Derived Words in Texts: Effects of text level and word frequency

The word family is the most frequently used counting unit in word lists (Coxhead, 2000; Nation, 2006), vocabulary tests (Nation & Beglar, 2007; Webb, Sasao, & Ballance, 2017), and studies of lexical coverage (Nurmukhamedov & Webb, 2019). Recently however it has been suggested that family-based vocabulary tests are flawed because learners may be unable to understand all the derived forms of basewords (McLean, 2018). This claim assumes that derived words contribute a substantial amount of text coverage, and the lack of understanding them will hamper comprehension. Conversely, Laufer & Cobb (2020) show that derived words constitute only ~3%-8% of text vocabulary, depending on text genre. But they did not study the distribution of derived words by text level, or word frequency, information that would show how morphological knowledge affects comprehension at different stages of learning.

We therefore investigated the effects of text level and word frequency on the morphological composition of texts. We used three tools from *lextutor.ca* to carry out (1) a between-text analysis by word family size and (2) a within-text analysis by word frequency level in graded readers, mid-frequency readers, novels, and academic texts (~ 1.5 million words). The between-text comparison was performed by the Nuclear Family Builder, which extracts recurring word families in texts showing the number of family members in each. The within-text comparison combined VocabProfile and Morpholex yielding an affix analysis per word frequency level.

The two analyses showed that the number of derived forms in word families is low in basic texts then grows as (1) the between-text language level increases and (2) the within-text word frequency level decreases. So the number of derived forms in texts is proportional to text level and lexical sophistication. We relate this to the contribution of morphological knowledge to comprehension and suggest a

References

Aviad-Levitzky, T., Laufer, B., & Goldstein, Z. (2019). The new computer adaptive test of size and strength (CATSS): Development and validation. *Language Assessment Quarterly*, *16*(3), 345–368.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Laufer, B., & Cobb, T. (2020). How much knowledge of derived words is needed for reading? *Applied Linguistics*, 41 (6), 971–998 doi:10.1093/applin/amz051

McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, *39*(6), 823–845.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review*, *63*(1), 59–82

Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, *31*(7), 9-13.

Nurmukhamedov, U., & Webb, S. (2019). Lexical coverage and profiling. *Language Teaching*, *52*(2), 188–200.

Ward, J., & Chuenjundaeng, J. (2009). Suffix knowledge: Acquisition and applications. *System*, 37(3), 461–469

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *ITL - International Journal of Applied Linguistics*, *168*(1), 33–69.