Structure and process-level lexical interactions in memory search: A case study of individuals with cochlear implants and normal hearing

Abhilasha A. Kumar[1], Mingi Kang[1], William G. Kronenberger[2], Michael N. Jones[2], David B. Pisoni[2]

[1]Bowdoin College, ME
[2]Indiana University, Bloomington

Abstract

Searching through memory is mediated by complex interactions between the underlying mental lexicon and the processes that operate on this lexicon. However, these interactions are difficult to study due to the effortless manner in which neurotypical individuals perform cognitive tasks. In this work, we examine these interactions within a sample of prelingually deaf individuals with cochlear implants who were administered the verbal fluency task. Specifically, we test how two potential candidates for mental lexicons constructed from textual (word2vec) versus speech-based (speech2vec) information, when combined with different process models of memory search account for search behavior within the verbal fluency task for CIs and normal hearing individuals. Our findings show that semantic and phonological information jointly influences search behavior at the structural and process level and highlight the delicate balance of different lexical sources that produces successful search outcomes.

Keywords: memory search; verbal fluency; computational modeling; word embeddings

Introduction

When individuals search their mental lexicon for items, they are bringing online several processes and knowledge structures. For example, if asked to produce as many *animals* as possible within a fixed duration, i.e., the verbal fluency task (VFT; Bousfield & Sedgewick, 1944), individuals first have to focus on a specific subset of their knowledge (i.e., *animals*). They also have to ignore other words that may be related or come to mind (e.g., where animals live, sounds animal make, characteristics of animals, etc.), but are not relevant to the task at hand. Then, they need to employ some type of search strategy to navigate this subspace of *animals*, which may involve attending to different lexical sources or characteristics to organize the search in a meaningful and efficient manner. Taken together, mental search is a result of interactions between structure-level representations and processes.

Several researchers have attempted to understand these structure-process interactions that occur during memory search from a computational perspective. This work has revealed that individuals tend to strategically retrieve items in "clusters" until they have exhausted the local neighborhood, at which point they "switch" to a new cluster, consistent with mechanisms of external search that rely on optimal foraging (Hills et al., 2012; Hills, Todd, & Jones, 2015; Zemla et al., 2023). However, in much of this work, the underlying semantic representations that contribute to the critical question of how clusters and switches are defined are either based on pre-existing norms (Troyer, 2000; Zemla et al., 2020) or derived from distributional semantic models (DSMs) that emphasize meaning-based relationships between words in a high-dimensional space (Hills et al., 2012). However, semantic information is part of an integrated mental lexicon, that also contains other lexical sources of information. Indeed, acquiring a word and its meaning likely involves complex interactions between semantics, phonology, orthography, and acoustic and speech-related information. Although the relationship between meaning and wordform (i.e., phonology/orthography) is typically thought to be arbitrary, there is some recent evidence to suggest that these cues may actually be correlated in natural language (Dautriche et al., 2017), largely due to the functional pressures that are associated with novel language acquisition, as well as semantically related words sharing common etiology (e.g., *conform*, *formulate*, *reform*). Additionally, speech patterns may be particularly critical in forming early mappings between form and meaning (Saffran et al., 1996; Hay et al., 2011). Overall, it is important to investigate how representations that are derived from *multiple* lexical sources contribute to meaning formation, and how this knowledge is used in downstream tasks such as memory search. In particular, comparisons between models of word representation that combine more than one lexical source may be particularly useful in delineating the contribution of lexical sources to semantic organization.

From a process-level perspective, accounts have mainly focused on how local (semantic) and global (word frequency) information may influence within- and between-cluster transitions (Hills et al., 2012; Hills, Todd, Lazer, et al., 2015). However, task-discrepant clustering has commonly been observed in VFT (Abwender et al., 2001). Recently, Kumar et al. (2022) provided some preliminary evidence that phonological similarity may be important in mediating local within-cluster transitions (e.g., *mouse* to *mole*), and individuals who produced the fluency lists considered phonologically similar items as part of the same cluster (Lundin et al., 2023). However, the representations were all derived from text-based models that emphasize semantic information, and it was therefore unclear whether phonology was important at the representational or process level during search.

To isolate the contribution of underlying representations (*structures*) from *processes* that act upon these representa-

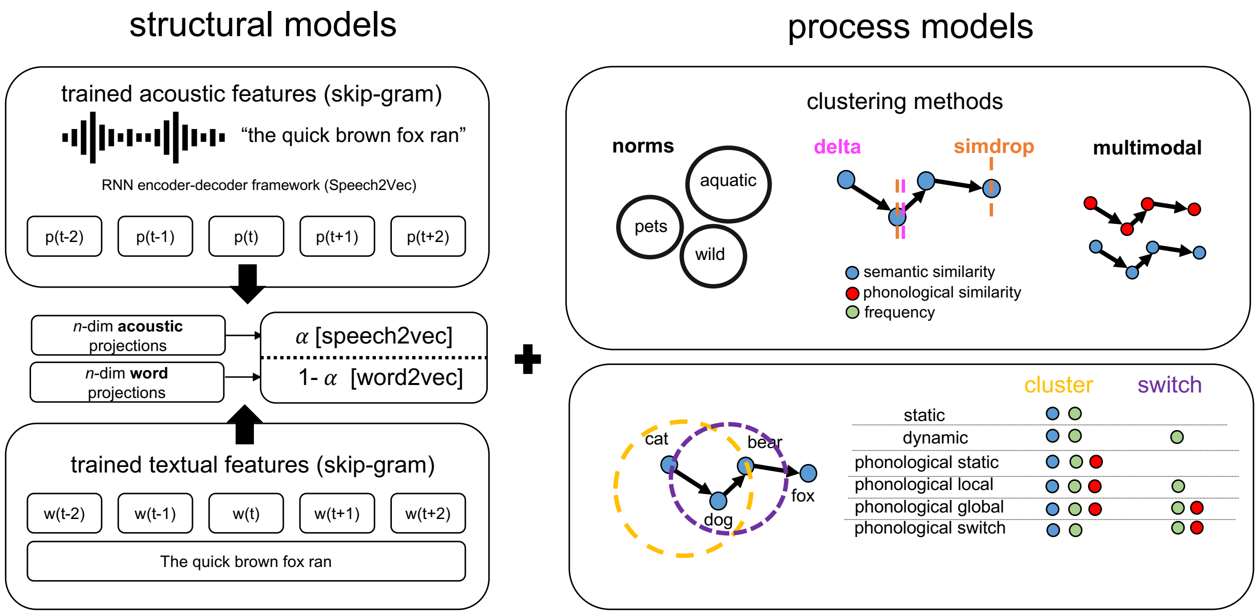

Figure 1: Modeling overview. Representations derived from two different language models trained on textual (*word2vec*) and acoustic (*speech2vec*) features were concatenated to produce multimodal vector representations for all words. These representations were then used in conjunction with process-level models of search that incorporated different clustering and switching methods as well as lexical sources to obtain likelihoods.

tions in the fairest way possible, we need a dataset where one of these variables can be "turned off" in principle. For example, to understand how speech and phonology impact the development of meaning-based representations, we would want a sample that does not have access to this information during early word learning. Although this manipulation is impossible to achieve in principle, many deaf children do indeed grow up in an environment with virtually no acoustic input, and experience difficulties achieving conventional language development milestones associated with vocabulary development and semantic organization in long-term memory (Coppens et al., 2013; Ormel et al., 2010). Even among deaf children who receive cochlear implants (CIs), language and other cognitive delays persist (Niparko et al., 2010; Bouchard et al., 2009). Whether these deficits arise due to compromised language *processing* or simply a result of impoverished *representations* remains unclear.

On one hand, it is possible that there are differences in how CIs organize the lexicon and this difference manifests in the nature of responses provided by CIs in language tasks. If semantic and phonological information are indeed correlated in natural language, one would expect that individuals who lack early phonological input have difficulties in acquiring semantic information, which in turn may influence their overall semantic organization and downstream retrieval processes. Kenett et al. (2013) examined fluency responses from a cohort of CIs using network analysis, and showed that responses from CIs were significantly more clustered and less well-integrated, compared to healthy controls, suggesting critical differences in semantic memory organization. However, Kenett et al. (2013) did not explore a computational model of memory search, so several interpretations are possible. One possibility is that the semantic representations among CIs are relatively comparable to neurotypical individuals, but the *processes* that enable retrieval of these representations are impaired. As such, Ormel et al. (2010) compared deaf and hearing children in their ability to perform the word–picture verification and found that hearing children automatically activated phonology during the task, regardless of whether it was relevant or not, whereas deaf children failed to do so in either case. Whether this tendency to not activate phonology in tasks where it may or may not be relevant extends to CIs is unknown. It is possible that access to speech may make phonology more salient for these individuals, in which case they may be *more* sensitive to it in language tasks. Alternatively, it may be the case that the lack of early phonological input may make this cue less salient during language processing. Overall, a deeper investigation of how search may differ across CIs and normal hearing individuals can yield novel insights about how these lexical sources interact during memory search. Within this context, atypical populations lacking early phonological input afford a unique opportunity to examine how semantic and phonological interactions during retrieval and the conditions under which these interactions may break down and lead to poorer retrieval performance.

In this work, we explored these questions within a sample of prelingually deaf individuals with cochlear implants (CIs) and neurotypical controls (NH) who completed the VFT. We

Table 1: Summary of demographic characteristics for cochlear implant (CI) and normal hearing (NH) groups in our sample.

Variables	**CI**	**NH**
Chronological age (years)	15.74 (9.86-26.66)	16.18 (10.2-27.07)
Age at implantation (months)	37.94 (11.07-75.76)	-
Duration of CI use (years)	12.58 (7.79-21.19)	-
Age of onset of deafness (months)	2.41 (0-24)	-
Income level	5 (1-8)	5.67 (1-10)
Standardized PPVT	84.69 (42-123)	108.63 (79-132)

implemented a series of *structural* models of semantic organization that were derived from purely textual, purely acoustic, or a combination of textual and acoustic information. We then combined the representations derived from these models with different process-level models of clustering and switching behavior in the VFT. We explored which type of structure + process model best fit the pattern of responses produced by participants in our groups (CI and NH).

Methods

Participants Participants were 30 prelingually deaf, early-implanted (less than 4 years), long-term (over 7 years) child and adolescent users of CIs, and compared with 30 age and nonverbal IQ-matched normal-hearing (NH) peers. Table 1 displays the demographic characteristics of both samples.

Verbal fluency The DKEFS (Delis et al., 2001) was administered. We only consider the "animals" category, given that this is the most rigorously investigated category in the VFT literature and also has the most extensive categorization norms. Individuals with cochlear implants (CIs) produced significantly fewer items than individuals with normal hearing (NH) in the current sample, $F(1,58) = 5.262$, $p = .024$.

Structure-level models We considered two underlying distributional semantic models as potential candidates for the lexicon, made available open-source by Chung & Glass (2018). *word2vec* is a standard next-word prediction-based language model that is trained on a large text corpus. *speech2vec* is a variant of *word2vec* that is trained on acoustic information in speech instead of text, where acoustic features are first processed through a recurrent neural network (RNN) encoder-decoder framework, which is subsequently integrated with a skip-gram neural network to predict upcoming acoustic sequences. Following previous work on multimodal DSMs (Kiela & Bottou, 2014), we concatenated the two representations (acoustics-based *speech2vec* and text-based *word2vec*, both trained on the same 500-hour LibriSpeech corpus) into a single vector (weighted by a tuning parameter α), which represented a jointly learned vector for each given word (see Figure 1):

$$v_{\text{word}} = \alpha \times v_{\text{speech2vec}} \parallel (1-\alpha) \times v_{\text{word2vec}}, \qquad (1)$$

We explored a wide range of parameter settings for α across four dimensions (50, 100, 200, 300), as well as baseline representations that only included one of the models and a model that averaged these representations.

Process-level models Our process models were based on prior work by Hills et al. (2012) and Kumar et al. (2022). Specifically, we examined different automatic clustering methods in conjunction with process models that differentiated which sources were used within and between clusters.

Four methods of defining clusters and switches were considered: norm-based, similarity drop, multimodal, and delta similarity. The norm-based method assigns clusters and switches based on pre-defined categories, which were obtained from SNAFU (Zemla et al., 2020). The similarity drop method assigns switches based on drops in semantic similarity, while the multimodal method assigns switches based on drops in a combined estimate of semantic and phonological similarity (weighted by a tuning parameter). Finally, the delta-similarity method assigns cluster and switch designations based on *relative* rises and drops in similarity, determined by *rise* and *fall* thresholds. All thresholds were parametrically varied to find the best-fitting models.

Three types of process models were considered in conjunction with the clustering methods. While the *static* model computes likelihoods of items produced based on a combination of semantic similarity and word frequency for all transitions (clusters and switches), the *dynamic* foraging model differentiates between clusters and switches and uses global word frequency for between-cluster transitions, and a combination of semantic similarity and word frequency for within-cluster transitions. We also evaluated a suite of foraging models that incorporated the role phonology in within and between-cluster transitions. Specifically, the *plocal* model uses frequency, semantic, and phonological similarity during within-cluster transitions and frequency during between-cluster transitions. The *pglobal* model uses frequency, semantic, and phonological similarity during within-cluster transitions, and frequency and phonological similarity during between-cluster transitions. Finally, the *pswitch* model uses only semantic similarity and frequency during within-cluster transitions and phonological similarity and frequency for between-cluster transitions (see Figure 1).

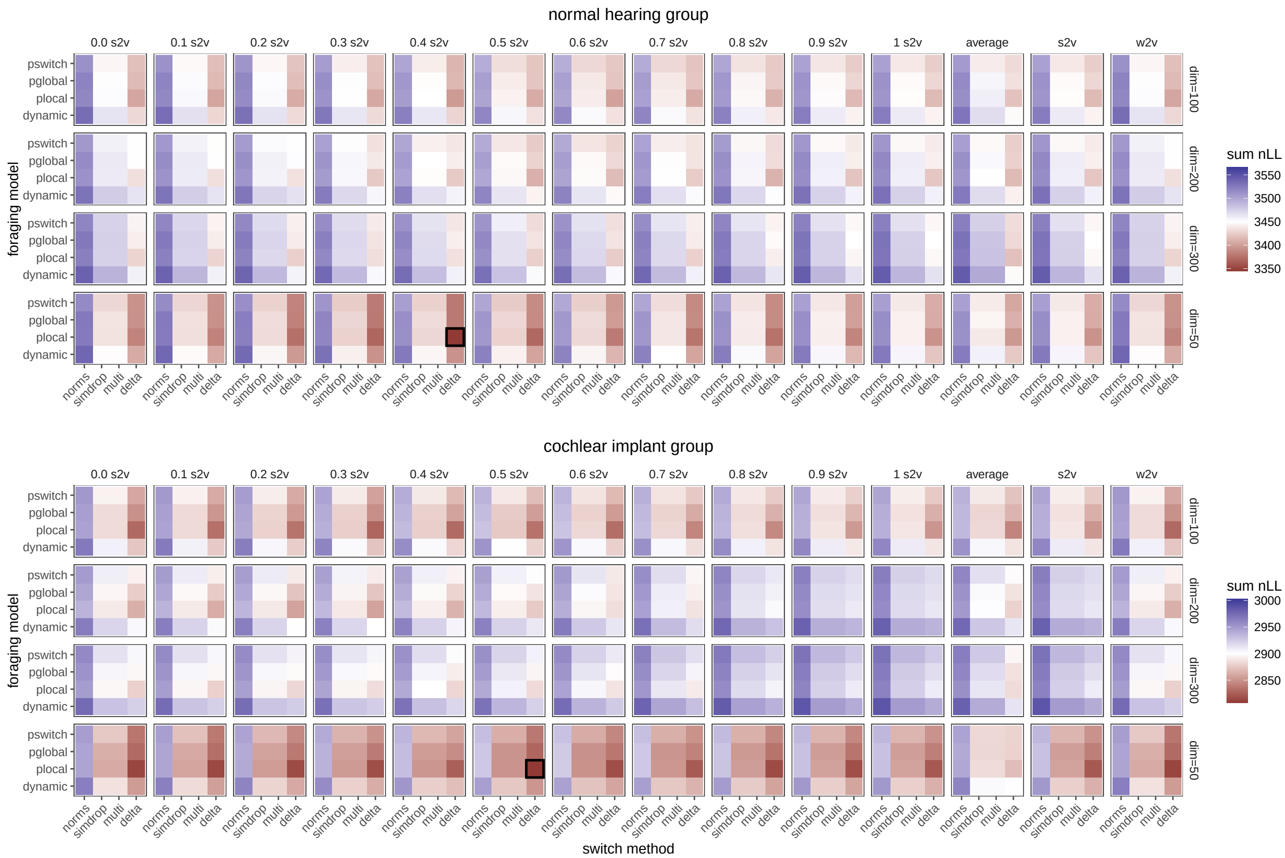

Figure 2: Negative log-likelihoods for models fit to verbal fluency data from the normal hearing and cochlear implant groups. Lower values indicate better fits. Squares outlined in black represent the best-fitting model in each group.

Creating embeddings We adapted *forager* (Kumar et al., 2023) to compute model likelihoods for different participants, using the representations derived from our structure-level models. A vocabulary of 463 animal words was constructed, using a combination of words present in the *word2vec* and *speech2vec* base models, animal words from SNAFU (Zemla et al., 2020) and words produced by participants in our sample. Suitable replacements were made for 50 words produced by participants that were not present in the base models and repetitions were excluded prior to any analyses.

Results

Normal hearing group First, we examined the extent to which different models were able to account for performance in our normal hearing sample. Figure 2 (top panel) displays the overall patterns across different dimensions and concatenated variants of the *structure*-level models, as well as the different foraging models and switch methods.

Overall, models with lower dimensions (i.e., 50) provided a better fit than models with higher dimensions. This is consistent with the patterns reported by Chung & Glass (2018), who found that the models with lower dimensions were best able to capture word similarity across a variety of benchmark tasks. Next, we found that the concatenated variants were able to perform better than the single-model or the "average" model, and the model that performed best overall assigned slightly lesser weight on embeddings derived from *speech2vec* than *word2vec* ($\alpha = 0.4$, see Equation 1).

The best-performing process model was the dynamic foraging model that incorporated semantic similarity, phonological similarity, and frequency in local transitions and frequency in global transitions (i.e., the *plocal* model in Figure 2), and used the delta similarity method to assign cluster-switch designations, replicating Kumar et al. (2022) and Lundin et al. (2023). This suggests that individuals with normal hearing treat local transitions within clusters differently than transitions between clusters. While within-cluster transitions emphasize *all* types of lexical sources, between-cluster transitions tend to emphasize word frequency.

Cochlear implant group Next, we examined the extent to which different models accounted for performance in our cochlear implant sample. As shown in Figure 2 (bottom panel), we again found that lower dimension models were better able to account for the behavioral patterns compared to higher dimension models. Interestingly, the *structure*-level model that best fit the CI data emphasized the *speech2vec* embeddings a little more than the *word2vec* embeddings ($\alpha =$

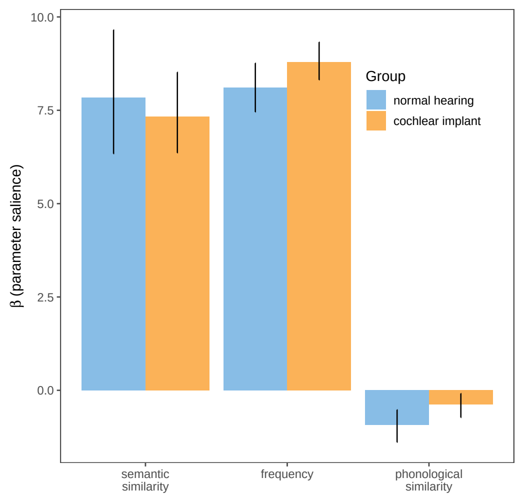

Figure 3: Parameter salience for lexical sources from the best-fitting models at the individual level. CIs showed greater reliance on frequency and phonological similarity. Error bars represent bootstrapped 95% confidence intervals.

0.5, see Equation 1), suggesting that CIs may be equally emphasizing the textual and speech-related features in their lexicon. At the process level, we found that the best-performing model for CIs was the same as their NH peers, i.e., the dynamic foraging model that incorporated all lexical sources in local transitions and frequency in global transitions (i.e., the *plocal* model).

Lexical sources We next investigated the relative usage of the three sources (word frequency, semantic similarity, and phonological similarity) across groups, by examining the best-fitting salience (β) parameters for each individual. As shown in Figure 3, CIs used phonological similarity (b= .55, t = 1.92, p = .06) and word frequency (b= .69, t = 1.60, p = .114) marginally more than NHs, while there were no differences in their use of semantic similarity (p = .62).

Fluency performance The finding that CIs were producing phonologically related words (or emphasizing phonology) more than their NH peers was surprising, as one might expect that the lack of exposure to sounds early in life may have had the opposite effect. One possibility is that these patterns may be confounded with the total number of items produced, i.e., since CIs produce fewer items overall, it is possible that phonologically similar items occur earlier in the lists and NHs continue to produce more dissimilar items over time, whereas CIs do not produce those remote items at all. To address this possibilty, we re-ran all analyses by truncating the CI/NH lists to the minimum number of items produced by either pair. The overall patterns were robust to list length, such that the best-fitting models for CIs were still ones that emphasized phonology more than NHs. Additionally, we examined whether the total number of items produced was related to the average consecutive phonological similarity across the two groups. As shown in Figure 4, producing more phonologically related items was related to producing more items overall among CIs, confirmed by a significant interaction, F(1,112)= 7.74; p = .006, and this pattern did not reliably vary by whether the analysis was conducted on truncated or complete lists (p = .60).

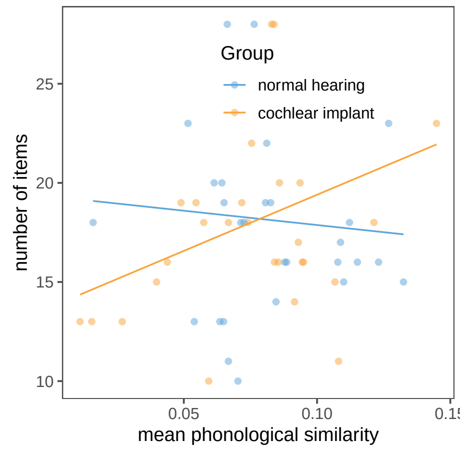

Figure 4: Total number of items produced as a function of mean consecutive phonological similarity, based on length-matched lists across both groups.

Discussion

In this work, we explored how different lexical sources interact at the *structural* and *process* level to produce search behavior within the verbal fluency task. While semantic and phonological information, in addition to word frequency have been previously implicated within the context of VFT (Abwender et al., 2001; Hills et al., 2012), separating the contribution of different lexical sources at the level of representation and processes has been difficult from a computational perspective. This issue is further compounded by the fact that neurotypical individuals are generally perform well in the VFT and it is therefore difficult to assess which of these sources are critical to successful performance in the task.

To better elucidate the potential underlying interactions between different lexical sources, we compared fluency lists from a sample of individuals with cochlear implants with individuals with normal hearing. Specifically, we explored how different concatenated variants of text and speech-based language models, when combined with a series of foraging models account for search behavior in the VFT. We found important similarities and differences at both the structure and process level between CIs and NHs. First, both groups were sensitive to representations derived from text and speech, suggesting that the lexicon is represented in a multimodal format across both groups. However, while CIs equally emphasized representations derived from speech and text, NHs deemphasized the speech-based representations in favor of the text-based representations. This may suggest that among neurotpyical individuals, speech-related cues may be overtaken by textual or linguistic cues over time, whereas CIs tend to rely on these cues a lot more than their peers. In line with this hypothesis, although both groups appeared to use phonology as a cue during search, CIs attended to phonology and word frequency marginally more than their NH peers, and this also improved their overall performance. It is possible that exposure to sound later in life heightens the salience of speech-related cues among CIs, and this cue is used in a compensatory manner during search tasks, when it is difficult to access semantically related information. Overall, these findings suggest that there may be differences in how concepts are organized (as suggested by Kenett et al., 2013) as well as how concepts are then retrieved in response to task constraints between CIs and NH peers.

Future work could further investigate these patterns in a larger sample and across diverse domains, to provide a more comprehensive picture of how different lexical sources are at play during searching through memory. Broadly, our work presents a novel perspective on how lexical sources may contribute to search behavior at *multiple* levels. Our findings highlight how memory search is a complex interplay of activating relevant semantic and phonological neighbors within the lexicon, and then navigating the lexicon through the use of these sources of information.

References

Abwender, D. A., Swan, J. G., Bowerman, J. T., & Connolly, S. W. (2001). Qualitative analysis of verbal fluency output: Review and comparison of several scoring methods. *Assessment*, *8*(3), 323–338.

Bouchard, M.-E., Ouellet, C., & Cohen, H. (2009). Speech development in prelingually deaf children with cochlear implants. *Language and Linguistics Compass*, *3*(1), 1–18.

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology*, *31*, 149.

Chung, Y.-A., & Glass, J. (2018). Speech2vec: A sequence-to-sequence framework for learning word embeddings from speech. *arXiv preprint arXiv:1803.08976*.

Coppens, K. M., Tellings, A., Verhoeven, L., & Schreuder, R. (2013). Reading vocabulary in children with and without hearing loss: The roles of task and word type.

Dautriche, I., Mahowald, K., Gibson, E., & Piantadosi, S. T. (2017). Wordform similarity increases with semantic similarity: An analysis of 100 languages. *Cognitive Science*, *41*(8), 2149–2169.

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). Delis-kaplan executive function system. *Assessment*.

Hay, J. F., Pelucchi, B., Estes, K. G., & Saffran, J. R. (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive psychology*, *63*(2), 93–106.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431.

Hills, T. T., Todd, P. M., & Jones, M. N. (2015). Foraging in semantic fields: How we search through memory. *Topics in cognitive science*, *7*(3), 513–534.

Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., & Couzin, I. D. (2015). Exploration versus exploitation in space, mind, and society. *Trends in cognitive sciences*, *19*(1), 46–54.

Kenett, Y. N., Wechsler-Kashi, D., Kenett, D. Y., Schwartz, R. G., Ben-Jacob, E., & Faust, M. (2013). Semantic organization in children with cochlear implants: Computational analysis of verbal fluency. *Frontiers in Psychology*, *4*, 543.

Kiela, D., & Bottou, L. (2014). Learning image embeddings using convolutional neural networks for improved multi-modal semantics. In *Proceedings of the 2014 conference on empirical methods in natural language processing (emnlp)* (pp. 36–45).

Kumar, A. A., Apsel, M., Zhang, L., Xing, N., & Jones, M. N. (2023). forager: A python package and web interface for modeling mental search. *Behavior Research Methods*, 1–17.

Kumar, A. A., Lundin, N. B., & Jones, M. N. (2022). Mouse-mole-vole: The inconspicuous benefit of phonology during retrieval from semantic memory. In *Proceedings of the annual meeting of the cognitive science society.*

Lundin, N. B., Brown, J. W., Johns, B. T., Jones, M. N., Purcell, J. R., Hetrick, W. P., ... Todd, P. M. (2023). Neural evidence of switch processes during semantic and phonetic foraging in human memory. *Proceedings of the National Academy of Sciences*, *120*(42), e2312462120.

Niparko, J. K., Tobey, E. A., Thal, D. J., Eisenberg, L. S., Wang, N.-Y., Quittner, A. L., ... others (2010). Spoken language development in children following cochlear implantation. *Jama*, *303*(15), 1498–1506.

Ormel, E., Hermans, D., Knoors, H., Hendriks, A., & Verhoeven, L. (2010). Phonological activation during visual word recognition in deaf and hearing children.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, *22*(3), 370–378.

Zemla, J. C., Cao, K., Mueller, K. D., & Austerweil, J. L. (2020). Snafu: The semantic network and fluency utility. *Behavior research methods*, *52*, 1681–1699.

Zemla, J. C., Gooding, D. C., & Austerweil, J. L. (2023). Evidence for optimal semantic search throughout adulthood. *Scientific Reports*, *13*, 22528. Retrieved from `https://doi.org/10.1038/s41598-023-49858-9` doi: 10.1038/s41598-023-49858-9