

Dyer, William

Oracle Corporation

Quan, Tracey

University of Delaware

Almost but not quite: Using formulaicity to assess L2 oral proficiency

The notion that native speakers tend to produce language that is more formulaic—words that tend to go together—than language learners is widely held (Wray & Perkins, 2000; Wray, 2002).

Research on second language proficiency has attempted to measure the formulaicity of learner language (cf., Wood, 2010, 2015), sometimes captured as the pointwise mutual information (MI) of contiguous words in an utterance (cf., Serrano et al., 2015; Quan, forthcoming). Yet such a linear approach to formulaicity falls short due to the variability of n-grams within a given phrase, as well as the propensity for language learners to make syntactic errors (Ellis, 2012). A dependency grammar framework provides an alternate solution to measuring formulaicity. Unlike a phrase-structure grammar, a dependency grammar is built from the relationship between a head word and its dependent words, which need not be adjacent (Tesnière, 1959). Further, recent SLA work has shown that a dependency approach is able to describe previously unexplored facets of L2 learner proficiency (Ouyang & Jiang, 2017). Therefore, this paper proposes that the MI of each pair of words within a dependency relationship can be used as a quantified measure of formulaicity.

The current exploratory study applies such an approach to the oral transcripts of five L2 Spanish learners of varying proficiency levels and five native Spanish speakers. Utilizing the Universal Dependencies corpora (Nivre et al., 2016), the findings show that the average MI of the native speaker data is higher than that of the L2 learners. In addition, more advanced L2 learners show higher average MI compared to their less advanced counterparts. In other words, as learners increase their L2 proficiency, the formulaicity of their speech also increases. These results provide initial empirical evidence for the validity of a dependency grammar approach to measuring formulaicity. Moreover, the approach provides insight into the assessment of L2 oral speech, as it seems to capture an aspect of the subtle differences between an advanced language learner and a target language speaker, warranting further research with a larger data set.

References:

- Ellis, N. C. (2012). Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics*, 32, 17–44.
- Nivre, J., M.-C. de Marneffe, F. Ginter, Y. Goldberg, et al. (2016). Universal dependencies v1: A multilingual treebank collection. *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016)*, 1659–1666.
- Ouyang, J. & Jiang, J. (2017). Can the probability distribution of Dependency Distance Measure Language Proficiency of Second Language Learners? *Journal of Quantitative Linguistics*, 1-19.

- Quan, T. (forthcoming). Acquisition of formulaic sequences in a study abroad context. *Study Abroad Research in Second Language Acquisition and International Education*.
- Serrano, R., Stengers, H., & Housen, A. (2015). Acquisition of formulaic sequences in intensive and regular EFL programmes. *Language Teaching Research*, 19(1), 89–106.
- Tesnière, L. (1959). *Éléments de syntaxe structural*. Paris: Klincksieck.
- Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence, and classroom applications*. London/New York: Continuum.
- Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction*. London: Bloomsbury Publishing.
- Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: an integrated model. *Language & Communication*, 20(1), 1–28.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.