SPEECH PERCEPTION AND THE ROLE OF SEMANTIC RICHNESS IN PROCESSING

Filip Nenadić[*][1], **Matthew C. Kelley** [†][1], **Ryan Podlubny** [‡][1], **and Benjamin V. Tucker** [§][1]
[1]Department of Linguistics University of Alberta

1 Introduction

Semantic richness is ordinarily defined as a construct encompassing indicators of variability in a given word's meaning. Higher semantic richness has consistently been connected to facilitation in processing, at least when recognition tasks are used [1]. These findings, however, have come largely from studies based on visual word recognition, and related studies focusing on the acoustic signal and speech perception are less common. Effects observed may vary across modalities, as is exemplified in the opposite effect of *phonological neighborhood density* when the visual and the auditory lexical decision tasks are compared [2]. Furthermore, models of spoken word recognition tend to focus on the activation-competition process concerning some representation of the acoustic input [3], and disregard how the activation of networks of meaning based on that very acoustic input shapes competition and facilitates or inhibits speech perception. In other words, very little is known about how semantic richness interacts with perception of the acoustic, speech signal.

One study that investigated the effects related to a multitude of semantic variables in the auditory modality was conducted by Goh et al. [4]. They found linear effects of *concreteness* and the *number of semantic features* and a quadratic effect of *valence* — all of which facilitated spoken word recognition in the auditory lexical decision task. *Arousal*, *semantic neighborhood density*, and *semantic diversity* had no effect. However, stimuli used by Goh et al. were a limited set of under 500 words, most of which were concrete nouns. The goal of the present study is to expand the scope of the analysis to a larger number of words by using the data collected in the Massive Auditory Lexical Decision (MALD) project [2] and to further explore semantic richness effects as they may pertain to the perception and processing of spoken language.

2 Method

2.1 Massive Auditory Lexial Decision

Responses were collected from 231 native monolingual listeners of western Canadian English (78% female, 22% male; age ranged from 17 to 29, M = 20.11, SD = 2.39). Participants were recruited at the University of Alberta and received partial course credit for participation.

The stimuli were 26,800 English words and 9,600 phonotactically licit pseudowords recorded by one male speaker of western Canadian English. Stimuli were further organized into 134 experimental lists, with each list containing 400 word and 400 pseudoword items.

Participants completed the experiment in E-Prime [5], in a sound-attenuated booth. Their task was to listen to a single MALD experimental list and decide whether each of the stimuli presented in random order is a word of English or not, and indicate so by pressing the appropriate button on the button-box. Participants could attend as many as three sessions and would provide responses for a single MALD list every time. The total number of completed sessions was 284.

2.2 Data preparation and analysis

For the analysis, we selected responses to those MALD words that had available estimates for the variables *concreteness*, *valence*, *arousal*, *semantic neighborhood density*, and *semantic diversity*. The variable *number of semantic features* was not included as the values were available only for the limited set of words used in Goh et al. [4]. The number of retained words was 9,086, which is an 18-fold increase in comparison to the Goh et al. study. Furthermore, the words retained in this sample had a much wider range of values on the semantic richness variables (whereas Goh et al. study mostly included, e.g., high *concreteness* and low *arousal* words).

We then removed all incorrect responses and all responses shorter than 500 ms or longer than 2,000 ms (5.75% of the data excluded). Log-transformed response latency was used as the dependent variable in a generalized additive mixed-effects model [6]. The model included smoothed effects of participant *age*, *trial* number, stimulus *duration* in milliseconds, *phonological neighborhood density*, *word run length* (i.e., the number of consecutive word stimuli prior to the trial), *concreteness*, *arousal*, *valence*, and *semantic diversity* as predictors. Additionally, we included *frequency weighted semantic neighborhood density* as a predictor. Weighting was performed by multiplying log-transformed *frequency* from the Corpus of Contemporary American English [7] and *semantic neighborhood density* to avoid issues of multicollinearity because the two variables correlated highly ($r = .78$). The new variable correlated highly with both logged *frequency* ($r = .95$) and *semantic neighborhood density* ($r = .91$). All predictor values were standardized. Finally, the model included participant *sex* as a factor, as well as random effects of word stimulus and random slopes of *trial* by participant.

3 Results

The results of the generalized additive mixed-effects model showed that there was no significant effect of participant *sex*, while there was a small effect of participant *age*, as older participants were slightly faster in their responses. The effects of non-semantic predictors matched the ordinary findings — response latency was shorter in later *trials*, shorter word *duration*, words with smaller *phonological neighborhood density*,

---

[*] nenadic@ualberta.ca
[†] mckelley@ualberta.ca
[‡] podlubny@ualberta.ca
[§] bvtucker@ualberta.ca

and longer *word run length*.

Where semantic richness variables are concerned, only the effect of *semantic diversity* was not significant. *Frequency weighted semantic neighborhood density* had a negative linear effect (Figure 1a). We also find that high *concreteness* is facilitatory, while low *concreteness* is inhibitory (Figure 1b). When *valence* is considered, extreme values (i.e., both negative and positive *valence*) are connected to shorter response latency (Figure 1c). High *arousal* predicted longer response latency, while there was no difference between words with average versus low *arousal* (Figure 1d).

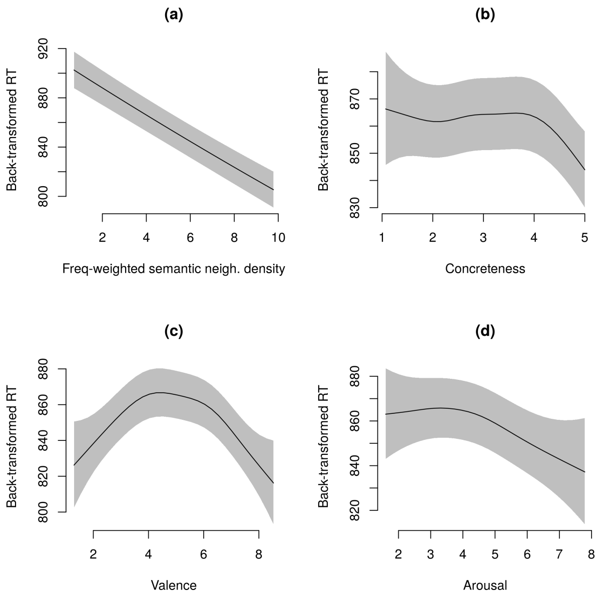

Figure 1: Effects of the four significant semantic predictors of response latency.

4 Discussion

Our results differ in some important ways from those of Goh et al. [4]. The effect of *valence* in our study is not exactly quadratic, but that is the smallest change we note. We find a non-linear effect of *concreteness* — besides in high values covered in Goh et al., we find the same effect in low values, but not in words with average *concreteness*. Furthermore, we record significant effects of *arousal* and *frequency weighted semantic neighborhood density*. The effect of *arousal* likewise seems to be explained by the distribution of this variable — we find an effect in higher *arousal* values, whereas the limited dataset previously used by Goh et al. mostly included words with low *arousal*. For *semantic neighborhood density*, since the distributions in the two datasets are comparable, perhaps the effect is significant simply due to larger sample size.

Although the relationship is not always linear or simple, we find that higher values in variables that operationalize semantic richness tend to be connected with faster response latency. Given these findings, we claim that at least some semantic characteristics of a word play a role in speech perception, facilitating processing. As the acoustic signal unfolds and competitor words are activated, word meanings are accessed, not just their form. This process would arguably be highly adaptive, as it may allow the listener to quickly access items that carry more (and more important) meaning. In comparison, most models of spoken word recognition treat the mental lexicon as a simple list of words [3], focusing solely on the process of matching (acoustic) input to abstract representation when describing speech perception.

5 Conclusions

Our results indicate that semantics can play a role in speech perception even when isolated words are presented. We argue that models of spoken word recognition should treat words in the mental lexicon as more than simple lists of items by taking into account their semantic richness and semantic connections. We also show that additional studies using large word samples with an encompassing variability of word characteristics are needed [8], especially given the discrepancies in the results between word sets of different sizes.

Acknowledgments

This project was funded by the Social Sciences and Humanities Research Council: Grant #435-2014-0678.

References

[1] Melvin J. Yap, Sarah E. Tan, Penny M. Pexman, and Ian S. Hargreaves. Is more always better? Effects of semantic richness on lexical decision, speeded pronunciation, and semantic classification. *Psychonomic Bulletin & Review*, 18(4):742–750, 2011.

[2] Benjamin V. Tucker, Daniel Brenner, D. Kyle Danielson, Matthew C. Kelley, Filip Nenadić, and Michelle Sims. The Massive Auditory Lexical Decision (MALD) database. *Behavior Research Methods*, 51(3):1187–1204, Jun 2019.

[3] Andrea Weber and Odette Scharenborg. Models of spoken-word recognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 3(3):387–401, 2012.

[4] Winston D. Goh, Melvin J. Yap, Mabel C. Lau, Melvin M. R. Ng, and Luuan-Chin Tan. Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in Psychology*, 7:976, 2016.

[5] Walter Schneider, Amy Eschman, and Anthony Zuccolotto. *E-Prime reference guide*. Psychology Software Tools, Incorporated, 2002.

[6] Simon N. Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1):3–36, 2011.

[7] Mark Davies. The 385+ million word corpus of contemporary american english (1990–2008+): Design, architecture, and linguistic insights. *International Journal of Corpus Linguistics*, 14(2):159–190, 2009.

[8] Emmanuel Keuleers and David A. Balota. Megastudies, crowdsourcing, and large datasets in psycholinguistics: An overview of recent developments. *Journal of Experimental Psychology*, 68(8):1457–1468, 2015.

[9] Lori Buchanan, Chris Westbury, and Curt Burgess. Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, 8(3):531–544, 2001.