling was due to the honing of a prediction mechanism (as proposed by neural embedding models) or whether it integrates some other information type into a distributional representation. It was found that the WW model had massive improvements in variance accounted for in word similarity data when negative sampling was integrated into its training methodology, which should not occur if the utility of negative sampling was to improve the predictions of a model, as the WW architecture is not predictive. Instead, it was shown that negative sampling allows for the base-rate occurrence of words to be integrated into a word's semantic representation, which allowed for the unique associations between words to be highlighted. This finding suggests that the advantages that neural embedding models provide over other frameworks are due to the complexities of the training methodologies employed, not necessarily the underlying processing and learning framework.

Johns et al. (2019) found that negative sampling vastly improved the WW model's performance. Additionally, they described two parameter-free analytical techniques that allowed for negative information to be integrated into a word's representation. The resulting model achieved similar levels of fit to neural embedding models, while eliminating a number of parameters, thus simplifying the approach. Importantly, the approach does not employ any probabilistic functions in its training materials, and in the implementation used here has only one hyperparameter: vocabulary size. Thus, when being trained on the 500 different user corpora, all aspects of the models can be consistent in the training methodologies used. This means that the resulting representations can be easily directly compared without worrying that some other aspect of the model's architecture is limiting the usefulness of the comparisons. Additionally, the use of the transformations with small, individualized corpora in the accounting for word similarity and free association data has been extensively evaluated recently by Johns (2023) to positive results, which strengthens the results found here.

There are three transformations that will be applied to the WW representation: (1) global negative (GN); (2) distribution of associations (DOA); and (3) combined GN + DOA. To apply the GN transformation, the first step is to construct a vector of base-rate of occurrence across the columns of the matrix. This is referred to as the GN vector. The first step to forming this vector is to compute the sum of each column in the matrix:

$$\mathbf{GN}_j = \sum_{i=1}^{n} \mathbf{M}_{i,j}, \tag{1}$$

where **GN** is the global negative vector, **M** is the word-by-word matrix, j is the column being calculated, and i increments through all n rows in the matrix. The elements in the GN vector are based on the frequency of a word across sentences and the window size used by a model (here window size will not be a parameter, as a sentence will be used to update a model's representation). The second step is to unit normalize the vector, making the magnitude of the vector 1, by diving each element in the GN vector by the vector's magnitude:

$$\mathbf{GN}_j = \frac{\mathbf{GN}_j}{\sum_{k=1}^{n} \mathbf{GN}_k}, \tag{2}$$

where k increments through each index in the **GN** vector.

The third step of the GN transformation is to balance the amount of positive and negative information in a model's representation. This is done by computing the sum of a word's row, which represents the amount of positive information a word received during training. This value will be used to add in an equivalent amount of negative information, accomplished with the following equation:

$$\mathbf{M}_i = \mathbf{M}_i - \left(\mathbf{GN} * \sum_{j=1}^{n} \mathbf{M}_{i,j}\right), \tag{3}$$

where $\mathbf{M}_i$ is a word's row in the matrix, and j goes through each column in the matrix. The result of the GN transformation is that there is an equivalent amount of negative and positive information contained in a word's representation. In the resulting matrix, if a value is positive it signals that the connection between two words exceeds the base-rate occurrences of those words. If it is negative, then it signals that the words do not have a unique association.

Instead of explicitly balancing the amount of negative and positive information like the GN transformation does, the DOA technique capitalizes upon the fact that the uniqueness of the connection between two words is already contained in the matrix. To take advantage of this situation, the first step in this transformation is to transform the columns of the WW matrix into z-scores:

$$\mathbf{M}_{i,j} = \frac{\mathbf{M}_{i,j} - \mu_j}{\sigma_j}, \tag{4}$$

where i represents a row in the matrix, j represents a column, μ_j represents the mean of the column, and σ_j represents the standard deviation of the column. The resulting values indicate how many standard deviations the association between two words is, compared to the other co-occurrence values in that column, which highlights the uniqueness of those two words.

However, the resulting scores are biased by word frequency, where words higher in frequency will have higher average association scores than lower-frequency words. To normalize this, each value in a row is normalized with a z-score:

$$\mathbf{M}_{i,j} = \frac{\mathbf{M}_{i,j} - \mu_i}{\sigma_i}, \tag{5}$$

where μ_i is the mean of a word's row, and σ_i is the standard deviation of that row. The result of this transformation are normalized association values, where the unique associations between words are highlighted.

In Johns et al. (2019) and Johns (2021b), it was found that the best-fitting models resulted from the GN transformation being applied first, followed by the DOA transformation, suggesting that the transformations are capitalizing upon somewhat different distributional information. Given that both Johns et al. (2019) and Johns (2021b) were applied to much larger corpora than used here, the impact of the three transformation types will be assessed independently in the results section to determine the impact of them when using smaller corpora. The result of the combined transformation on the WW model is that each word is represented by its association to every other word in the model's vocabulary, with each association value being a z-score indicating how much above (or below) that word's association is compared to the baseline level of association. As demonstrated by Johns et al. (2019), most of the z-scores in a word's representation end up around 0, with a small subset of positive values which serve to highlight the important associations that a word has.

The initial vocabulary size for the model will be 80,150 for the WW model (thus, the model's memory matrix has a dimensionality of 80,150×80,150). This will change in the results section, however, as the specific words that are consistent across all users are identified and analyzed with the semantic alignment measures defined below.

1.4. Semantic alignment measures

There will be two semantic alignment measures used in this article—direct and indirect semantic alignment. Indirect semantic alignment will be equivalent to the methodologies employed by Thompson et al. (2020), while the direct metric will be unique to this article, made possible through the use of the above-outlined distributional framework.

The direct semantic alignment measure will be the average similarity of a word's representation derived from each user corpus. The similarity metric employed here, and as is standard across distributional modeling, is a vector cosine (normalized dot product), which returns a value between -1 and 1. Vector cosine is calculated as follows:

$$S\ (\mathbf{x}, \mathbf{y}) = \frac{\sum_{j=1}^{N} \mathbf{x}_j * \mathbf{y}_j}{\sqrt{\sum_{j=1}^{N} \mathbf{x}_j^2}\sqrt{\sum_{j=1}^{N} \mathbf{y}_j^2}}, \tag{6}$$

Word = "Freedom"

Comparison	Cosine
User1 – User2	.2
User1 – User3	.15
User1 – User4	.25
User2 – User3	.1
User2 – User4	.3
User3 – User4	.18
Sum =	1.18
Direct Alignment =	1.18/6 = .197

Fig. 4. An example of calculating direct alignment for the word freedom for four users.

where N is the size of the vectors. To compute the direct semantic alignment of a word, the similarity between each user's representation was computed and these similarity values were averaged. Thus, the direct semantic alignment measure reflects how similar the representation of a word is across the 500 user corpora. If every user had produced a word, there is a maximum of 124,750 comparisons possible for that word.[2] Fig. 4 displays a hypothetical example calculating the direct alignment for the word *freedom* when only four users are included in the computation. This figure shows that direct semantic alignment is the average cosine of all of the pairwise cosine similarity for each user pair.

As explored in the results section, the number of comparisons will be controlled by a frequency parameter, where for a user to be included in the computation of the semantic alignment of a word, they would need to use a word a minimum number of times. This parameter will be included to ensure that each user has enough samples of a word's usage to form a stable representation of a word's meaning, and the impact of this parameter will be explored in the results section. The frequency parameter will also be used when computing indirect semantic alignment, thus ensuring that both measures are using the same underlying representations, but in a different fashion.

The indirect semantic alignment measure will employ the same methodology as Thompson et al. (2020). The computation of this measure occurs across multiple steps. Fig. 5 contains an example of the computation of the indirect alignment of a single word (*freedom* in this example) for two users (see Thompson et al., 2020 for additional illustrations and more specific details of how indirect semantic alignment is calculated). As is shown in the figure, the first step in calculating the measure is to determine the n nearest neighbors for a word from the first representation. A vector of size n is then formed, consisting of the vector cosine of these neighbors. A second vector is formed by taking these same words and assessing the similarity values for those same words for the second representations, resulting in a second vector of size n composed of those cosine values (the similarity of the same words as is contained in the first vector). Then, the similarity between these two vectors is computed with a vector cosine. This is the first value used in the resulting semantic alignment measure. The second value is simply the reverse of this process, where the n nearest neighbors of the second representation

Step 1 – Determine Nearest Neighbors

Word = "Freedom"

User1	User2
Religion	Speech
Liberty	Protecting
Rights	Idea
Expression	Free
Absolute	Belong
Mobility	Privacy
Violation	Defending
Institution	Loving
Assembly	Crowded
Economics	Restricted

Step 2 – Calculate User1 to User2 Alignment

Word	User1	User2
Religion	.3	.15
Liberty	.28	.17
Rights	.25	.12
Expression	.21	.2
Absolute	.2	.22
Mobility	.14	.08
Violation	.11	.11
Institution	.09	.13
Assembly	.06	.07
Economics	.05	.09

Cosine = .85

Step 3 – Calculate User2 to User1 Alignment

Word	User1	User2
Speech	.15	.27
Protecting	.17	.26
Idea	.09	.23
Free	.14	.2
Belong	.11	.18
Privacy	.2	.15
Defending	.18	.14
Loving	.08	.11
Crowded	.07	.1
Restricted	.1	.08

Cosine = .93

Step 4 – Average Alignment Values

Alignment = (.85 + .93)/2 = .89

Fig. 5. An example of calculating indirect alignment for the word freedom for two users. This process is repeated for each user−user pair and the final indirect alignment is the average of all of those comparisons.

was used to drive the comparison, resulting in a second neighborhood similarity value. The final semantic alignment is the average of these two values. This process is then repeated for each possible comparison—languages in the work of Thompson et al. (2020) and user corpora here—and averaged for each word that was analyzed. The final value of neighborhood size (the *n* parameter above) that was used by Thompson et al. was 100, which will be the value used here.

As stated, the use of indirect semantic alignment by Thompson et al. was necessary based upon their research goals. Here, there is no such limitation, so one of the first analyses performed in the results section will be to contrast and compare these two metrics, and to determine their connection to various psycholinguistic variables. This will allow for a determination of what the best methodology to measure semantic alignment within a language is.

1.5. Implementation details

The computation of the semantic alignment values used below faced a number of technical challenges, due to the number of corpora used here and the number of words that will be analyzed.[3] To compute both the indirect and direct semantic alignment measures, the first step was to compute the WW model transformed with the combined GN and DOA transformations for each user corpus, which were then stored separately as individual files (stored as a matrix). To compute the direct semantic alignment measures, two user matrices were held in main memory (which required substantial amounts of RAM given the size of the matrices) and the alignment of words was attained between those two users was recorded. This was repeated for each possible user pair, resulting in 124,750 comparisons across all users, a task which took over a week to compute using a 12-core Intel Xeon processor put into parallel using 256GB of RAM. The alignment measures were then averaged for each word.

To compute the indirect semantic alignment measure, a similar strategy was employed. However, the first step was for each user to compute the similarity matrix for all words in the representation. It is necessary to determine all similarity values, not just the 100 nearest neighbors, due to it not being possible to know what cosine values will be necessary in the second step of computing the indirect alignment measure. Then, the indirect semantic alignment for each word was calculated for each user pair (with the same number of comparisons as for the direct measure), and averaged.

All resulting metrics used in this article are available at https://osf.io/rsjkc. Due to the size of the various corpora and resulting derived representations (in total, the corpora and derived representations and similarity matrices require about 1TB of hard drive space), as well as privacy concerns, it is not feasible to share them publicly; however, they are available upon request from the author. Code written in the Java programming language that was used to derive the semantic alignment measures from these materials is also available upon request.

2. Results

There will be nine different simulations reported. Simulation 1 will demonstrate that the user corpora are extracting a sufficient amount of semantic information to be useful in examining alignment within lexical semantic memory. Simulation 2 will examine the impact of the transformations on the fit of the user models. Simulation 3 will provide an examination of the two semantic alignment metrics, followed by Simulation 4 which will determine their respective fits to multiple psycholinguistic variables, in order to determine the best semantic alignment measure. Following this, the impact of different user metadata on semantic alignment will be done. Specifically, Simulation 5 will determine the impact that the similarity of discourse communication patterns has on across user semantic alignment. Simulation 6 will examine the semantic alignment of users to themselves, while Simulation 7 will examine the impact of discourse on within user semantic alignment. Finally, Simulation 8 will examine the impact of syntactic category on semantic alignment, while Simulation 9 will examine semantic categories.

2.1. Simulation 1: Small corpora and similarity performance

Compared to previous studies utilizing distributional models to examine language processing, one unique aspect of the following analyses is the use of many different small corpora that are relatively small in size. One issue that must be considered when examining the semantic representations that are derived from small corpora is whether the word representations that are constructed are stable and contain enough semantic information in them to drive a meaningful semantic alignment measure. This concern is due to the limited number of training samples available for each word that using small corpora afford, which may result in the word representations not containing enough semantic information about a word's meaning to be reliable.

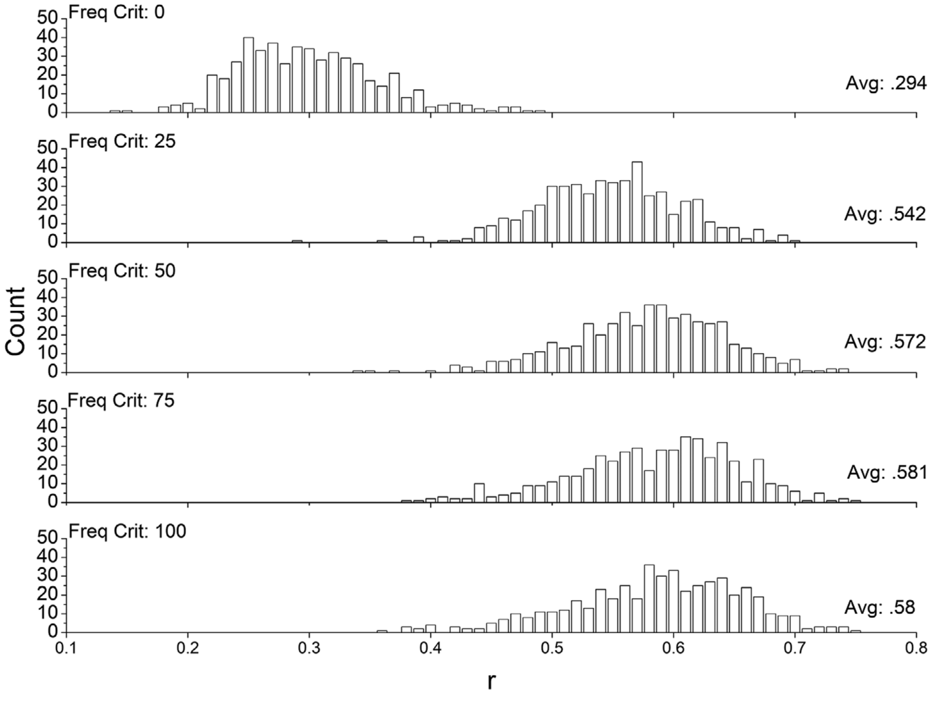

Fig. 6. Histograms of the Spearman correlations for the word pair similarity values for each user corpus to the Bruni et al. (2012) word similarity data, manipulated by the frequency criterion. Each panel corresponds to a different frequency criterion. The frequency criterion sets the minimal level of sentences that each word in a pair must appear in to be included in the analysis.

To determine the impact that the number of training samples has on model performance, an initial simulation was done using the word similarity data of Bruni, Boleda, Baroni, and Tran (2012), a standard dataset for evaluating distributional model performance. In a word similarity task, participants are prompted with pairs of words and must rate how similar they believe they are on a given scale. This dataset was used here because it consists of 3000 word-pairs, a relatively large number of pairs for this kind of dataset. Each user corpus was assessed independently. To evaluate the impact of the number of training sentences on model performance, a frequency criterion was set such that if both words appeared in a number of sentences that exceeded this criterion, then that pair was included in the analysis. Otherwise, the pair was not included. The criterion was set at 0 (meaning that each word had to have occurred at least once), 25, 50, 75, and 100 sentences. Repeated uses within a sentence were ignored. Increasing the criterion caused fewer word pairs to be included but it will allow for a determination as to what impact receiving more training exemplars for each word has on model performance. The simulation was run with the combined GN+DOA matrix transformations described above.

The results of this simulation are contained in Fig. 6, with each criterion level having its own histogram of Spearman correlations across the 500 user corpora. For the 0, 25, 50, 75, and

100 comparison levels, there were 2545.9, 906.85, 604.16, 437.68, and 340.96 average number pairs for each corpus included in the analysis, respectively. This figure shows that when the frequency criterion is set at 0 (where each word just needed to occur once), indicating that when some word pairs had minimal levels of experience, the model struggled. However, when the criterion was increased to 25, the model achieved a relatively high level of performance. Model fits plateaued at a criterion of 50, suggesting that at this level of sampling the derived word representations are relatively stable.

To compare the performance of the individual user corpora to a larger corpus, performance was calculated when the model was trained on all 500 user corpora at once. When the model was trained on all corpora at once, the Spearman correlation was $r = .764$, $p < .001$ signaling that including all corpora provides an improvement in model performance (the best user corpus achieved a correlation of $r = .739$, $p < .001$ with a frequency parameter of 50; however, as Fig. 6 shows, most user corpora fell far below this level of fit).

This simulation establishes that even though the various user corpora are relatively limited in size, and likely contain less diverse language than other sources of written language due to being from a single individual, they still contain enough semantic information to give reasonable fits to word similarity data. This result provides confidence that the resulting semantic alignment values will be based upon realistic representations of word meanings. Johns (2023) provides a much broader examination of the ability for distributional models trained on small corpora to account for lexical semantic behavioral data.

2.2. *Simulation 2: Impact of transformations*

A second question that needs to be answered before examining the semantic alignment results is how the various transformation techniques deal with the smaller user corpora, as they have previously only been applied to large corpora (Johns et al., 2019; Johns, 2021b). To determine the impact of the various transformations, Fig. 7 contains histograms of the Spearman correlations for the 500 user corpora for the model with no, GN, DOA, and GN+DOA transformations. A frequency criterion of 50 was used in this simulation. This figure shows results that are coherent with past results, as both the GN and DOA transformations provide substantial improvements over the untransformed representations. However, here the DOA transformation provides a much larger improvement over the GN transformation, suggesting that this is the better technique when dealing with small corpora. As described in Johns et al. (2019), the GN transformation is most analogous to the negative sampling procedure used by word2vec, which has been shown to struggle with small corpora (Levy et al., 2015), so the increased advantage for the DOA transformation makes sense in this context. Overall, the best fit is given with the combined GN and DOA transformation, again consistent with past results, and thus this will be the model that will be used through the remaining analyses.

2.3. *Simulation 3: Examination of semantic alignment measures*

A necessary step before analyzing the results of the two semantic alignment variables is to determine the words that will be examined here. Two variables that need to be taken into account when forming the word list are: (1) the frequency criterion described above, and (2)

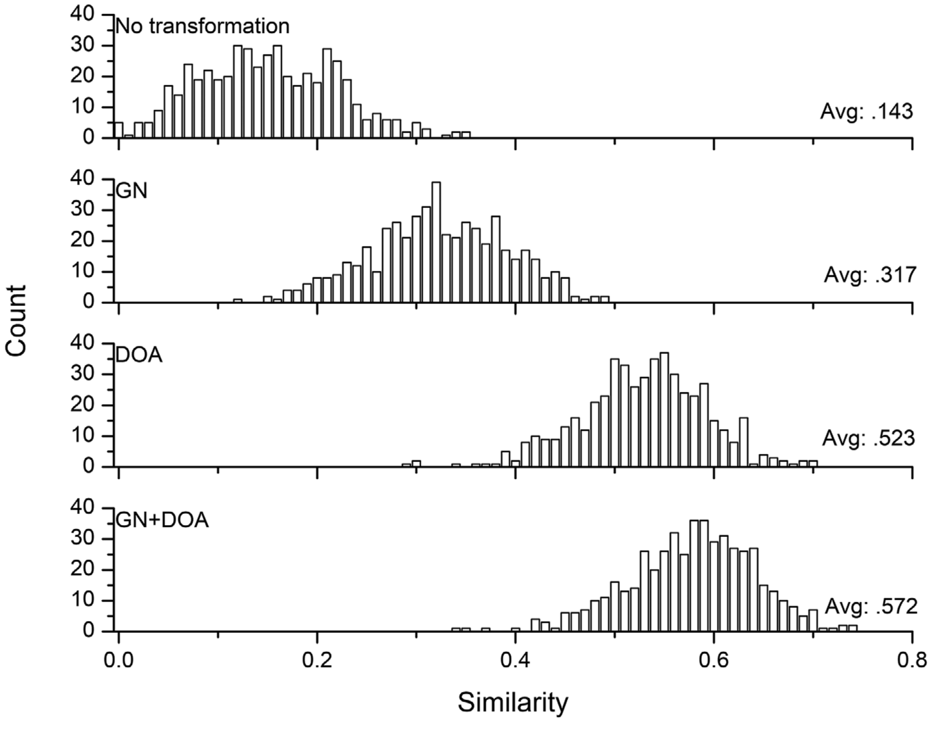

Fig. 7. Histograms of the Spearman correlations for the word pair similarity values for each user corpus to the Bruni et al. (2012) word similarity data for the four different transformations of the WW model. The results displayed in this figure are consistent with past findings, showing that the combined GN+DOA transformation offers the best performance.

the number of users who produced a word. It is necessary to consider these two variables to ensure that there are enough samples to provide a reliable estimate of semantic alignment. It was determined that for a word to be included on the word list, it had to be used at least 50 times (this criterion was decided based on the results of the simulation contained in Fig. 6) by 25 different users.

The use of these criteria resulted in a word list of 10,870 words. Although the selection of 25 different users is somewhat arbitrary, this allowed for at least 300 user−user comparisons for each word contained in the list (most words will have substantially more than this). A total of 674 words were used by all users (mostly high-frequency function words), with a declining usage after this point. To visualize the number of comparisons that was done for each word, Fig. 8 contains the number of comparisons that was done across words, ranked from most to least. This shows that most words receive a substantial number of samples, which provides confidence that the alignment numbers are reliable. The median number of comparisons for a word was 7348, signaling that most words received a substantial number of comparisons in the construction of their semantic alignment values. Overall, there were 352,304,340 comparisons done across words for both the direct and indirect semantic alignment metrics.

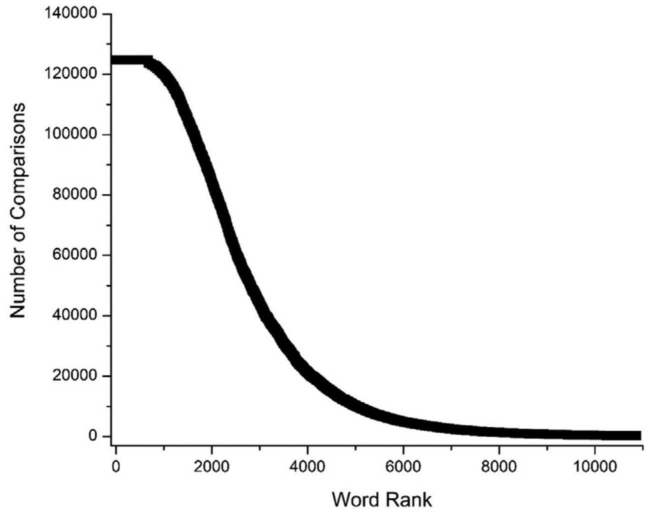

Fig. 8. A ranked list of the number of comparisons that each word received when computing the two semantic alignment values.

To visualize the two alignment measures, Fig. 9 contains histograms of the direct and indirect semantic alignment values for the 10,875 words analyzed. The first noticeable trend in this figure is that the indirect semantic alignment values are shifted negatively compared to the direct semantic alignment values. However, it is difficult to know what exactly this finding represents as they are measuring somewhat different aspects of a word's representation. The Pearson correlation for all words between the direct and indirect values is $r = .745$, $p < .001$, signaling that the two measures are quite associated with each other but that there are some divergences between them.

To gain a better understanding of these differences, Fig. 10 contains the ranked direct and indirect alignment values with standard deviation error bars included for each word. In this figure, words are ranked by their magnitude in the respective alignment measures (and thus do not map onto the same words across the two panels). The standard deviation represents the variability in the semantic alignment across user pairs. This figure shows that the indirect comparison technique has much more variance across words compared to the direct comparison. This finding suggests that the more limited semantic space that the indirect measure is capturing leads to an increase in the variance in the alignment values across words.

As Fig. 9 displays, the scale of the semantic alignment values is relatively limited, and the values are quite close to zero. It is unclear whether this is due to the mathematical underpinnings of the modeling approach or whether the alignment metrics are failing to assess similarity in the usage of the same word across individual language users. To ensure that the alignment measures used here are actually capturing similarity in word usage, an additional simulation was conducted where random semantic alignment values for different words were compared against the semantic alignment values for the same words. To accomplish this, the average semantic alignment value for every user to every other user was calculated

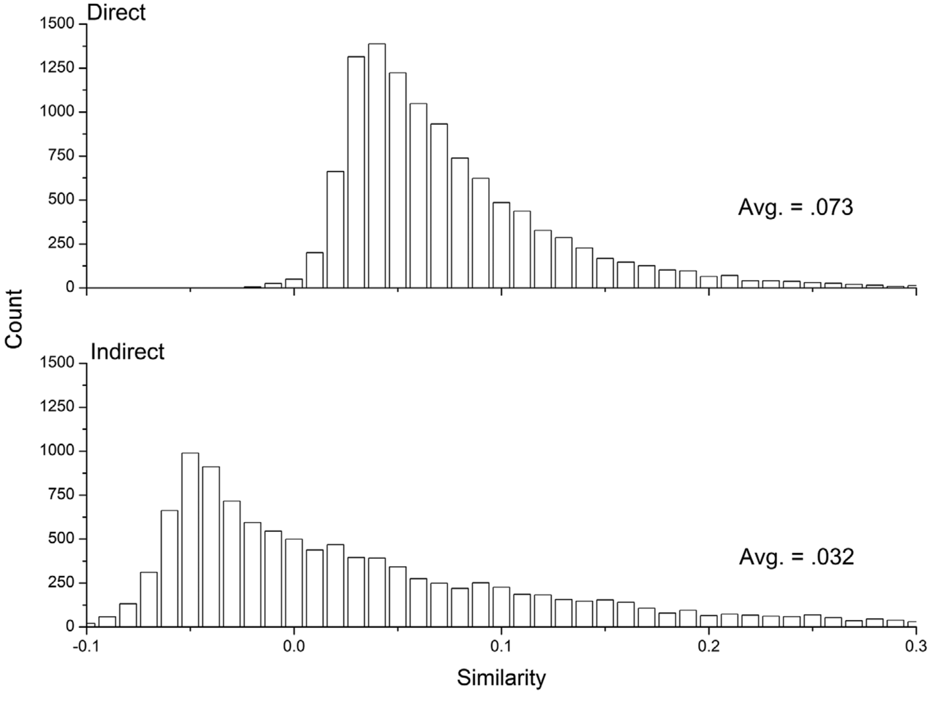

Fig. 9. Histograms of direct and indirect semantic alignment values for the 10,875 selected words.

(i.e., the average of all the semantic alignment value for each word between each user pair was assessed and averaged). To construct randomized semantic alignment values, a Monte Carlo simulation was done where for each user pair the average semantic alignment value was taken between 100,000 random word pairs (where both words in the pair exceeded the frequency parameter). Only the direct semantic alignment measure was used in this simulation.

The histogram of these values is contained in Fig. 11, which shows that the average semantic alignment for the different word comparisons is very distinct from the same word comparisons. Specifically, this figure shows that all the average different word comparisons are very close to a value of 0, signaling no semantic alignment when using direct semantic alignment. However, the average same word comparisons are all well above this level, demonstrating that there is a systematic semantic alignment of words across individual language users. This simulation demonstrates that there is consistency in the semantic representations of words used across individuals, although there is significant variability in this consistency.

To further examine the impact that semantic alignment is having on the semantic representation of words across user models, an additional simulation was done looking at the consistency of semantic neighborhoods for a word that has a high level (*solar*; direct alignment = .355), a medium level (*economy*; direct alignment = .149), and a low level (*concept*; direct alignment = .097) of direct semantic alignment. Semantic neighborhood consistency was determined by computing the 50 nearest neighbors for each word across all users. Then, the

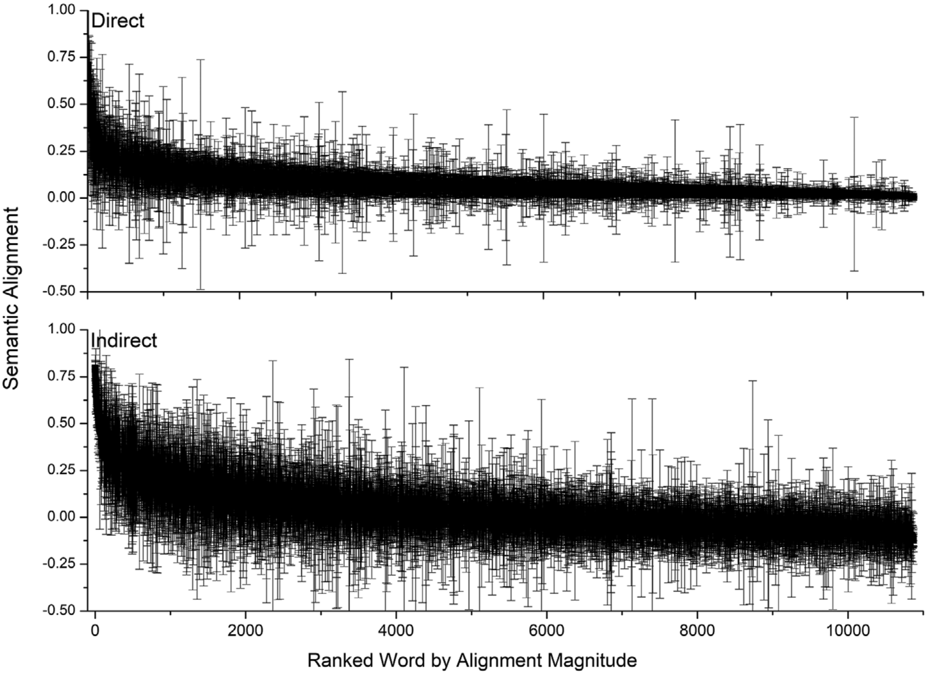

Fig. 10. The average direct (top panel) and indirect (bottom panel) semantic alignment value for each word assessed, ranked by alignment magnitude (not word frequency). Error bars are standard deviation. This figure shows that the use of indirect semantic alignment results in considerably more variance across users compared to direct semantic alignment.

percent probability that a given word was included in a user's semantic neighborhood was calculated. These percentages for the three target words for the top 25 included neighbors are displayed in Fig. 12. This figure shows that for the highly aligned word *solar*, there was considerable consistency in the semantic neighborhoods of different users. For example, the word *panels* was a near neighbor of *solar* for approximately 80% of users. However, as the level of semantic alignment went down, so did the consistency of the semantic neighbors across users, with their being low levels of agreement in the semantic neighborhoods for the words *economy* and *concept*. This suggests that these words are being used by individuals in appreciably different ways across their comments, leading to divergences in their nearest neighbors across users.

To determine if this variability is due to the connection to other psycholinguistic properties of a word, the correlations between the alignment measures and other word-level psycholinguistic variables will be assessed next.

2.4. *Simulation 4: Connection to psycholinguistic variables*

For semantic alignment to be considered a useful theoretical construct to psycholinguistic researchers, the metrics should have some connection to previously proposed variables. That

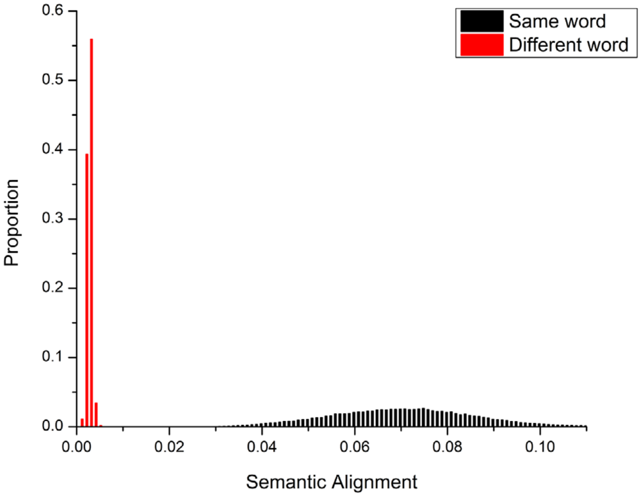

Fig. 11. A histogram comparing the average user−user direct semantic alignment value for randomly selected different word pairs versus the same word. The same word values are the average semantic alignment for each user pair, while the different word values are the result of a Monte Carlo simulation examining semantic alignment of randomly selected word pairs across user pairs. The results show that same word alignment is much more consistent for the same word versus randomly selected words.

is, there are obvious differences in the meanings that people derive from their experiences, but these meanings should be at least somewhat connected to other variables that have been shown to influence word processing.

To determine how the two alignment measures are connected to different psycholinguistic variables, the correlations between the direct and indirect semantic alignment measures and various psycholinguistic variables were taken. The variables included were: (1) concreteness (Brysbaert, Warriner, & Kuperman, 2014); (2) body−object interaction ratings (Pexman, Muraki, Sidhu, Siakaluk, & Yap, 2019); (3) semantic diversity (Hoffman, Ralph, & Rogers, 2013); (4) Reddit word frequency (Johns, 2021a); (5) contextual diversity (the UCD-SD measure from Johns, 2021; the best fitting current contextual diversity measure); and (6) word length. Concreteness ratings assess how grounded a word's representation is in the perceptual environment. Body−object interaction ratings assess the ease with which the human body can interact with a word's referent and is designed to assess the embodied nature of a word's semantic representation. Semantic diversity assesses the variability in the meaning of the different contexts that a word occurs in. Contextual diversity measures how many different semantic context types that a word occurs in. Word frequency is the number of times a word occurred in the Reddit corpus from Johns (2021a), and word length is related to a number of structural properties of a word, such as morphological and phonological complexity. Notably, Thompson et al. (2020) did not find a significant connection between their alignment measure and concreteness[4] across languages, so this analysis will shed light on some differences in the semantic alignment of words within a language versus across languages.

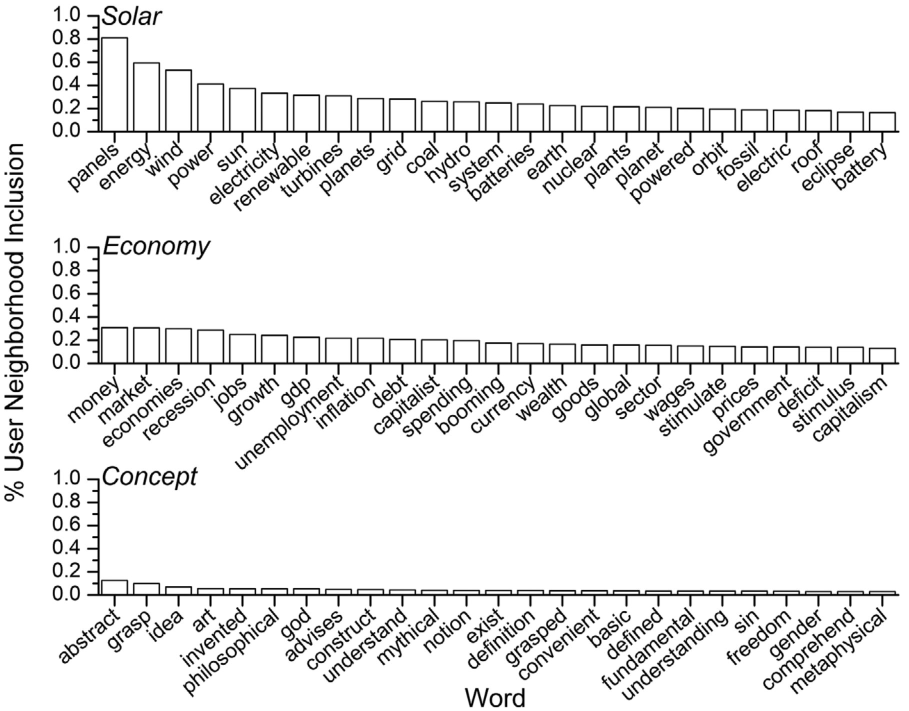

Fig. 12. Results of a simulation examining the consistency of the semantic neighborhoods of a word that has high semantic alignment (*solar*), medium semantic alignment (*economy*), and low levels of semantic alignment (*concept*). This simulation demonstrates that semantic alignment is measuring appreciable differences in the consistency of the usage of words across individual users of a language.

Table 1
Correlations between the alignment measures and the different lexical strength variables

Measure	2	3	4	5	6	7	8
1. Direct	.752	.301	.197	.−.378	.059	−.169	−.204
2. Indirect	.	.229	.062	−.235	.391	.03	−.252
3. Concreteness	.	.	.775	−.369	−.015ϕ	.028 ϕ	−.388
4. Body-object	.	.	.	−.378	−.2	−.157	−.018
5. Sem diversity	.	.	.	.	.428	.575	.008
6. WF	.	.	.	.	.	.816	−.383
7. UCD-SD	.	.	.	.	.	.	−.352
8. Length	.	.	.	.	.	.	.

Note. $N = 7399$ for concreteness, $N = 9860$ for SemD, $N = 3369$ for body−object interaction, and $N = 10{,}237$ for all other variables. ϕ = not significant, all other correlations significant at $p<.01$.

The resulting correlations are displayed in Table 1. This table shows that both the direct and indirect alignment measures correlate positively with concreteness and body−object interaction ratings, suggesting that words that are higher in alignment tend to also be more perceptually grounded. Additionally, both measures are negatively correlated with semantic diversity (a measure of the variability of the meanings of the contexts that a word occurs in), suggesting that words that are more semantically aligned tend to have less semantic diversity (meaning that they tend to occur in more consistent contexts), a comforting finding. In both cases, the direct measure has a stronger correlation to these variables. The only correlation where the direct and indirect alignment measures diverge is in the lexical strength variables, where the indirect measure has a moderate positive correlation to word frequency, while the direct measure does not. The direct measure has a small negative correlation to the UCD-SD variable, but this is likely a reflection of the direct measure's increased sensitivity to semantic diversity, as the UCD-SD measure also takes into account the semantic diversity of the contexts that a word occurs in when calculating a word's contextual diversity.

Taken together, the results displayed in Table 1 suggest that in the case of the small corpora used in this analysis, the direct semantic alignment maps onto other lexical measures, particularly concreteness and semantic diversity, better than indirect semantic alignment. This is likely due to the fact that the direct measure takes into account the entire semantic space of a word (as it is a direct examination of all features contained in two word's representation), while the indirect measure is considering only a limited subsection of it (due to necessarily limiting the analysis to the nearest neighbors of a word). For the remainder of this article, the direct semantic alignment measure will be the sole metric utilized.

An important issue to consider with this analysis is what the psycholinguistic variables indicate about the alignment measures. Although the two measures are rather highly correlated (as one would expect), they do have some divergences. The clearest divergence is the connection to WF, where the direct variable is not correlated to WF, but the indirect is (indeed, it is the highest correlation to the various psycholinguistic variables for either measure). This is not necessarily a positive—being connected to WF is likely an artifact of using a limited sample of cosines in the calculation of the indirect measure, with high-frequency words likely being preferentially selected in semantic neighborhoods. However, the increased connection between the direct alignment measure and concreteness is a positive one, as it is intuitive that words that are more perceptually grounded should have more similar representations across people given that perceptual experience is likely less variable than linguistic experience. This speaks to a general issue of using corpus-based measures to explain linguistic behavior, where the connections between derived variables and behavior need to be crouched within cognitive theory.

Additionally, the finding that both the direct and indirect measures have a significant correlation to psycholinguistic variables such as concreteness and body−object interaction ratings, which Thompson et al. (2020) did not find, suggests that there are different impacts of word-levels variables on word representations within a language as opposed to across languages. This will be discussed further in the General Discussion.

2.5. Simulation 5: Impact of user similarity

One unique aspect of using Reddit as a corpus is that discourse information can be attained for each comment produced, as previously explored by Johns (2021a, 2021b) in lexical organization and lexical semantics. Here, discourse information will allow a determination of the

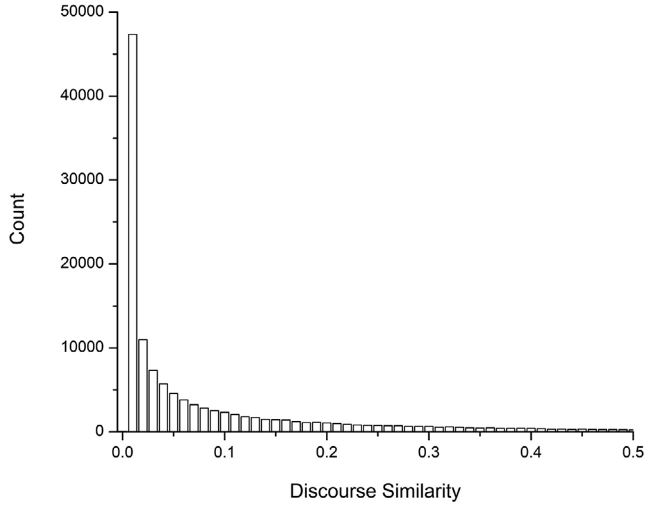

Fig. 13. A histogram of the discourse similarity of all user pairs. This figure shows that most users have relatively distinct commenting patterns on the forum, but with a long tail of users who tend to comment in similar discourses.

similarity of users based on their shared communication patterns across discourse types. The ability to extract discourse-level information is due to the existence of subreddits, with each subreddit being a topic of discussion around a certain topic (e.g., r/CogSci features discussions about cognitive science).

In order to compute user similarity, all subreddits that had five or more of the selected users communicated in were recorded. This resulted in a list of 20,307 subreddits. Each user was then represented by a vector of length 20,307, where each element is the number of comments that the user made in each subreddit. Then, the pairwise similarity was taken between each user pair by taking the vector cosine between each discourse vector, giving a discourse similarity value between 0 and 1 (due to the vectors not containing any negative numbers). This resulted in 124,750 discourse similarity values across all users.

Fig. 13 contains the histogram of the discourse similarity values. This figure shows that most users are relatively dissimilar in their communication patterns, indicating that most of the users selected are unique in terms of the subreddits that they communicate in. However, there is a long tail of higher similarity values suggesting that there are some users who have a considerable overlap in the subreddits that they comment in.

In order to determine the impact of the discourse similarity measure on the semantic alignment values, the average semantic alignment between each user pair was calculated (i.e., for each user pair, the average alignment across all words for those users that exceeded the frequency criterion was computed, equivalent to what was done in Fig. 11 for the same word comparison) and correlated to the discourse similarity values for that user pair, for a total of 124,750 comparisons. The Pearson correlation coefficient between discourse similarity and direct semantic alignment was $r = .359$, $p < .001$, while the Spearman correlation was

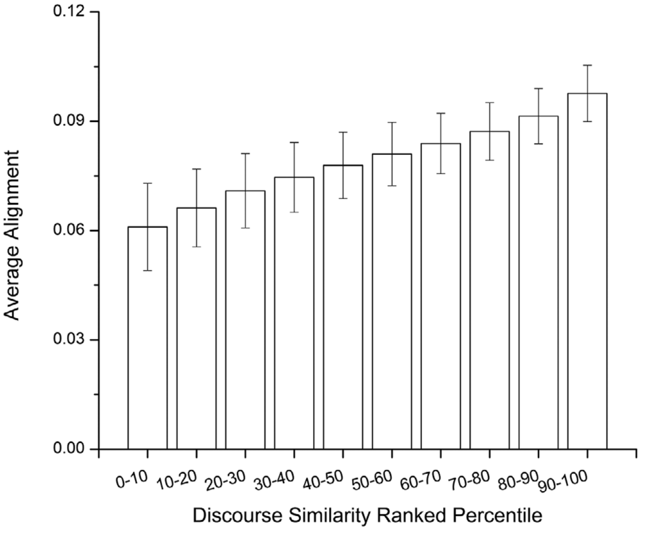

Fig. 14. The average alignment of each user−user pair, placed into 10% ranked bins based upon the discourse similarity of the two users. This figure shows that users who communicate in similar discourses tend to have a greater degree of semantic alignment. Error bars are standard deviation.

$r = .47$, $p < .001$, suggesting that a ranked transformation of the values provides a closer correspondence between the variables. This disparity between correlation metrics is likely due to the alignment values and the discourse similarity values having considerably different distributions (see Figs. 9 and 13, respectively).

To gain an understanding of the shape of the connection between semantic alignment and user discourse similarity, the average semantic alignment values were calculated in ranked bins. The ranked bins contained 10% of each discourse similarity value, ranked from 0 to 100 in steps of 10. The result of this analysis is contained in Fig. 14. This figure shows that there is a linear increase in the semantic alignment values as a function of the discourse similarity values. This suggests that the types of discourses in which an individual communicates within impacts the alignment that two language users have, with users that have similar communicative requirements having more semantic alignment, on average. This finding indicates that there are social and communicative pressures on the word meanings that different individuals derive, a finding that will be discussed further in the General Discussion.

2.6. *Simulation 6: Within versus across user semantic alignment*

So far, semantic alignment has only been analyzed across individual language users. A related, and important, question is how aligned an individual's own usage of words is across time? To determine this, each user corpus was temporally organized by comment date and split in half. Then, semantic representations were derived for each half and compared to each other to generate a within-user semantic alignment value. This was repeated for all users

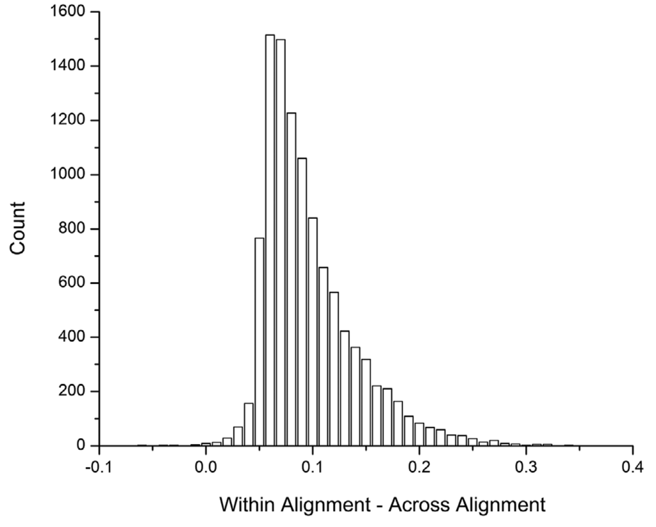

Fig. 15. A histogram of the difference scores for within and across alignment. This figure shows that an individual's usage of language is more consistent compared to language usage across individuals.

and the values were averaged. The resulting semantic alignment values will be referred to as within alignment (vs. across alignment from the previous analyses). Given that within alignment necessary limits the number of comparisons possible (down to a maximum of 500), the frequency criterion was relaxed to 25. In addition, a minimum number of 10 users had to have semantic alignment values for a word to be included in the resulting analysis. The use of these parameters led to within alignment values for 10,616 out of the previous 10,870 words examined.

The average within semantic alignment values was .165 (SD = .075), a considerable increase over the across semantic alignment values (M = .073, SD = .06), with this difference between the two alignment measures was highly significant [$t(10{,}615) = 215.67$, $p<.001$]. To gain insight into how the within semantic alignment values compared to across semantic alignment, Fig. 15 contains a histogram of difference scores (within alignment – across alignment) for each of the 10,616 words. This figure shows that for almost all words the within alignment values are considerably higher than the across alignment values (indeed, only 12 words had an across alignment score that was higher than their within alignment score). This finding suggests that word usage within individuals is much more stable than across individuals, an intuitively satisfying result. However, across user alignment had a strong correlation to within user semantic alignment, with an $r = .812$, $p<.001$, suggesting that the alignment of the words is consistent across and within users, but with an increased magnitude for the within alignment values. This entails that words that have consistent alignment across users also tend to have consistent alignment within a user. That is, words that are used in a similar way by different people across contexts tend to be used by the same person in a similar way across contexts. This is likely related to the attachment to concreteness and semantic diversity that both variables share.

Table 2
Correlations between the within and across direct alignment measures and the different lexical strength variables

Measure	Within alignment	Across alignment
Concreteness	.402	.3
Body-object	.296	.202
Sem diversity	−.468	−.379
WF	−.201	.062
UCD-SD	−.359	−.169
Length	−.205	−.202

Note. $N = 7257$ for concreteness, $N = 9661$ for SemD, $N = 3369$ for body−object interaction, and $N = 10{,}021$ for all other variables. All correlations significant at $p < .001$.

To determine how connected the within semantic alignment values are to the previously outlined psycholinguistic variables, Table 2 contains the Pearson correlations for both within and across alignment for all words available. This table shows that for every variable tested, the within semantic alignment values has a stronger fit compared to the across alignment values. This finding suggests that the impact of the various psycholinguistic measures, such as concreteness and semantic diversity, exerts more influence on the consistency of an individual's usage of language, compared to across individuals.

2.7. *Simulation 7: Impact of within user discourse communication patterns*

Fig. 14 demonstrates that users who communicate within the same discourse types tend to be more semantically aligned. A related question is what impact communicating in different discourses has on an individual's usage of language. That is, if communicative requirements of different discourses shift the meaning of words across users, is there an equivalent pattern found within a single individual language user, where the meaning of words that a single individual uses changes depending on where and with whom they are communicating.

To answer this question, all users who communicated within a single discourse on Reddit between 35% and 65% of the time were found. This criterion was set so that a user corpus could be split into a single discourse corpus (the subreddit where the person predominantly communicated within) versus an across discourse corpus (all other comments that the user made on the site). Overall, 113 users fit this criterion, which will reduce the number of comparisons possible in this analysis compared to previous ones. Two within user semantic alignment measures were then constructed from these corpora: (1) within discourse alignment and (2) across discourse alignment. Within discourse semantic alignment was calculated by splitting the single discourse corpus in two, with a temporal split being employed, and assessing the similarity of words from the model trained separately on the two halves. Across discourse semantic alignment was calculated by training the model on the single discourse corpus and the across discourse corpus separately and computing the semantic alignment of words. As in the previous simulation, a frequency criterion of 25 was used and a word had to have at least 10 users had to have used a word to be included in the resulting analysis. This resulted in 4971 words, a reduction over previous levels, but still a sufficient number of words to provide

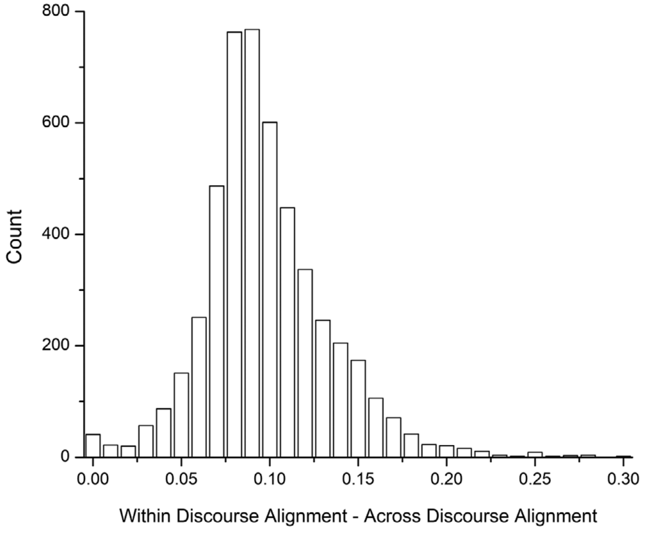

Fig. 16. A histogram of the difference scores for within discourse and across discourse alignment. This figure shows that the semantic alignment for an individual is greater when communicating within a single discourse than across discourses.

concrete evidence about how semantic alignment changes depending on the discourse a user is communicating within.

The average within discourse semantic alignment values was .238 (SD = .074), while the across discourse semantic alignment values was .145 (SD = .073), signaling that there is more within user semantic alignment when a user is communicating within one discourse compared to across discourses. This difference is highly significant [$t(4984) = 172.61$, $p < .001$]. There were only 41 words that had a higher within discourse semantic alignment compared to across discourse semantic alignment. Fig. 16 contains the average within discourse minus across discourse difference scores for all words. This figure shows that for the majority of words, there is a fairly large and consistent difference in these measures, signaling that individuals shift the meaning of the words that they are using when consistently communicating within a single discourse as compared to different discourses.

To further evaluate the semantic representations being developed by the user models trained on the within versus across discourse corpora, an additional simulation was conducted examining the semantic neighborhoods of words trained on the different corpus types. If words are being used more consistently within a discourse, then this should lead to more similar semantic neighborhoods as compared to words being used across discourses, where the words may be used more flexibly in relation to the communicative requirements of the different discourses. To accomplish this, each of the within and across discourse corpora were split into two random sections and representations were generated from each portion, leading to four models for each user (two within and two across models). Then, for each of the above specified 4971 words, the 100 closest neighbors were found for the different models.

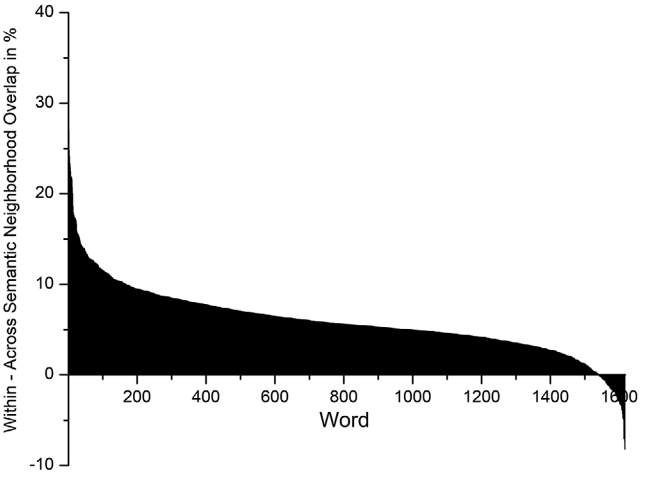

Fig. 17. A simulation examining the difference in neighborhood overlap for the within discourse corpora compared to the across discourse corpora. Each bar represents the percentage difference in within discourse neighborhood overlap over across neighborhood overlap for a single word. This figure demonstrates that the within discourse corpora are generating more consistent word meanings than the across discourse corpora.

However, the semantic neighborhoods were only generated for each word that exceeded a frequency parameter of 100. This increased frequency criterion was used to ensure that a target word had a strong representation, given that each model is only being trained on a quarter of already small corpora. This led to a set of 1617 words being used in the following analysis. For each of these words, the closest 100 neighbors were calculated across the four models. Then, the percent overlap between the neighborhoods for the two within discourse representations was calculated, as well as the percent overlap between the neighborhoods for the two across discourse representations. These values were then summed across all users to attain an average overlap value.

Overall, it was found that for the within discourse representations, there was an average of 22.9% overlap in the semantic neighborhoods of words, while for the across discourse representations, there was an average of 17.03% overlap. A paired-sample t-test showed that this was a highly significant difference [$t(1616) = 64.23$, $p < .001$]. Fig. 17 contains the average difference between the within discourse overlap percentage and the across discourse overlap percentage for all words. This figure shows that almost all words had a higher within discourse neighborhood overlap, with only 77 words having more neighborhood overlap with the across discourse representations. This provides further evidence that word meaning relativity does not hold just across different individuals but also within an individual as well, where the flexibility of word meaning representations allows for words to be used by the same user in different ways across discourses. These differences are likely due to communicative requirements when communicating with different sets of people about different topics.

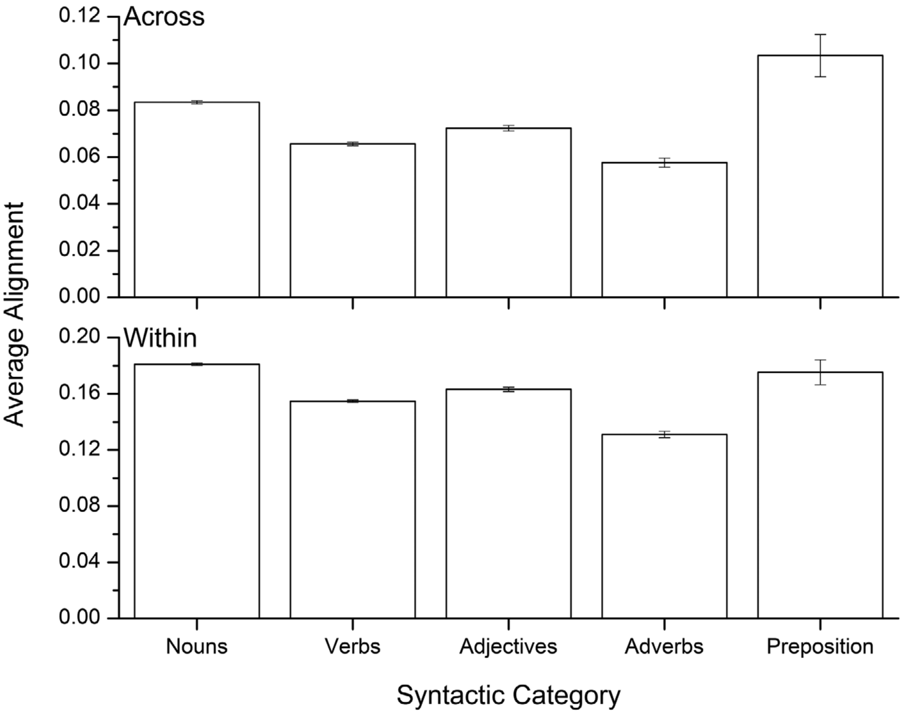

Fig. 18. Average within and across semantic alignment for five different syntactic categories.

2.8. *Simulation 8: Impact of syntactic category*

The next question that will be examined is how semantic alignment is impacted by the syntactic category of a word. To assess this, the across and within direct semantic alignment values were assessed for five different syntactic categories: (1) nouns ($n = 7152$); (2) verbs ($n = 5345$); (3) adjectives ($n = 2484$); (4) adverbs ($n = 899$); and (5) prepositions ($n = 118$). The syntactic category for a word was retrieved from the MRC psycholinguistic database (Wilson, 1988). Words could be assigned to more than one syntactic category.

The results of the simulation are contained in Fig. 18, with the data for the across semantic alignment values being displayed in the top panel, while the within user semantic alignment values are contained in the bottom panel. This figure shows that nouns are more aligned than verbs, with the modifiers of those categories showing the same pattern. The only divergence of the two metrics is for prepositions, where the across values for this category had the highest average alignment, while for the within alignment metric, it was slightly less than what was found for nouns. This suggests that there could be differences in the alignment within and across users, but this may be due to the small sample size of prepositions compared to the other syntactic categories.

Overall, this simulation suggests that nouns (and modifiers of nouns) are more consistent in meaning across individuals than verbs. However, this could be related to the connection

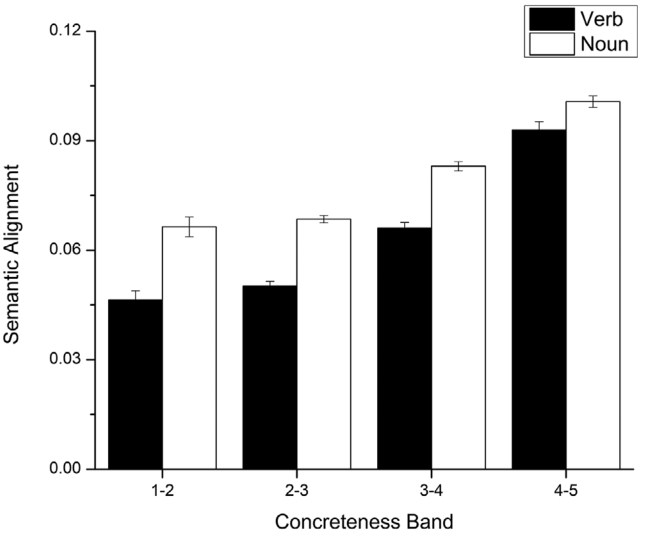

Fig. 19. Average alignment values for nouns and verbs across concreteness bands.

between semantic alignment and concreteness previously found. To determine whether the greater alignment for nouns was due to concreteness and not semantic alignment, the concreteness values for all nouns ($n = 4986$) and verbs ($n = 2894$) from the above-analyzed categories were attained. Then, average semantic alignment was calculated at four bins of concreteness ratings: 1−2, 2−3, 3−4, and 4−5. Fig. 19 contains the results of this analysis and shows that across all concreteness bands, the nouns have significantly higher alignment than verbs ($p<.001$). However, in the 4−5 band, the difference in magnitude between nouns and verbs is reduced, suggesting that very concrete words have greater levels of alignment across the two syntactic categories.

2.9. Simulation 9: Impact of semantic category

One of the main points of analysis from Thompson et al. (2020) was to determine how universal the concepts contained in semantic categories are. To accomplish this, they generated a set of categories that were likely to be used across cultures. Given that the aim of this article is about understanding the alignment within a language, these categories are not necessarily relevant to the current analysis. Instead, here category production data from the standard Van Overschelde, Rawson, and Dunlosky (2004) norms will be used to examine the semantic alignment of different semantic categories.

Specifically, 59 categories were extracted from these norms. The first analysis will examine the average alignment values for each category. For a category to have an average alignment assessed, they needed to contain at least 10 words from the word list. There were 29 categories that reached this criterion, consisting of 501 words. The average alignment values from the within and across measures were highly correlated, with an $r = .899$, $p < .001$, so only the across alignment values will be analyzed. The resulting averages are contained in Fig. 20. This

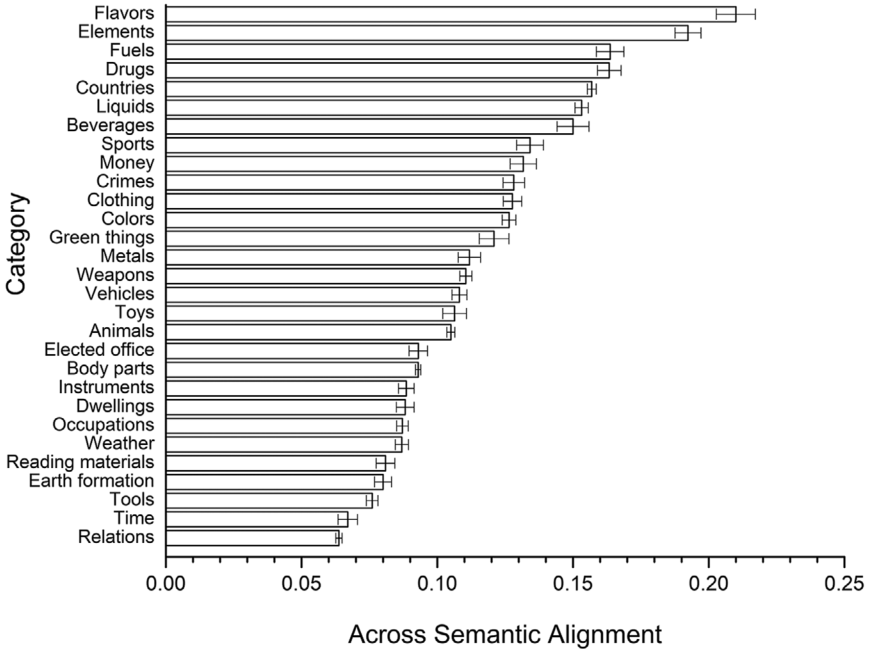

Fig. 20. The average across alignment values for 29 categories from the Van Overschelde et al. (2004) norms.

figure shows that different semantic categories vary widely in their alignment values, with categories like *food flavors* and *chemical elements* having a high level of alignment compared to categories like *family relations* and *time measurements*, which have a relatively low level of alignment. This is somewhat in contrast to the results of Thompson et al., who found that the time category had a high level of semantic alignment across languages, suggesting that there can be differences in semantic alignment for within a language as compared to across languages.

A further piece of data that will be analyzed is the connection between category production probability (the probability of a category member being produced by participants when given a category cue) and semantic alignment. Due to different categories having a different range of probabilities and semantic alignment values, all data were transformed to ranks within a category (e.g., the most prototypical category member was given the rank of 1, the second most prototypical category member the rank of 2, etc…), so that both data and corresponding semantic alignment values were on the same scale. Additionally, the requirement for at least 10 category members to be included was dropped, so that all categories were analyzed. This raised the number of words in the analysis to 672. Strong correlations were found between the category production data and both across and within semantic alignment, with an $r = .601$, $p < .001$ for across semantic alignment and an $r = .478$, $p < .001$ for within semantic alignment. This signals that words that are more prototypical of a category are also the most semantically aligned across individuals. The strong correlation for across semantic alignment signals that words that are more prototypical of a category have more consistency across individual language users, suggesting that categorical structure has an impact on the language environment that different individuals are apart of, with more prototypical category exemplars having a more consistent usage pattern.

3. General discussion

The overall goal of this study was to determine the relativity of word meanings across individual language users. This follows previous work by Thompson et al. (2020) examining the relativity of word meanings across languages. To accomplish this, 500 corpora of prominent users on the internet forum Reddit were constructed, following recent work using Reddit to examine distributional properties of words in a communicative environment to examine lexical organization and lexical semantics (Johns, 2021a, 2021b, 2022, 2023; Johns & Jones, 2022). A count-based distributional framework (first described in Johns et al., 2019 and further explored by Johns, 2021b, 2023), which integrates optimization procedures from neural embedding models, was used to derive semantic representations for words from the user corpora. These corpora were relatively small compared to most modern corpora used in distributional modeling, ranging between approximately 3.2 and 32 million words each, but it was found that each individual user corpus contained sufficient amounts of linguistic knowledge to provide adequate fits to word similarity data (see Johns, 2023 for a further evaluation of aggregating individualized semantic models to fit to lexical semantic data).

The motivation for this research was multifold. As stated, one motivation was to build upon the work of Thompson et al. (2020) in examining the overall relativity of the language environment through an examination of the word meanings extracted from individualized models of distributional semantics. An additional motivation was to extend distributional modeling down to smaller scales to examine the cognitive plausibility of these models at a realistic level of experience (see also Johns, 2023 for related work on this issue). This approach runs counter to current trends in distributional and embedding models, where the focus has been on building very large models of word meanings, resulting in the current class of Large Language Models (e.g., the GPT class of language models; Binz & Schulz, 2023). Although these models show very impressive performance across a range of tasks, they lack cognitive plausibility due to their large-scale training materials which vastly eclipse the amount of material any human could process (i.e., the entire internet). From a cognitive perspective, the original goal of distributional modeling as set forth by Landauer and Dumais (1997) was to understand how word meanings can be acquired by simple learning mechanisms trained with realistic levels of lexical experience. One key aspect of this line of research is to understand the individual differences that are inherent in the word meanings that different individuals have, due to their differential experience with language (Johns et al., 2019). The results found in this article provide a starting point for the theoretical development of individualized models of word meanings through the determination of the lexical landscape of word meanings across and within language users.

To examine the relativity of word meanings, two measures of semantic alignment were used. The first followed the methodology employed by Thompson et al. (2020), where the semantic consistency of the closest neighbors for the same word, trained from two different user corpora, was calculated. In this article, this measure was entitled indirect semantic alignment, as it relies upon second-order properties of a word's representation to assess the similarity of two words. The other measure, entitled direct semantic alignment, simply calculated the similarity between two word's representations that were trained on different corpora directly, which was then calculated across every user pair to determine a word's overall semantic alignment. This ability comes from the user corpora analyzed here using a common language, which was not possible for Thompson et al. (2020) given the cross-language goals of that study, and the stable training procedure offered by the distributional framework used here. It was found that the direct semantic alignment measure offered considerably higher correlations to a variety of psycholinguistic variables, such as word concreteness, suggesting that it is capturing meaning variability better than indirect semantic alignment. This is an unsurprising finding as the direct measure is capturing the totality of a word's representation, while the indirect measure is only capturing a limited subset of the space of a word's meaning.

Direct semantic alignment was found to offer a moderate correlation to concreteness (Brysbaert et al., 2014) and body−object interaction ratings (Pexman et al., 2019), suggesting that words that are more perceptually grounded are used more consistently across individuals. This finding differs from that of Thompson et al. (2020), who did not find a relationship between concreteness and semantic alignment in their across language analysis. This difference between results suggests that there may be differential impacts on the perceptual environment on within versus across language word meanings, where within a language perceptual information is used to coordinate word meanings compared to across languages. This could potentially be due to differences in the physical referents of words across cultures. However, there are considerable variations across these two studies that make any definitive conclusion difficult to make. For instance, there were significantly more word types examined here as compared to Thompson et al. (2020), due to the limitation in the number of words that can be compared across languages. This increased sample size enabled a more complete examination into word relativity and perceptual groundedness.

The positive correlation between direct semantic alignment and word concreteness and body−object interaction ratings is an important one as it suggests that words that are more perceptually grounded are more likely to have a common internal representation across individuals. This entails that the physical environment may be less variable than the linguistic environment, implying that the physical referents of words may have consistent levels of exposure across individuals. Developing semantic alignment measures from distributional models that integrate perceptual and grounded information into their representations, a current trend in distributional semantics in cognitive modeling and machine learning (e.g., Bruni, Tran, & Baroni, 2014; De Deyne, Navarro, Collell, & Perfors, 2021; Johns & Jones, 2012; Lazaridou, Marelli, & Baroni, 2017), could potentially provide a better ability to determine the contribution of the perceptual environment to word meaning relativity.

The other psycholinguistic variable that semantic alignment was correlated to is the semantic diversity measure of Hoffman, Ralph, & Rogers (2013). This metric measures the variability of the meaning of contexts that a word appears in. The finding that this variable is related to semantic alignment suggests that words that are more promiscuous in their context-to-context usage have less agreement across individual language users. This means that they have a greater degree of acceptability in terms of how these words can be used across different contexts and by different users. Direct semantic alignment also has a significant correlation to the contextual diversity variables from Johns (2021a), which is a count of the number of different contexts that a word occurs in normalized by the semantic diversity of those contexts, reinforcing this finding. This finding suggests that contextual variability in word usage impacts the semantic representation that people derive from those words and may lead to different people acquiring different semantic representations if they are biased toward experiencing those words within certain context types.

A unique advantage of using Reddit to drive distributional models is that there are a number of available metadata for each comment produced on the forum, which is what enabled the construction of individual user corpora in the first place. An additional type of data that can be extracted for each comment is the discourse (subreddits) that the comment was produced in. This information enabled an examination into the impact of user similarity. User similarity was computed by constructing a vector of the number of times each user commented across different subreddits and taking the similarity between each user's corresponding vector. There was a strong connection found between user similarity and the average semantic alignment of users, suggesting that users who tend to communicate in the same discourses also have more aligned word meanings. The impact of user similarity on word relativity suggests social and communicative pressures on word meanings: as people need to communicate with the same groups of individuals, the usage of words that they employ becomes more similar. That is, words are communicative tools (see Nagy & Townsend, 2012) that have specific uses across different contexts.

To examine within user semantic alignment, each user corpus was split in half temporally, and then the direct semantic alignment of words between the two halves was calculated. It was found that the within user semantic alignment measures were significantly greater than the across individual alignment values, signaling that individual language users are much more aligned with themselves than others, a comforting result and one that has been previously detailed at a different level of analysis (Johns & Jamieson, 2018; Johns et al., 2020). However, word meanings were still not completely aligned, as there was significant variability across words, suggesting that there is still fluidity in word meanings within an individual. Within user semantic alignment was found to be more connected to the above outlined psycholinguistic variables than across user alignment, suggesting that these variables exert more influence on the consistency of an individual's usage of language.

To examine within user semantic alignment more closely, the discourses that users tend to communicate within were examined. Specifically, a subset of users who communicate in a single discourse a large portion of the time (users who communicated between 35% and 65% in a single discourse) were identified. Within discourse semantic alignment and across discourse semantic alignment were calculated from this subset of users. It was found that within discourse semantic alignment was significantly greater than across discourse semantic alignment, signaling that the users are changing their language patterns based upon the discourse that they are communicating within. This suggests that word meanings are dependent on who one is communicating with and within which discourse. There are several possible reasons for this pattern—for example, it may be related to topic expertise, where some of the users are experts in certain technical fields. It could also be an adaptation effect, where communication patterns become standardized based upon repeated interactions within the same context and with similar sets of people. Future research involving a more targeted examination of user comments may provide more evidence about the possible reasons for this finding.

The final set of simulations examined the impact of syntactic and semantic category on semantic alignment. As was found in Thompson et al. (2020), there was natural variation found across both of these categories. For syntactic categories, it was found that nouns (and modifiers of nouns) are more aligned than verbs (and modifiers of verbs), which is likely related to concreteness. However, distributional models are not typically tested on verbs (Qiu & Johns, 2020), and thus it may be the case that this model type is not as adept at constructing the meaning of these words compared to nouns, which may make this result difficult to interpret. For semantic categories, the semantic alignment of the categories from the Van Overschelde et al. (2004) was used. It was found that there was also natural variation in the alignment across different categories. Intriguingly, it was found that category member prototypicality had a strong correlation to semantic alignment, indicating that words that are more common members of a category are used more consistently across individuals. This suggests that the linguistic environment may be more stable for highly prototypical category exemplars, leading to more shared experience across people.

Overall, this study suggests that there are considerable amounts of relativity in word meanings across individuals within a single language. This fluidity is likely the result of the differential experience that different individuals have with language. We all experience different language statistics, leading to different representations being formed from that experience. Additionally, this experience is crouched within higher-order properties of linguistic context, including communicative and social aspects of language usage. That is, word meanings seem to be context-dependent, with the context being multidimensional. The end result of this differential experience with the linguistic world that different people have is that the word meanings derived from different user corpora contain significantly different representations. Importantly, these differences are partly explained by other factors of language, such as concreteness, semantic diversity, and social aspects of language usage, such as the discourses one communicates within. This demonstrates that the semantic alignment values derived here are grounded in well-studied psycholinguistic variables. An important goal for future research is to determine the various other components of the relativity of word meanings.

The work of Thompson et al. (2020) provided the methodological and theoretical inspiration for the research reported here. This study used a distributional model to determine the relativity of word meanings across 41 languages and found that there was relatively little alignment across the languages, with the exception of a few semantic categories. Additionally, they found that the amount of alignment of two languages was related to the cultural similarities of where those languages are used, suggesting an experiential basis to differences in the lexical semantic system of different peoples. The theoretical aims of Thompson et al. were to examinate universalists (which claims a common set of meanings should occur across cultures; Gleitman & Fisher, 2005; Li & Gleitman, 2002; Pinker, 1994) versus relative (which claims that experience drives meaning and thus should differ across cultures; Lupyan, 2016; Lupyan & Dale, 2016) perspectives on lexical semantics. The results of Thompson et al. pointed toward a relative view of word meanings. The results reported here validate and extend this work by demonstrating that word meaning relativity extends to within a single language, as it showed that different individuals can differ significantly in the word meanings that they have acquired.

The finding of significant levels of relativity in word meanings across individuals has wide-ranging implications (see Andrews, 2012 for an early discussion of these issues in relation to word recognition) for theories of memory and language. In terms of computational models of cognition, it suggests future directions for how to account for variability in memory and language performance. Standard practice in cognitive modeling (e.g., Lewandowsky & Farrell, 2010) is to account for changes in the behavior of participants under different experimental conditions by manipulating the parameters of the model (e.g., decision criterion, encoding probability, drift rate, etc…), with a variety of optimization procedures being developed to fit these parameters (Shiffrin, 2010; Shiffrin et al., 2008). The existence of significant differences in the semantic memory of individuals within a population of language users suggests that it is likely possible to account for variability in behavior by assuming variability in the representation of words that the models utilize to simulate behavior, a possibility previously explored by Johns et al. (2019). However, this requires for a model to have integrated representation and processing assumptions. A current trend in cognitive modeling is to use a distributional model as the representation for a processing model (e.g., Hills, Jones, & Todd, 2012; Johns, Jones, & Mewhort, 2012, 2019; Mewhort, Shabahang, & Franklin, 2018; Osth, Shabahang, Mewhort, & Heathcote, 2020; see Johns et al., 2023 for a recent review) which would allow for the variability of representations to be included in the model's assumptions.

For the development of future distributional models, the work here presents a challenge. Namely, it suggests that not only do individuals have different knowledge bases due to differential experience with language, but it is also possible that the same individual can shift the type of meaning that they use depending on communicative context (i.e., the representation that they utilize to comprehend language is context-dependent). However, more research needs to be done on the latter point, utilizing much finer-grained controls than was utilized here.

Most distributional models are centralized (or prototype) models, which means that they construct a singular representation of word meaning, learned across contexts (Jamieson, Avery, Johns, & Jones, 2018). A different model type, employed widely across cognitive psychology, are instance-based models (for a recent review, see Jamieson et al., 2022). Multiple instance-based distributional models have been proposed (e.g., Kwantes, 2005; Jamieson et al., 2018; Johns & Jones, 2015; Johns, Jamieson, Crump, Jones, & Mewhort, 2020; see Jones, 2019 for a thorough discussion of these approaches). This distributional model type stores each experience with language that the model has in a separate location in memory and uses retrieval operations on the fly to construct representations of word meaning. Importantly, the representations that the model produces can be sensitive to context, given that this type of information can be integrated into the retrieval operation used to retrieve traces from memory, for example, by integrating discourse-based information into the traces that are stored in memory. This would provide the model with flexibility in terms of the meaning that it retrieves from memory, resulting in different meanings being retrieved in response to different discourses. Other frameworks, such as neural embedding models (e.g., Mikolov et al., 2013), could provide a similar mechanism by having a separate input layer that signals the discourse one is currently processing. The best way to account for word relativity in distributional models is an important topic for future research.

The goal of this article was to determine the semantic alignment of words across individuals and determine the lexical factors that influence word meaning alignment. However, one aspect that was not explored here was alignment at different levels of analysis. For example, when examining within user alignment, it was found that users were more aligned when communicating within a single discourse as compared to across discourses. However, the impact of the content of the discourse was not included in examining the alignment values. One could analyze user-discourse alignment by computing word representations for both individuals (as was done here) and discourses, to determine how the semantic construction of a discourse modulates an individual's usage of language. Additionally, this type of analysis could be used to explore the dynamics of individual language change, as one could examine how a user adapts their language usage to the communicative requirements of individual discourses across time. This type of analysis is beyond the scope of the current article but is an important area of future research.

A related issue with the work presented here is that the comments that the analyzed users produced were self-selected in terms of the discourses that they communicated within. That is, each user has their own interests (e.g., sports, politics, computer programming, etc…) and so the lack of relativity across users may simply be due to the differences in the discourses that users choose to communicate in. This was found to be somewhat true given the results of Fig. 14, where it was found that as the similarity between two users' discourse communication pattern increased, so did their semantic alignment. However, it is worth noting that even at the highest level of user−user discourse similarity, the alignment values were still dwarfed by the average within user alignment values, which means that users with similar discourse communication patterns are not at the ceiling of possible alignment values. Additionally, the correlation between user semantic alignment and discourse similarity was only moderate. This suggests that the communicative discourses that users communicate within have a clear impact on how they use language, there are still considerable differences in the alignment of users even with highly similar preferences for where those users choose to communicate.

Given the clearly exploratory and uncontrolled nature of this research, it is difficult to get a very definite answer on this question, which is quite fundamental—that is, does communicative discourse or individual perspective/experience dictate the malleability of language? The underlying theoretical position that this article takes is more influenced by the latter notion (see Johns et al., 2023), but multiple simulations in this article point to the general importance of discourse in the differences in word meanings across users. This argument reflects similar work in lexical organization that has compared individual user and discourse lexical statistics in accounting for word recognition data, where it has been found that user-based information seems to provide a better accounting than discourse information in general analyses (Johns, 2021a) but that finding is sensitive to task and participant population (Johns, 2022; Johns, Taler, & Jones, 2022). Tackling this question in a more controlled fashion is an important topic for future research; however, the utilization of large-scale communication patterns along with metadata that provides information about communicative context provides a promising starting point for this kind of research.

A methodological concern about the finding of meaning relativity across individuals is related to individual differences and replication issues in empirical data collection, currently a major focus in the psychological sciences (e.g., Shrout & Rodgers, 2018). If there are significant levels of variability in word meanings across individuals, and an experiment's design is reliant on the selection of specific words as stimuli, it follows that relativity could impact a study's ability to replicate and/or find consistent results.

A good example of the impact of individual differences on lexical semantic processing is given by the Semantic Priming Project (Hutchison et al., 2013), which was a mega study designed to examine semantic priming across a large number of participants collected at a variety of laboratories across the United States. This study found a variety of individual differences and reliability issues across stimuli (Yap, Hutchison, & Tan, 2016), replicating previous findings about variability in semantic priming (Stolz, Besner, & Carr, 2005), and established across a reanalysis of a number of different semantic priming experiments (Heyman, Bruninx, Hutchison, & Storms, 2018). Given that semantic priming effects rely upon the association strength of words (Hutchison, 2003), and association is likely to be at least partly determined through environmental occurrence patterns of words, word relativity is likely related to issues of reliability and replication in this literature. Indeed, Aujla (2021) recently demonstrated that individual experience with language predicts levels of semantic priming.

To gain a more concrete understanding of the impact of individual experience on replication in lexical semantics, consider mediated semantic priming. In mediated priming, the association between the target and prime (e.g., *turtle-slow*) is through a mediated word (e.g., *fast*). Mediated priming is often a subtle finding, with relatively low effect sizes leading to reliable effects only being found under certain experimental conditions (Balota & Lorch, 1986; McNamara & Altarriba, 1988). Many process models have been proposed of this finding (e.g., Balota & Lorch, 1986; Jones, 2010; McKoon & Ratcliff, 1992), and it has also been explained with distributional models (Jones, Kintsch, & Mewhort, 2006), but here it will be used as an example of a finding that is relatively delicate in order to see what impact individual variability in lexical semantics could be playing in explaining such empirical results.

In Balota and Lorch (1986), a set of 48 target words with related (strong associates) and mediated primes was assembled. These pairs will be used to run a Monte Carlo analysis to determine the impact of individual variability on finding a significant effect in a mediated priming experiment. Only the mediated primes will be analyzed here, as the related primes were found to be universally connected to the target across the user corpora, so there is a lack of interesting behavior across the user models to analyze for these primes. Due to the relatively limited scale of the cosine values from the model used here, the priming effect was calculated as a percent increase over the cosine of unrelated words, with unrelated words here being a randomly selected mediated prime of a different target word. Across the 500 user corpora, an average cosine increase of 28% from the mediated primes over the unrelated primes was found, a significant increase [$t(499) = 8.693$, $p < .001$].

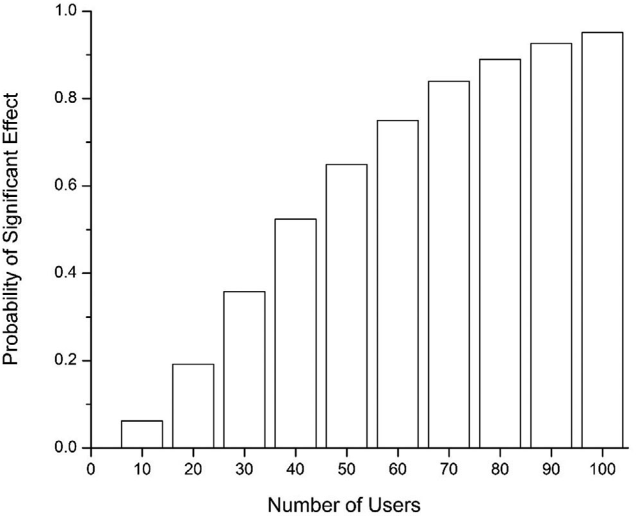

Fig. 21. Results of a Monte Carlo simulation predicting whether a mediated priming effect was found depending on how many user corpora were included in calculating the amount of priming.

However, the key question is what happens to model performance when only a sample of the users were used to calculate average priming. To accomplish this, groups of 10−100 users, in steps of 10, were selected randomly. Then, the average priming level across the group was calculated and tested for significance with a one-sample *t*-test to determine if the group achieved a significant level of semantic priming (i.e., a significant increase over 0% priming). A relatively conservative alpha of .01 was used to test significance to ensure that the effect was strong. For each group size, 10,000 random samples were conducted and the average probability of finding a significant level of priming was calculated. The results are displayed in Fig. 21, which shows that at a small number of samples, the probability of finding a significant priming effect was relatively low (e.g., only 36% of the time when a sample of 30 users was used). This probability increased considerably with larger samples (e.g., 95% of the time with a sample of 100), but finding an effect was still not ensured.

The simulation contained in Fig. 21 demonstrates the potential impact of word relatively on empirical data collection in psycholinguistics, as it suggests that word-level variability in meaning across individuals could impact finding a significant result just due to the stimuli that were selected, not necessarily because of the experimental design or underlying theoretical construct. The semantic alignment values reported in this article could aid in the construction of more stable stimuli, through the selection of words that are high in semantic alignment across individuals. They could also serve as a potential tool to integrate stimuli selection in determining the statistical power of an experiment, by providing a measure that predicts how much variability should be seen for the selected stimuli in a given experiment.

The results of this article, as well as other recent studies utilizing social media to examine different aspects of cognition (e.g., Herdağdelen & Marelli, 2017; Johns, 2019, 2021a, 2021b; Otto, Devine, Schulz, Bornstein, & Louie, 2022; Otto & Eichstaedt, 2018), suggest that social media resources provide a testbed to evaluate the impact of different aspects of the language environment on cognitive processes. The impact of the widespread use of social media has been popular in other areas of psychology, such as clinical and social psychology (see Appel, Marker, & Gnambs, 2020 for a review), but also provides unique opportunities for the cognitive and language sciences. One unique aspect of these materials is that they allow for metadata, such as who produced a piece of language and in what discourse it was produced in, to be extracted. This provides information about the surrounding social and communicative environment that language was produced in, which in turn allows for unique analyses on the impact of the social environment on language and memory processes to be tested (see Johns, 2021a, 2022; Johns & Jones, 2022). Additionally, given the increasing importance of the online world to everyday life, it will also be important to understand how the pattern of linguistic information on social media differs from other, more traditional, information sources.

Even though using social media as a text source for cognitive modeling has some positives, there are some issues with the materials used here. All corpora have their drawbacks—for example, it is unclear if Wikipedia articles accurately account for cultural knowledge, or if movie and television subtitles accurately described how individuals communicate, even though they have been used as such across multiple studies. Using Reddit as a language source has similar issues. One major drawback is that Reddit users are likely not representative of the general population, as the average Reddit user tends to be young, male, white, American, and more educated than the population at large (Barthel, Stocking, Holcomb, & Mitchell, 2016). An additional issue is that discussions on an internet forum are less naturalistic than communications that take place during everyday life, as the discussions that take place on the site are prompted by the discourse one is communicating in.

However, there are corresponding advantages to using Reddit as a datasource in terms of ecological validity (at least compared to other linguistic corpora), as many of the topics discussed on the site map onto an individual's everyday experiences, ranging from occupationally organized discussions (e.g., r/Plumbing, r/ComputerEngineering, r/InvestmentBanking/, etc…), entertainment and sports (e.g., r/Hockey, r/Television, r/GameBoy, etc…), or household activities (e.g., r/Cleaning, r/InteriorDesign, r/Cooking, etc…). Additionally, given the massive number of comments that each individual studied here has produced on the site, it can be assumed that a large part of their everyday life involves communicating on Reddit. Finally, at the moment there are few alternatives to obtaining massive samples of a single individual's language usage that can be used to examine word relativity.

The determination of large-scale relativity in word meanings across language users demonstrates, along with the results of Thompson et al. (2020), the usefulness of distributional models (and experiential accounts of cognition in general) in examining historically important problems in the cognitive and language sciences. Individual experience is inherently variable, and this variability has necessary consequences on the meanings that people derive from the environment that they occupy. Understanding the nature of this experiential variability, and its impact on human behavior, should be a central focus for computational accounts of lexical semantics in the future.

Acknowledgments

This research was supported by the Natural Science and Engineering Research Council of Canada (NSERC) Discovery Grant RGPIN-2020-04727.

Open practices statement

All resulting metrics used in this article are available at https://osf.io/rsjkc.

Notes

1 The API documentation can be found at: https://www.reddit.com/dev/api/; however, this API has recently been closed to the public.
2 The diagonal half of the matrix calculated with $\frac{n^2-n}{2}$, where n is the number of columns in the matrix.
3 Thompson, Roberts, and Lupyan (2020) analyzed 1010 words, which were words that had translations across the various languages that were available. Given that this article is only analyzing English, there is no necessary limitation on the number of words, outside of their availability in the user corpora, that could be examined.
4 One might expect semantic alignment and concreteness to be correlated as words that have more concrete referents should not differ as much across people and cultures compared to more abstract words, given a relatively consistent perceptual world.

References

Andrews, S. (2012). Individual differences in skilled visual word recognition and reading: The role of lexical quality. In J. S. Adelman (Ed.), *Visual word recognition volume 2* (pp. 163–184). Psychology Press.

Appel, M., Marker, C., & Gnambs, T. (2020). Are social media ruining our lives? A review of meta-analytic evidence. *Review of General Psychology*, *24*, 60–74.

Aujla, H. (2021). Language experience predicts semantic priming of lexical decision. *Canadian Journal of Experimental Psychology*, *75*, 235–244.

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L., Simpson, G. B., Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459.

Balota, D. A., & Lorch, R. F. (1986). Depth of automatic spreading activation: Mediated priming effects in pronunciation but not in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 336–345.

Barthel, M., Stocking, G., Holcomb, J., & Mitchell, A. (2016). Reddit news users more likely to be male, young and digital in their news preferences. Pew Research Center.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. *14*, pp. 830–839).

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., Holland, J., Ke, J., Larsen-Freeman, D., & Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning*, *59*, 1–26.

Bhatia, S., & Aka, A. (2022). Cognitive modeling with representations from large-scale digital data. *Current Directions in Psychological Science*, *31*, 207–214.

Binz, M., & Schulz, E. (2023). Using cognitive psychology to understand GPT-3. *Proceedings of the National Academy of Sciences*, *120*, e2218523120.

Bruni, E., Boleda, G., Baroni, M., & Tran, N. K. (2012). Distributional semantics in technicolor. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1* (pp. 136–145).

Bruni, E., Tran, N. K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.

Brysbaert, M., Mandera, P., McCormick, S. F., & Keuleers, E. (2019). Word prevalence norms for 62,000 English lemmas. *Behavior Research Methods*, *51*, 467–479.

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

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). How many words do we know? Practical estimates of vocabulary size dependent on word definition, the degree of language input and the participant's age. *Frontiers in Psychology*, *7*, 1116.

Bullinaria, J. A., & Levy, J. P. (2007). Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior Research Methods*, *39*, 510–526.

Bullinaria, J. A., & Levy, J. P. (2012). Extracting semantic representations from word co-occurrence statistics: Stop-lists, stemming, and SVD. *Behavior Research Methods*, *44*, 890–907.

Chang, M., Jones, M. N., & Johns, B. T. (2023). Comparing word frequency, semantic diversity, and semantic distinctiveness in lexical organization. *Journal of Experimental Psychology: General*, *152*, 1814–1823.

Christiansen, M. H., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences*, *31*, 489–509.

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*, 159–190.

De Deyne, S., Navarro, D. J., Collell, G., & Perfors, A. (2021). Visual and affective multimodal models of word meaning in language and mind. *Cognitive Science*, *45*, e12922.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological Review*, *62*, 369.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*, 429–448.

Frimer, J. A., Aujla, H., Feinberg, M., Skitka, L. J., Aquino, K., Eichstaedt, J. C., & Willer, R. (2022). Incivility is rising among American politicians on Twitter. *Social Psychological and Personality Science*, 14, 259-269.

Gleitman, L., & Fisher, C. (2005). Universal aspects of word learning. In J. McGilvray (Ed.), *The Cambridge companion to Chomsky* (pp. 123–144). New York: Cambridge University Press.

Goldberg, Y., & Levy, O. (2014). word2vec Explained: Deriving Mikolov et al.'s negative-sampling wordembedding method. arXiv preprint arXiv:1402.3722.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*, 787–814.

Griffiths, T. L., Steyvers, M., & Tenenbaum, J. B. (2007). Topics in semantic representation. *Psychological Review*, *114*, 211–244.

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*, 1006–1033.

Herdağdelen, A., & Marelli, M. (2017). Social media and language processing: How Facebook and Twitter provide the best frequency estimates for studying word recognition. *Cognitive Science*, *41*, 976–995.

Heyman, T., Bruninx, A., Hutchison, K. A., & Storms, G. (2018). The (un) reliability of item-level semantic priming effects. *Behavior Research Methods*, *56*, 2173–2183.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*, 431–438.

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*, 718–773.

Hutchison, K. A. (2003). Is semantic priming due to association strength or feature overlap? A microanalytic review. *Psychonomic Bulletin & Review*, *10*, 785–813.

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

Jamieson, R. K., Avery, J. E., Johns, B. T., & Jones, M. N. (2018). An instance theory of semantic memory. *Computational Brain & Behavior*, *1*, 119–136.

Jamieson, R. K., Johns, B. T., Vokey, J. R., & Jones, M. N. (2022). Instance theory as a domain-general framework for cognitive psychology. *Nature Reviews Psychology*, *1*, 174–183.

Johns, B. T., & Jones, M. N. (2010). Evaluating the random representation assumption of lexical semantics in cognitive models. *Psychonomic Bulletin & Review*, *17*, 662–672.

Johns, B. T., & Jones, M. N. (2012). Perceptual inference through global lexical similarity. *Topics in Cognitive Science*, *4*, 103–412.

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2012). A synchronization account of false recognition. *Cognitive Psychology*, *65*, 486–518.

Johns, B. T., & Jones, M. N. (2015). Generating structure from experience: A retrieval-based model of language processing. *Canadian Journal of Experimental Psychology*, *69*, 233–251.

Johns, B. T., & Jamieson, R. K. (2018). A large-scale analysis of variance in written language. *Cognitive Science*, *42*, 1360–1374.

Johns, B. T., & Jamieson, R. K. (2019). The influence of time and place on lexical behavior: A distributional analysis. *Behavior Research Methods*, *51*, 2438–2453.

Johns, B. T., Mewhort, D. J. K., & Jones, M. N. (2019). The role of negative information in distributional semantic learning. *Cognitive Science*, *43*, e1273.

Johns, B. T., Jones, M. N., & Mewhort, D. J. K. (2019). Using experiential optimization to build lexical representations. *Psychonomic Bulletin & Review*, *26*, 103–126.

Johns, B. T., Dye, M., & Jones, M. N. (2020). Estimating the prevalence and diversity of words in written language. *Quarterly Journal of Experimental Psychology*, *73*, 841–855.

Johns, B. T., Jamieson, R. K., & Jones, M. N. (2020). The continued importance of theory: Lessons from big data approaches to cognition. In S. E. Woo, R. Proctor, & L. Tay (Eds.), *Big data methods for psychological research: New horizons and challenges* (pp. 277–295). APA Books.

Johns, B. T., Jamieson, R. K., Crump, M. J. C., Jones, M. N., & Mewhort, D. J. K. (2020). Production without rules: Using an instance memory model to exploit structure in natural language. *Journal of Memory and Language*, *115*, 104165.

Johns, B. T. (2019). Mining a crowdsourced dictionary to understand consistency and preference in word meanings. *Frontiers in Psychology*, *10*, 268.

Johns, B. T. (2021a). Disentangling contextual diversity: Communicative need as a lexical organizer. *Psychological Review*, *128*, 525–557.

Johns, B. T. (2021b). Distributional social semantics: Inferring word meanings from communication patterns. *Cognitive Psychology*, *131*, 10144.

Johns, B. T., & Jones, M. N. (2022). Content matters: Measures of contextual diversity must consider semantic content. *Journal of Memory and Language*, *123*, 104313.

Johns, B. T. (2022). Accounting for item-level variance in recognition memory: Comparing word frequency and contextual diversity. *Memory & Cognition*, *50*, 1013–1032.

Johns, B. T. (2023). Computing word meanings by aggregating individualized distributional models: Wisdom of the crowds in lexical semantic memory. *Cognitive Systems Research*, *80*, 90–102.

Johns, B. T., Jamieson, R. K., & Jones, M. N. (2023). Scalable cognitive modelling: Putting Simon's (1969) ant back on the beach. *Canadian Journal of Experimental Psychology*, *77*, 185–201.

Jones, L. L. (2010). Pure mediated priming: A retrospective semantic matching model. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 135.

Jones, M. N. (2019). When does abstraction occur in semantic memory: Insights from distributional models. *Language, Cognition and Neuroscience*, *34*, 1338–1346.

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

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*, 1–37.

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*, 287–304.

Kumar, A. A. (2020). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*, 40–80.

Kramsch, C. (2014). Language and culture. *AILA Review*, *27*, 30–55.

Kwantes, P. J. (2005). Using context to build semantics. *Psychonomic Bulletin & Review*, *12*, 703–771.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–224.

Lazaridou, A., Marelli, M., & Baroni, M. (2017). Multimodal word meaning induction from minimal exposure to natural text. *Cognitive Science*, *41*, 677–705.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *Advances in neural information processing systems* (pp. 2177–2185).

Levy, O., Goldberg, Y., & Dagan, I. (2015). Improving distributional similarity with lessons learned from word embedding. *Transactions of the Association for Computational Linguistics*, *3*, 211–225.

Lewandowsky, S., & Farrell, S. (2010). *Computational modeling in cognition: Principles and practice*. SAGE Publications.

Li, P., & Gleitman, L. (2002). Turning the tables: Language and spatial reasoning. *Cognition*, *83*, 265–294.

Lupyan, G. (2016). The centrality of language in human cognition. *Language Learning*, *66*, 516–553.

Lupyan, G., & Dale, R. (2016). Why are there different languages? The role of adaptation in linguistic diversity. *Trends in Cognitive Sciences*, *20*, 649–660.

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.

Martí, L., Wu, S., Piantadosi, S. T., & Kidd, C. (2023). Latent diversity in human concepts. *Open Mind*, *7*, 79–92.

McNamara, T. P., & Altarriba, J. (1988). Depth of spreading activation revisited: Semantic mediated priming occurs in lexical decisions. *Journal of Memory and Language*, *27*, 545–559.

McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 1155–1172.

Mewhort, D. J. K., Shabahang, K. D., & Franklin, D. R. J. (2018). Release from PI: An analysis and a model. *Psychonomic Bulletin & Review*, *25*, 932–995.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111–3119).

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*, 91–108.

Osth, A. F., Shabahang, K. D., Mewhort, D. J., & Heathcote, A. (2020). Global semantic similarity effects in recognition memory: Insights from BEAGLE representations and the diffusion decision model. *Journal of Memory and Language*, *111*, 104071.

Otto, A. R., Devine, S., Schulz, E., Bornstein, A. M., & Louie, K. (2022). Context-dependent choice and evaluation in real-world consumer behavior. *Scientific Reports*, *12*, 17744.

Otto, A. R., & Eichstaedt, J. C. (2018). Real-world unexpected outcomes predict city-level mood states and risk-taking behavior. *PLoS One*, *13*, e0206923.

Pexman, P. M., Muraki, E., Sidhu, D. M., Siakaluk, P. D., & Yap, M. J. (2019). Quantifying sensorimotor experience: Body–object interaction ratings for more than 9,000 English words. *Behavior Research Methods*, *51*, 453–466.

Pinker, S. (1994). *The language instinct*. New York: William Morrow & Co.

Qiu, M. & Johns, B. T. (2020). Semantic diversity in paired-associate learning: Further evidence for the information accumulation perspective of cognitive aging. *Psychonomic Bulletin & Review*, *27*, 114–121.

Recchia, G. L., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information to latent semantic analysis. *Behavior Research Methods*, *41*, 657–663.

Senaldi, M. S. G., Titone, T., & Johns, B. T. (2022). Determining the importance of frequency and contextual diversity in the lexical organization of multiword expressions. *Canadian Journal of Experimental Psychology*, *76*, 87–98.

Shaoul, C., & Westbury, C. (2010). *The Westbury Lab Wikipedia Corpus 201*. Edmonton, AB: University of Alberta.

Shiffrin, R. M. (2010). Perspectives on modeling in cognitive science. *Topics in Cognitive Science*, *2*, 736–750.

Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*, 1248–1284.

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, *69*, 487–510.

Simon, H. A. (1969). *The sciences of the artificial*. Boston, MA: MIT Press.

Stolz, J. A., Besner, D., & Carr, T. H. (2005). Implications of measures of reliability for theories of priming: Activity in semantic memory is inherently noisy and uncoordinated. *Visual Cognition*, *12*, 284–336.

Taler, V., Johns, B. T., & Jones, M. N. (2020). A large scale semantic analysis of verbal fluency across the aging spectrum: Data from the Canadian longitudinal study on aging. *Journal of Gerontology: Psychological Sciences*, *75*, e221–e230.

Thompson, B., Roberts, S. G., & Lupyan, G. (2020). Cultural influences on word meanings revealed through large-scale semantic alignment. *Nature Human Behaviour*, *4*, 1029–1038.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Tiv, M., Kutlu, E., Gullifer, J. G., Feng, R. Y., Doucerain, M. M., & Titone, D. (2022). Bridging interpersonal and ecological dynamics of cognition through a systems framework of bilingualism. *Journal of Experimental Psychology: General*, *151*, 2128–2143.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the norms. *Journal of Memory and Language*, *50*, 289–335.

Yap, M. J., Hutchison, K. A., & Tan, L. C. (2016). Individual differences in semantic priming performance: Insights from the Semantic Priming Project. In *Big data in cognitive science* (pp. 212–235). Psychology Press.

Wang, X., & Bi, Y. (2021). Idiosyncratic tower of Babel: Individual differences in word-meaning representation increase as word abstractness increases. *Psychological Science*, *32*, 1617–1635.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers*, *20*, 6–10.