

Beyond Vocabulary Size: Individual Differences in the Organisation of the Mental Lexicon

Bálint József Ugrin (ugrin.balint.jozsef@ttk.bme.hu)¹, Péter Rácz¹, Ágnes Lukács¹

¹Department of Cognitive Science, Faculty of Natural Sciences,
Budapest University of Technology and Economics

The mental lexicon varies in size [1,2] and internal organisation [3]. Vocabulary size is well documented and tied to cumulative language exposure [4,5,6], but considerably less is known about individual differences in lexical connectivity – how strongly words and multiword expressions are associated in the lexicon [cf. 7]. We investigated both dimensions, alongside idiom knowledge, in 130 native Hungarian speakers aged 19–73 years (median age = 27).

Vocabulary size was assessed using a large-scale familiarity task in which participants identified familiar words among panels containing both real words and nonce words. Responses were collected for 500 words varying in frequency. Idiom knowledge was measured using a similar familiarity task involving 45 lexicalized figurative expressions. To assess sensitivity to lexical association strengths, participants completed a forced-choice preference task using 50 semantically transparent adjective–noun combinations. In each trial, participants selected the most natural-sounding expression among semantically similar alternatives and a lure item. Association strengths were parametrised using corpus-based [8] metrics of lexical association, including mutual information and logDice [9].

Results showed a strong positive correlation between vocabulary size and idiom knowledge ($r = .51$), which may reflect the influence of cumulative language exposure as a shared underlying factor, as both measures increased with age across the sample. Furthermore, sensitivity to lexical associations was positively associated with vocabulary size ($r = .35$), while a weaker but still significant relationship was observed for idiom knowledge ($r = .18$). Importantly, unlike vocabulary size and idiom knowledge, sensitivity to lexical associations did not show significant age-related effects. Although individuals with larger vocabularies tend to exhibit stronger lexical-semantic associations, the relationship is unlikely to be explained solely by cumulative language exposure. Tentative support for this interpretation comes from the comparison of different vocabulary measures: vocabulary scores derived from age-balanced lexical items showed a substantially stronger relationship with sensitivity to lexical associations ($r = .35$) than vocabulary scores based on age-skewed lexical items ($r = .12$). Together, these findings suggest that the size and structure of the mental lexicon may represent a partly independent dimension of language ability, motivating future research into the stable individual factors underlying the dissociation between lexical growth and lexical-semantic organisation.

References

- [1] Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). WordBank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44, 677–694. <https://doi.org/10.1017/S0305000916000209>
- [2] Keuleers, E., Stevens, M., Mandera, P., & Brysbaert, M. (2015). Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology*, 68(8), 1665–1692. <https://doi.org/10.1080/17470218.2015.1022560>
- [3] Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism*, 2(3), 219–253. <https://doi.org/10.1075/lab.2.3.01dab>
- [4] Dąbrowska, E. (2018). Experience, aptitude and individual differences in native language ultimate attainment. *Cognition*, 178, 222–235. <https://doi.org/10.1016/j.cognition.2018.05.018>
- [5] Hintz, F., Dijkhuis, M., Van 'T Hoff, V., McQueen, J. M., & Meyer, A. S. (2020). A behavioural dataset for studying individual differences in language skills. *Scientific Data*, 7(1), 429. <https://doi.org/10.1038/s41597-020-00758-x>
- [6] Kidd, E., & Donnelly, S. (2020). Individual Differences in First Language Acquisition. *Annual Review of Linguistics*, 6(1), Article 1. <https://doi.org/10.1146/annurev-linguistics-011619-030326>
- [7] McConnell, K., & Blumenthal-Dramé, A. (2021). Usage-Based Individual Differences in the Probabilistic Processing of Multi-Word Sequences. *Frontiers in Communication*, 6, 703351. <https://doi.org/10.3389/fcomm.2021.703351>
- [8] Nemeskey, D. M. (2020). *Natural Language Processing Methods for Language Modeling* [PhD Thesis]. Eötvös Loránd University.
- [9] Rychlý, P. (2008). A lexicographer-friendly association score. *Proceedings of Recent Advances in Slavonic Natural Language Processing, RASLAN*.