SEMANTIC DIVERSITY IS NOT A CONTINUOUS MEASURE OF SEMANTIC AMBIGUITY

Benedetta Cevoli (Psychology Department, Royal Holloway, University of London), Chris Watkins (Computer Science Department, Royal Holloway, University of London) and Kathleen Rastle (Psychology Department, Royal Holloway, University of London)
Benedetta.Cevoli.2018@live.rhul.ac.uk

Previous research has speculated that semantic diversity and lexical ambiguity may be closely related constructs. Our research sought to test this claim in respect of the semantic diversity measure proposed by Hoffman et al. (2013). To this aim, we replicated the procedure described by Hoffman et al. (2013) for computing multidimensional representations of contextual information using Latent Semantic Analysis, and from these we derived semantic diversity values for 28,555 words. We then replicated the faciliatory effect of semantic diversity on word recognition using existing data resources and observed this effect to be greater for low frequency words. Yet, simulation analyses of Rodd et al. (2002) and Armstrong and Plaut (2016) on sematic diversity showed no relationship between this measure and lexical ambiguity effects in word recognition. Further analysis of the LSA-based contextual representations used to compute Hoffman et al.'s (2013) measure of semantic diversity revealed that they do not capture the distinct meanings of ambiguous words. Instead, these contextual representations appear to capture general information about the topics and types of written material in which words occur (Figure 1). These analyses suggest that the semantic diversity metric previously proposed by Hoffman et al. (2013) facilitates word recognition because high diversity words are likely to have been encountered no matter what one has read, whereas many participants may not have encountered lower diversity words simply because the topics and types of written material in which they occur are more restricted. In other words, semantic diversity facilitates recognition because it measures whether a word is a generalist (occurring in many topics and types of documents) or a specialist (occurring only in a few topics and types of document).

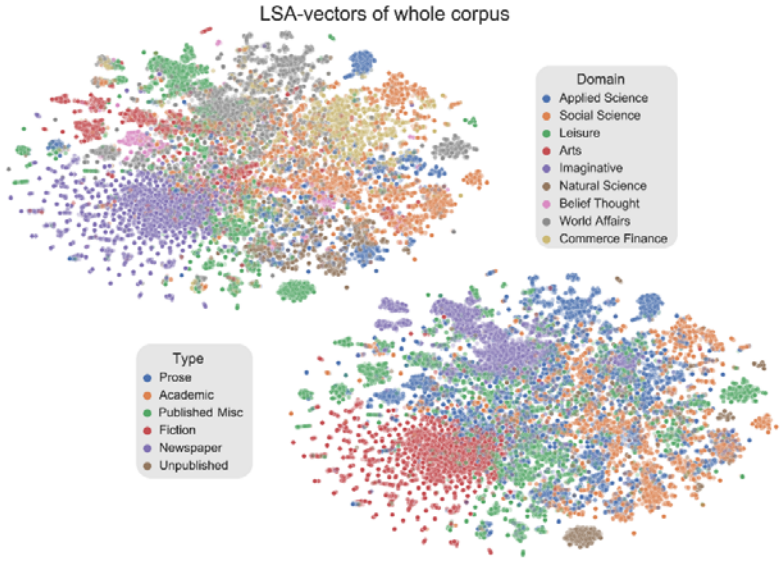

Figure 1. t-SNE plots of LSA-context vectors of the whole corpus by domain and type of written material.

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966.

Hoffman, P., Lambon Ralph, M. A., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*(3), 718–730.

Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making Sense of Semantic Ambiguity: Semantic Competition in Lexical Access. *Journal of Memory and Language*, *46*(2), 245–266.