A dynamic usage-based investigation of formulaicity and lexical complexity development: The case of sojourners

L2 learners struggle to learn how to combine words in authentic, native-like ways on the way to proficiency. When it comes to this challenging aspect of L2 development, learners might benefit from meaningful authentic input to develop their uses of formulaic expressions. Operationalizing formulaicity as *conventionalized ways of saying things* (Smiskova-Gustafsson et al., 2012), this study asserts that learning contexts are a crucial part of the picture as some are deprived of authentic input, while some others provide a wealth of examples, such as the study abroad context.

Taking a dynamic usage-based perspective to L2 development (Verspoor & Behrens, 2011), this study investigates how the study abroad context interplays with formulaicity and lexical complexity development. Assuming that input and meaningful interaction are key components of SLD, this study also aims to explore the learning trajectories of a group of sojourners as to lexical diversity, variation, sophistication, and formulaicity. The researchers analyzed written performance data in the form of weekly diary entries, as part of a larger project (Pérez-Vidal, 2014). The participants are 26 Catalan/Spanish tertiary level English learners who studied abroad in an Anglophone country for 14-17 weeks. The SALA diary corpus is compiled of a total of 383 weekly diary entries (~250K words) about sojourners' experiences related to language use, interaction, and host culture over the course of their stay abroad. The dataset was coded for various measures of lexical complexity (lexical density, sophistication, and diversity) using automated tools. To determine how formulaic each weekly entry was, the dataset was also analyzed through IdiomSearch (Colson, 2016) for multi-word constructions and by human raters for holistic-formulaicity. The results of the linear mixed-effects models showed that sojourners made significant gains towards holistic-formulaicity but using significantly less diverse words and fewer multi-word constructions per tokens, suggesting a discrepancy between automated and human ratings.

References

Colson, JP. (2017). The *IdiomSearch* experiment: Extracting phraseology from a probabilistic network of constructions. In: Mitkov, R. (Ed), *Computational and Corpus-Based Phraseology*. EUROPHRAS 2017 Proceedings. Springer: Cham. https://doi.org/10.1007/978-3-319-69805-2_2.

Pérez-Vidal, C., & Barquin, E. (2014). Comparing progress in writing after formal instruction and study abroad. In C. Pérez-Vidal (Ed.), *Language acquisition in study abroad and formal instruction contexts* (pp. 217–234). Amsterdam: John Benjamins. https://doi.org/10.1075/aals.13

Smiskova-Gustafsson, H., Verspoor, M., & Lowie, W. (2012). Conventionalized ways of saying things (CWOSTs) and L2 development. *Dutch Journal of Applied Linguistics*, *1*(1), 125-142.

Verspoor, M.H., & Behrens, H. (2011). Dynamic systems theory and a usage-based approach to second language development. In M. H. Verspoor, K. de Bot, & W. Lowie (Eds.), A dynamic approach to second language development (pp. 25-38). Philadelphia: John Benjamins. https://doi.org/10.1075/lllt.29