<u>Analysis of Diachronic Language Change</u>

There are two accounts of how meaning representations are formed. According to one account, the structure of semantic knowledge results from direct perceptual experience and abstract reasoning about unseen properties of the world (Kim, 2019). Therefore, changes in the statistical patterns of the language should have little effect on how an individual's semantic representations evolve as they age. In another account, meaning representations are shaped by language experience: listeners learn word meanings in part by tracking their patterns of co-occurrence with other words (Lewis, 2019). Since the statistical patterns of words can change across decades (Hamilton, 2016; Michel, 2011), the experience-based account would suggest that older adults will arrive at different semantic representations relative to their younger counterparts, based on words whose co-occurrence patterns have changed over their lifetime. Indeed, previous work indicates that lexico-semantic processing changes with age (Dubossarsky, 2017; Federmeier, 2010).

Beginning with how meaning changes over time, I will use the free-association data from the Small World of Words (SWOW; De Deyne, 2019) and the HistWords toolkit (Hamilton, 2016), a set of word meaning vectors pre-trained on data from specific time periods. First, co-occurrence matrices for several age groups will be generated from SWOW cues and responses, which will be processed to generate representational dissimilarity matrices (RDM; Kriegeskorte, 2008) for each age group. HistWords word vectors will be used to generate per-decade RDMs using the same distance metric over the same set of words. Representational similarity analysis (RSA, Spearman rank correlation) will be used to compare the individual or combinations of HistWords matrices to the SWOW matrices, with the final goal of finding the combination of HistWords matrices which best fit to the human data.

Besides RSA, a linear model will also be used to analyze the fit between data. This could be averaging the decade representations evenly over an individual's lifespan (e.g., average from 50s onward for a 70 y.o. individual) or weighting the earlier decades more heavily to reflect the primacy of early language learning (e.g., assigning a higher weight to 60s and 70s for a 70 y.o. individual). In addition, the results will be used to identify words whose meanings have changed over the last 60 years (e.g., "Script" as in "Letter" vs "Computer program"), as well as words whose meanings appear to have remained stable.

Next, electroencephalography (EEG) will be used to observe and compare the semantic representations of individuals with different linguistic experience (i.e., from different age groups), since studies have shown that the N400 reflects semantic processing (Federmeier, 2010; Kutas, 2000). I will collect EEG data from young and older adults as they are processing words selected from the first part. On each trial, participants will hear a word and then, after a brief delay, answer a question about that word. After the EEG data are preprocessed and time- locked to the critical word, average voltage over a typical auditory N400 time window will be extracted for each channel and averaged across 10 repetitions, creating a vector of length 64 for each word. EEG-based RDM will be created for each participant using pairwise cosine distance of EEG-based word vectors.

After these two stages, RDMs will be generated to capture semantic representations from the three sources; behavioral (SWOW free-association), corpora (HistWords), and neural (EEG). RSA will be used to see if diachronic change in lexico-semantic representations is reflected in the neural representational spaces of individuals receiving language input over their lifetime spanning different time periods.

In conclusion, this research will expand on the relationship between semantic processing and linguistic experience by connecting EEG-based measures of semantic representation to those derived from large language data sets (corpora, free-association norms), in addition to focusing on diachronic differences. The results will also contribute to the debate about the importance of linguistic structure and experience in forming semantic representations, through explicit testing and manipulation of the co-occurrence statistics.

References:

De Deyne, S., Navarro, D. J., Perfors, A., Brysbaert, M., & Storms, G. (2019). The "Small World of Words" English word association norms for over 12,000 cue words. Behavior Research Methods, 51(3), 987–1006. https://doi.org/10.3758/s13428-018-1115-7

Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. Developmental Psychology, 53(8), 1560–1570. https://doi.org/10.1037/dev0000347

Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. Brain and Language, 115(3), 149–161. https://doi.org/10.1016/j.bandl.2010.07.006

Hamilton, W. L., Leskovec, J., & Jurafsky, D. (2018). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. ArXiv:1605.09096 [Cs]. http://arxiv.org/abs/1605.09096

Kim, J. S., Elli, G. V., & Bedny, M. (2019). Knowledge of animal appearance among sighted and blind adults. Proceedings of the National Academy of Sciences, 116(23), 11213–11222. https://doi.org/10.1073/pnas.1900952116

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. Frontiers in Systems Neuroscience. https://doi.org/10.3389/neuro.06.004.2008

Lewis, M., Zettersten, M., & Lupyan, G. (2019). Distributional semantics as a source of visual knowledge. Proceedings of the National Academy of Sciences, 116(39), 19237–19238. https://doi.org/10.1073/pnas.1910148116

Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2011). Quantitative Analysis of Culture Using Millions of Digitized Books. Science, 331(6014), 176–182. https://doi.org/10.1126/science.1199644