CompCog: Modeling Search within the Mental Lexicon

Richard Lim, Class of 2027

Memory search and retrieval is foundational to almost all cognitive tasks, but its underlying processes are not well understood. The Semantic Fluency Task (SFT) is a classic psychology task used to study lexical retrieval, where participants are asked to name all the words they can in a particular semantic category (eg. animals). Subjects typically produce multiple words within a category in short bursts ("clusters," eg. cat -> dog -> hamster), before moving on to another category ("switch," eg. hamster -> lion -> cheetah).

In the **existing literature**, there are two main frameworks to explain this behavior. The *optimal foraging* model compares lexical search to animals foraging for food, where they actively search for and exploit resource-rich clusters before switching to another cluster (Hills et al., 2012). The *random walk* model argues that an undirected exploration of one's lexicon would produce similar clustering patterns in the SFT because words that are similar tend to be stored "close" together in one's lexicon (Abbott et al., 2015). In age-related literature, Zemla et. al (2023) demonstrated that there are age effects on lexical retrieval (increased time in each cluster with age) but found behavior consistent with an optimal search strategy across ages.

The **goals** of our present research are to (1) build a predictive process model to identify possible mechanisms in which age affects lexical retrieval and (2) contextualize age-related findings within existing literature on both the optimal foraging and random walk models.

We used two main **methods**: quantitative statistical analysis of produced fluency lists and process modeling.

We computed correlations of age with various lexical measures using the Zemla et al. data, controlling for income and education (see Table 1). Semantic similarity was quantified using a vector representation of each word obtained via the Universal Sentence Encoder (Cer et. al, 2018). Phonological similarity was quantified using the normalized edit distance of the phonemic transcriptions of word pairs. We identified clusters and switches using hand-coded classifications (Troyer et. al, 1997). These were computed using the *forager* package (Kumar et al., 2023).

Variable	Partial Correlation	p-value
Mean pairwise semantic similarity	-0.00832	0.849
Mean pairwise phonological similarity	**-0.122**	**0.00529**
Mean frequency	-0.00134	0.976
Number of items per list	-0.0243	0.579
Number of switches	-0.00131	0.976
Mean cluster size	-0.0270	0.538
Semantic spread (normalized variance vector of embeddings)	-0.0569	0.194
Mean response time	0.0104	0.812

ıble 1: Partial correlations of lexical metrics with ag controlling for income and education

Our **preliminary results** show a significant negative partial correlation of mean pairwise phonological similarity with age after controlling for income and education, $r(522)=-.12$, p=.006. We also found, using a median split, that older participants (0.27, SD = 0.07) produced more words in common with their peers than younger participants (0.26, SD = 0.07) t (61947)=-2.03, $p < .001$. These findings imply that in younger adults, there is greater diversity in the produced fluency lists and a greater salience of phonological cues during retrieval.

In creating our process model, we produced a predictive algorithm to perform optimal foraging on a semantic network structure. Our *network forager* process model attempts to search for items within a local cluster, and switches clusters only after retrieval failures. On a semantic network with weights purely based on semantic similarities, our model appears promising, producing fluency lists with similar clustering behavior to human participants, although "atypical" animals are often listed, such as in the following excerpt produced by our model: *buck, cattle, heifer, elk, deer, red deer, white tailed deer, antelope, wildebeest, macaque, primate*.

Next steps for our research include incorporating phonological and frequency information to produce more realistic lists, and formal testing to validate our model, before examining age effects on model parameters. We also envision our model being used to examine individual-level and group-level differences in memory search.

Faculty Mentor: Abhilasha Kumar
Funded by the Gibbons Summer Research Program

References

Abbott, J. T., Austerweil, J. L., & Griffiths, T. L. (2015). Random walks on semantic networks can resemble optimal foraging. *Psychological Review*, *122*(3), 558–569. https://doi.org/10.1037/a0038693

Cer, D., Yang, Y., Kong, S., Hua, N., Limtiaco, N., St. John, R., Constant, N., Guajardo-Cespedes, M., Yuan, S., Tar, C., Strope, B., & Kurzweil, R. (2018). Universal sentence encoder for English. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. https://doi.org/10.18653/v1/d18-2029

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–440. https://doi.org/10.1037/a0027373

Kumar, A.A., Apsel, M., Zhang, L., Xing, N., Jones. M.N. (2023). forager: A Python package and web interface for modeling mental search

Troyer, A. K., Moscovitch, M., & Winocur, G. (1997). Clustering and switching as two components of verbal fluency: Evidence from younger and older healthy adults. *Neuropsychology*, *11*(1), 138– 146. https://doi.org/10.1037//0894-4105.11.1.138

Zemla, J. C., Gooding, D. C., & Austerweil, J. L. (2023). Evidence for optimal semantic search throughout adulthood. *Scientific Reports*, *13*(1), 1–12. https://doi.org/10.1038/s41598-023- 49858-9