

THE EFFECTS OF LEXICAL FREQUENCY AND MORPHEME TYPE ON ANTICIPATORY VOICE ASSIMILATION IN BULGARIAN OBSTRUENTS

BISTRA ANDREEVA, MITKO SABEV

SAARLAND UNIVERSITY

andreeva@lst.uni-saarland.de, msabev@lst.uni-saarland.de

This study investigates the relationship between the frequency (or predictability) of linguistic units and the anticipatory voice assimilation of obstruents in Bulgarian. Using a speech corpus and word-level language models, we calculate the surprisal of word forms ending in the consonants /t/, /f/, and /x/ in inflectional and lexical morphemes. We then employ a linear mixed model (LMM) to analyse how surprisal, the voicing of the following obstruent, and morpheme type affect the voicing realised in the examined consonants. The findings demonstrate that linguistic predictability affects articulatory precision by modulating the degree of anticipatory assimilation, with different effect sizes for different morpheme types. More specifically, obstruents in less predictable words are articulated with greater precision than those in more predictable words, especially within lexical morphemes. This reveals a complex interaction between information density, morphology and articulation, providing avenues for further research in phonetics and phonology.

Keywords: lexical frequency; surprisal; regressive assimilation; voicing

REFERENCES

- Aylett, Turk 2004: *Aylett, M., A. Turk*. The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. – *Language and Speech*, 47, pp. 31–56.
- Aylett, Turk 2006: *Aylett, M., A. Turk*. Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei. – *Journal of the Acoustical Society of America*, 119 (1), pp. 30–48.
- Bates et al. 2015: *Bates, D., M. Maechler, B. Bolker, S. Walker, R. H. Bojesen Christensen, N. Singmann, B. Dai*. Package 'lme4': Linear mixed-effects models using Eigen and S4, February 20, 2015. [Online]. <http://cran.rproject.org/web/packages/lme4/lme4.pdf>.
- Benner et al 2007: *Benner, U., I. Flechsig, G. Dogil, G., B. Möbius*. Coarticulatory resistance in a mental syllabary. – In: *Proceedings of the International Congress of Phonetic Sciences (Saarbrücken)*, pp. 485–488.
- Boersma 2022: *Boersma, P., D. Weenink*. Praat: doing phonetics by computer, version 6.4.17 (computer program).

- Bybee, Scheibman 1999: *Bybee, J., J. Scheibman, J.* The effect of usage on degrees of constituency: the reduction of don't in English. – *Linguistics*, 37, pp. 575–596. doi: 10.1515/ling.37.4.575.
- Carreiras, Perea 2004: *Carreiras, M., M. Perea.* Naming pseudowords in Spanish: effects of syllable frequency. – *Brain and Language*, 90 (1-3), pp. 393–400. doi: 10.1016/j.bandl.2003.12.003.
- Cholin et al. 2006: *Cholin, J., W. J. Levelt, N. O. Schiller.* Effects of syllable frequency in speech production. – *Cognition*, 99 (2), pp. 205–235. doi: 10.1016/j.cognition.2005.01.009.
- Cohen Priva 2015: *Cohen Priva, U.* Informativity affects consonant duration and deletion rates. – *Laboratory Phonology*, 6 (2), pp. 243–278.
- Croot, Rastle 2004: *Croot, K., K. Rastle, K.* Is there a syllabary containing stored articulatory plans for speech production in English? – In: *Proceedings of the 10th Australian International Conference on Speech Science and Technology (Sydney)*, pp. 376–381.
- Hale 2001: *Hale, J.* A probabilistic Earley parser as a psycholinguistic model. – In: *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics, Stroudsburg, PA.* Association for Computational Linguistics, pp. 159–166. doi: 10.3115/1073336.1073357.
- Hateva, Mitankin, Mihov 2016: *Hateva, N., P. Mitankin, S. Mihov.* BulPhonC. Bulgarian Speech Corpus for the development of ASR technology. – In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pp. 771–774.
- Ibrahim et al. 2022: *Ibrahim, O., I. Yuen, M. van Os, B. Andreeva, B. Möbius.* The combined effects of contextual predictability and noise on the acoustic Realization of German syllables. – *Journal of the Acoustical Society of America*, 152 (2), pp. 911–920.
- Jaeger 2010: *Jaeger, T. F.* Redundancy and reduction: Speakers manage syntactic information density. – *Cognitive Psychology*, 61 (1), pp. 23–62.
- Jurafsky et al. 2001: *Jurafsky, D., A. Bell, M. Gregory, W. D. Raymond.* Probabilistic relations between words: evidence from reduction in lexical production. – In: Bybee, J., P. Hopper (eds.) *Frequency and the Emergence of Linguistic Structure*. Amsterdam, Benjamins, pp. 229–254.
- Kuznetsova et al. 2016: *Kuznetsova, A., P. Bruun Brockhoff, R. Haubo Bojesen Christensen,* lmerTest: Tests in Linear Mixed Effects Models, r package version 2.0-33. [Online]. <<https://CRAN.R-project.org/package=lmerTest>>.
- Levelt 1999: *Levelt, W. J. M.* Producing spoken language: a blueprint of the speaker. – In: Brown, C. M., P. Hagoort (eds.). *The Neurocognition of Language*. Oxford, Oxford University Press, pp. 83–122. doi: 10.1093/acprof:oso/9780198507932.003.0004.
- Levelt, Wheeldon 1994: *Levelt, W. J., L. Wheeldon.* Do speakers have access to a mental syllabary? – *Cognition*, 50, pp. 239–269.
- Levy 2008: *Levy, R.* Expectation-based syntactic comprehension. – *Cognition*, 106 (3), pp. 1126–1177. doi: 10.1016/j.cognition.2007.05.006.
- Losiewicz 1992: *Losiewicz, B. L.* The Effect of Frequency on Linguistic Morphology. University of Texas at Austin. PhD thesis.

- Malisz 2018: *Malisz, Z., E. Brandt, B. Mobius, Y. Oh, B. Andreeva*. Dimensions of segmental variability: interaction of prosody and surprisal in six languages. – *Frontiers in Communication*, 3, pp. 1–18.
- Mitankin, Mihov, Tinchev 2009: *Mitankin, P., S. Mihov, T. Tinchev*. Large vocabulary continuous speech recognition for Bulgarian. – In: *Proceedings of Recent Advances in Natural Language Processing (RANLP) 2009*, pp. 246–250.
- Pierrehumbert 2001: *Pierrehumbert, J. B.* Exemplar dynamics: Word frequency, lenition and contrast. – In: Bybee, J., P. Hopper (eds.). *Frequency and the Emergence of Linguistic Structure*. Amsterdam, Benjamins, pp. 137–157.
- R Core Team 2024: *R Core Team*. R: A Language and Environment for Statistical Computing, version 4.4.1. Vienna. <https://www.R-project.org/>.
- Vu et al. 2010: *Vu, N.T., T. Schlippe, F. Kraus, T. Schultz*. Rapid bootstrapping of five Eastern European languages using the rapid language adaptation toolkit. – In: *Proceedings of Interspeech*, 2010, pp. 865–868, doi: 10.21437/Interspeech.2010-292.
- Whiteside, Varley 1998: *Whiteside, S. P., R. A. Varley*. Dual-route phonetic encoding: some acoustic evidence. – In: *Proceedings of the 5th International Conference on Spoken Language Processing (Sydney)*, pp. 3155–3158.

✉ *Prof. Dr. Bistra Andreeva*
Department of Language Science and
Technology, Saarland University
Building C7.2, Room 5.02
66123 Saarbrücken, Germany

✉ *Dr. Mitko Sabev*
Department of Language Science and
Technology, Saarland University
Building C7.2, Room 4.04
66123 Saarbrücken, Germany