Psycholinguistic evidence against frequency effects for multi-morphemic sequences in Japanese

Hikaru Hotta
University of Tokyo, hihikahihi@g.ecc.u-tokyo.ac.jp

Keywords: Usage-based models of language, Frequency, Multi-word sequences, Multi-morphemic sequences, Japanese

Usage-based theories of language predict that speakers store linguistic patterns of varying sizes and schematicity, including lexically specific constructions such as collocations, of which meanings are predictable from their respective component parts (e.g., Wray 2002; Goldberg 2019). One question that has garnered attention is what factors are involved in the vast storage of lexically specific constructions. Among a variety of factors that have been investigated, the effects of token frequency have been particularly studied and discussed. Some psycholinguistic studies show that speakers are sensitive to frequency information of multi-word sequences (e.g., Arnon & Snider 2010; Tremblay et al. 2011), while others cast doubt on the causal effect of frequency in the processing of multi-word sequences (e.g., Jolsvai et al. 2020). Due to a lack of cross-linguistic research on this topic, the processing of multi-morphemic sequences in agglutinative languages has not been previously investigated (Arnon 2021).

Building on this background, the present study investigates frequency effects for multi-morphemic sequences in Japanese. Specifically, the study aims to examine whether highly frequent multi-morphemic sequences are processed faster than their less frequent counterparts in an experimental setting. 24 item pairs of multi-morphemic sequences that differ in the first morpheme and the whole-string frequency (e.g., 基本-的-に-は [basis like COP TOP] (high), 技術-的-に-は [technology like COP TOP] (low)) are used in the experiment. In addition to controlling for frequencies of the first unigrams (e.g., basic [基本] and technology [技術]), other potentially confounding variables such as the number of characters, morae, Kanji (Chinese character), and compositionality of meaning are all matched within an item pair. Processing speed is measured in a phrasal decision task in which participants are asked to judge whether the sequences that appear on the computer screen are possible sequences as accurately and quickly as possible (Arnon & Snider 2010).

The results show that there was little difference in reaction times between the high-frequency condition (mean 645.1 ms) and the low-frequency condition (mean 651.7 ms) in the phrasal decision task. A linear mixed-effects model corroborates the descriptive result and shows that frequency is not a strong or significant factor in predicting the reaction times. A post-hoc item analysis further reveals that the difference in mean reaction times between the high-frequency condition and the low-frequency condition is substantially different depending on the target item. These results suggest the existence of another variable impacting the previous studies' findings of frequency effects for multi-word sequences. In light of the fact that highly frequent multi-word sequences and multi-morphemic sequences are often associated with discourse functions that may arguably be non-compositional (Biber et al. 2004; Kaneyasu 2012), the present results call for taking context-dependent meanings into account when studying the processing of recurrent sequences.

Gloss abbreviations
COP copula
TOP topic marker

References
Arnon, Inbal. 2021. The starting big approach to language learning. *Journal of Child Language,* 48(5), 937-958. https://doi.org/10.1017/S0305000921000386

Arnon, Inbal., & Snider, Neal. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62, 67-82. https://doi.org/10.1016/j.jml.2009.09.005

Biber, Douglas., Conrad, Susan., & Cortes, Viviana. 2004. If you look at …: Lexical bundles in university teaching and textbooks. *Applied Linguistics*, 25(3), 371-405. https://doi.org/10.1093/applin/25.3.371

Goldberg, Adele E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton: Princeton University Press.

Jolsvai, Hajnal., McCauley, Stwart M., & Christiansen, Morton H. 2020. Meaningfulness beats frequency in multiword chunk processing. *Cognitive Science*, 44(10), e12885-n/a. https://doi.org/10.1111/cogs.12885

Kaneyasu, Michiko. 2012. *From Frequency to Formulaicity: Morphemic Bundles and Semi-Fixed Constructions in Japanese Spoken Discourse*. [Doctoral dissertation, UCLA]. Retrieved from https://escholarship.org/uc/item/1zp613xj

Tremblay, Antonie., Derwing, Bruce., Libben, Gray., & Westbury, Chris. 2011. Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569-613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Wray, Alison. 2002. *Formulaic language and the lexicon*. Cambridge: Cambridge University Press.