THE EFFECT OF LEXICAL FREQUENCY ON ANTICIPATORY VOICING ASSIMILATION IN BULGARIAN OBSTRUENTS

Mitko Sabev, Bistra Andreeva, Bernd Möbius, Ivan Yuen, Omnia Ibrahim

Department of Language Science and Technology, Saarland University
msabev@lst.uni-saarland.de

Abstract: This study investigates the relation between the surprisal (or unpredictability) of linguistic items and anticipatory voicing assimilation in Bulgarian obstruents. Using a corpus of speech read by 140 Bulgarian speakers and word-level language models, we calculated unigram surprisal for word forms ending in obstruents followed by a word-initial obstruent of the opposite underlying [±voice] specification. Percentage of voicing was computed for 9,712 word-final obstruents. Linear mixed models were used to determine the effect of surprisal on the percentage of voicing in assimilating obstruents. The results confirm that Bulgarian obstruents do indeed in general assimilate to the voicing of a following obstruent: voiceless obstruents become voiced before voiced ones, while voiced obstruents are devoiced before voiceless ones. Crucially, however, surprisal had a significant effect on the percentage of voicing found in assimilating obstruents: in words with higher surprisal values, we found significantly lower degrees of voicing in voiceless obstruents before voiced ones, as well as significantly less devoicing of voiced obstruents before voiceless ones. This shows that assimilation is stronger in low-surprisal words, while in high-surprisal words speakers attempt to maintain the underlying [±voice] specification of an obstruent to a higher degree. Our findings add to a growing body of research that demonstrates that processes once thought of as entirely categorical in fact exhibit gradient variation in fine phonetic detail, which is attributable to speakers' awareness of statistical patterns in language use and their response to the predictability of linguistic items in maintaining a balance between phonetic encoding and information density.

1 Introduction

Language offers speakers a multitude of choices in how to encode messages: from the durational, spectral, and prosodic properties of segments and syllables, to word selection, the ordering of syntactic constituents, and the arrangement of sentences in discourse. In recent years, information theory has increasingly been employed to gain a better understanding of the effect of information density on variability in language. Research has shown that speakers and listeners have access to probability distributions of linguistic units [1, 2]. Easily predictable units tend to be weakened or even deleted, while less expected items remain more distinct [3, 4, 5, 6, 7].

One quantitative measure of information density is *surprisal*. Surprisal theory [8, 9] has been very successful at explaining difficulties in the perception, processing, storage, retrieval and production of linguistic items. The theory posits that difficulty is directly proportional to surprisal, or the unexpectedness of a linguistic item, which is typically quantified as negative log

probability. Language models are often trained at the level of word form but can, in principle, be designed for units at various levels of linguistic structure.

It has been established that speech production is sensitive to word frequency (absolute number of occurrences in a corpus). For instance, the speech production model advanced by Levelt and colleagues suggests an interaction between frequency and the encoding of articulatory processes [10, 11]. High-frequency syllables, which typically occur in high-frequency words, are articulated faster than low-frequency ones [12, 13]. English high-frequency past-tense verb forms are produced faster than lower-frequency ones [14]. Frequent words and collocations (such as *don't* in *don't know*) undergo more reduction and lenition than rarer items [15, 16]. High-frequency syllables exhibit more coarticulation than less frequent syllables [17, 18]. Experimental work on coarticulatory effects in read speech has reported significant differences between high- and low-frequency syllables in words and nonsense words in terms of several variables defining vowels in the perceptual space, including formant trajectories and formant transitions [19].

The present article explores the relationship between predictability and anticipatory voicing assimilation in Bulgarian obstruents, which are assumed to assimilate to the [±voice] feature of a following obstruent across word and morpheme boundaries [20]. Our hypothesis is that speakers modulate the details of phonetic implementation to maintain a balance between phonetic encoding and information density, resulting in a negative correlation between articulatory and acoustic precision, on the one hand, and predictability of linguistic items, on the other. We thus expect that speakers will attempt to maintain the underlying voicing of assimilating obstruents in unexpected items to a measurably larger extent than in more predictable items.

2 Material and methods

The material analysed is continuous speech from the Bulgarian Phonetic Corpus *BulPhonC*, version 3, consisting of 319 phonetically rich sentences read by 140 speakers (81 female and 59 male), with a mean speaker age of 37 (SD 16, median 30) [21]. The corpus was designed for the development of ASR technology. The recordings were made in an echo-cancelling studio with a Sennheiser MK 4 omnidirectional microphone on a TASCAM DP32 digital recorder at a sampling rate of 48 kHz and 24 bits, filtered and down-sampled to 16 kHz. Canonical transcriptions and automatic phoneme segmentation are available [22].

Unigram surprisal refers to the ratio of occurrences of a linguistic unit (here a word form) among all units of the same type (word forms) in a training text (as opposed to the absolute word frequency count). It is a measure of the predictability of a linguistic unit expressed as the amount of conveyed information and is calculated as shown in Equation (1), where S is surprisal, X is the linguistic unit, and P is the probability of that unit occurring in the language or corpus; the binary logarithm renders the measure in bits.

$$S(X_i) = -\log_2 P(X_i) \tag{1}$$

Easily predictable units have low surprisal, while unexpected units have high surprisal. In other words, predictable linguistic expressions convey less information than surprising expressions. To test our hypothesis, we analysed obstruent voicing in 9,712 two-word combinations (4,527 word-final voiced obstruents preceding voiceless obstruents and 5,185 word-final voiceless obstruents preceding voiced obstruents). Unigram surprisal was calculated for each first word in a two-word combination.

The voicing profiles of word-final obstruents were computed using a Praat [23] script[1] that

[1] `http://menzerath.phonetik.uni-frankfurt.de/tools/Pitch_5_0_0.praat`
`http://www.phonetik.uni-frankfurt.de/tools/Pitch_manual.pdf`

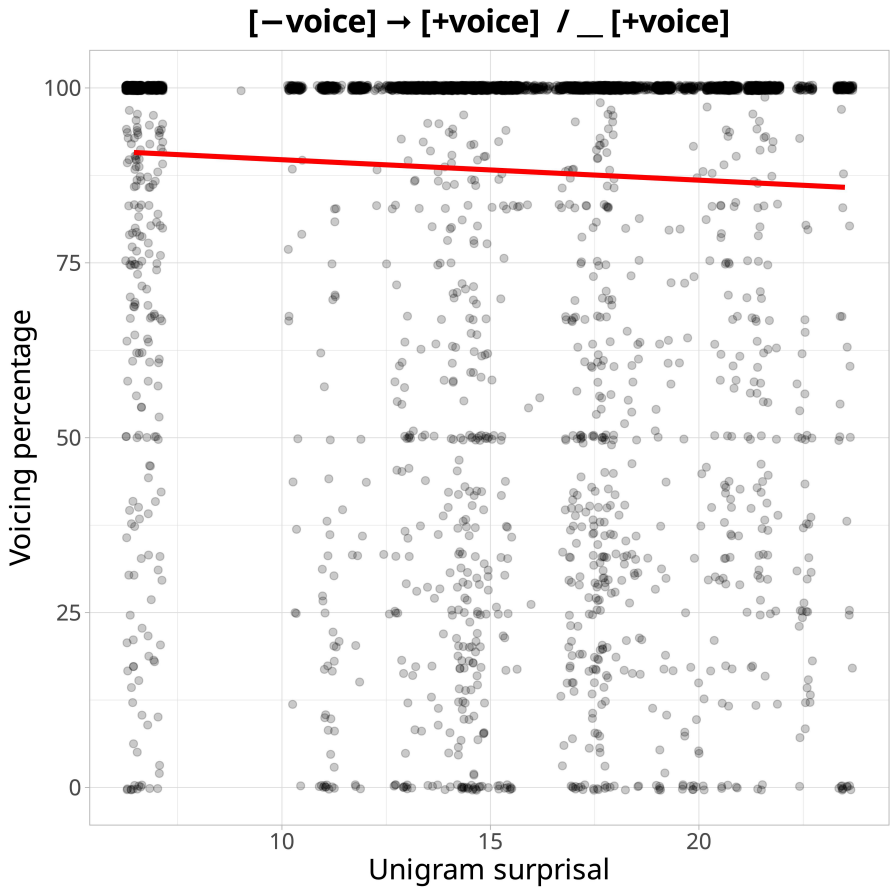

Figure 1 – Percentage of voicing in underlyingly voiceless obstruents preceding voiced obstruents. The red regression line shows that voicing decreases as surprisal increases.

calculates the proportion of voicing within a given interval by identifying voiced and voiceless frames based on a fixed step rate (e.g. 5 ms). The script determines the overlap of the interval with the first and last voiced frames and sums the durations of fully voiced frames within the interval. The percentage of voicing is then calculated as the total voiced duration divided by the interval duration, multiplied by 100. This approach also accounts for partial frame overlaps at the boundaries of the interval.

Linear mixed effects models were fitted, with VOICING PERCENTAGE in the first obstruent as the response variable, UNIGRAM SURPRISAL as the fixed effect, and SPEAKER as a random effect. All statistical analyses and graphs were made in R [24].

3 Results

3.1 Voicing assimilation in voiceless obstruents (obstruent voicing)

The mean percentage of voicing in underlyingly voiceless obstruents that are followed by a voiced obstruent is 88.26%, which is evidence that a process of voicing takes place. Table 1 shows the results of the linear mixed effects model for the effect of surprisal on the voicing of underlyingly voiceless obstruents in a context before voiced obstruents. The intercept estimate of 92.36% indicates that the percentage of voicing at a hypothetical surprisal value of 0 is very high. Surprisal had a significant main effect: voicing is estimated to decrease by 0.30% with each point of increase in surprisal. A scatter plot of voicing against surprisal is presented in Figure 1. The fitted regression line also shows that voicing decreases as surprisal increases.

Table 1 – Effect of surprisal on the percentage of voicing in underlyingly voiceless obstruents preceding voiced obstruents. Model: VOICING ∼ SURPRISAL + (1|SPEAKER). $r^2 = 0.11$.

	Estimate	**SE**	**DF**	***t***	***p***
(Intercept)	92.36	1.35	1060	68.40	0.0000
Surprisal	−0.30	0.07	5072	−4.12	0.0000

3.2 Voicing assimilation in voiced obstruents (obstruent devoicing)

The mean percentage of voicing in underlyingly voiced obstruents that are followed by a voiceless obstruent is 27.56%, which points to devoicing. The results of the linear mixed effects model estimating the effect of surprisal on the voicing of underlyingly voiced obstruents preceding voiceless obstruents are reported in Table 2. The percentage of voicing is low at low surprisal levels, as evidenced by the intercept estimate of 26.29%. Surprisal had a significant main effect on voicing here as well. The estimated coefficient is positive (0.21), which means that, with the increase of surprisal, a greater percentage of voicing is retained in underlyingly voiced obstruents. In other words, obstruent devoicing decreases as surprisal increases. The effect is also observable in the regression line fitted to the scatter plot in Figure 2.

Table 2 – Effect of surprisal on the percentage of voicing in underlyingly voiced obstruents preceding voiceless obstruents. Model: VOICING ∼ SURPRISAL + (1|SPEAKER). $r^2 = 0.12$.

	Estimate	**SE**	**DF**	***t***	***p***
(Intercept)	26.29	1.30	448	20.28	0.0000
Surprisal	0.21	0.10	4409	2.14	0.0320

4 Discussion and conclusion

Neutralising phonological processes, such as various assimilations, have traditionally been framed in categorical, typically binary terms. For example, anticipatory obstruent voicing assimilation is often formalised with a rewrite rule of the type shown in (2) below, or as a spreading process, whereby a feature, such as [±voice], spreads from one segment to another that is adjacent at some level of representation.

$$[-\text{sonorant}] \rightarrow [\alpha\,\text{voice}] \,/\, \_\_ \begin{bmatrix} -\text{sonorant} \\ \alpha\,\text{voice} \end{bmatrix} \tag{2}$$

The results reported in this study have confirmed that, when followed by a voiced obstruent, Bulgarian voiceless obstruents do become largely voiced, maintaining voicing throughout 88.26% of their duration on average. Our results have also validated the opposite process, in which voiced obstruents are devoiced (retaining only 27.56% of voicing on average) before a voiceless obstruent. At the same time, our findings have unequivocally demonstrated that the variation in voicing that is present in assimilating obstruents is not merely random but significantly related to surprisal in the way we predicted at the outset. In words with higher surprisal values, we found significantly lower degrees of voicing in voiceless obstruents before voiced ones, as well as significantly less devoicing of voiced obstruents before voiceless ones. From a broader perspective, this demonstrates that, overall, there is significantly less obstruent voicing assimilation in unexpected words than in more predictable ones.

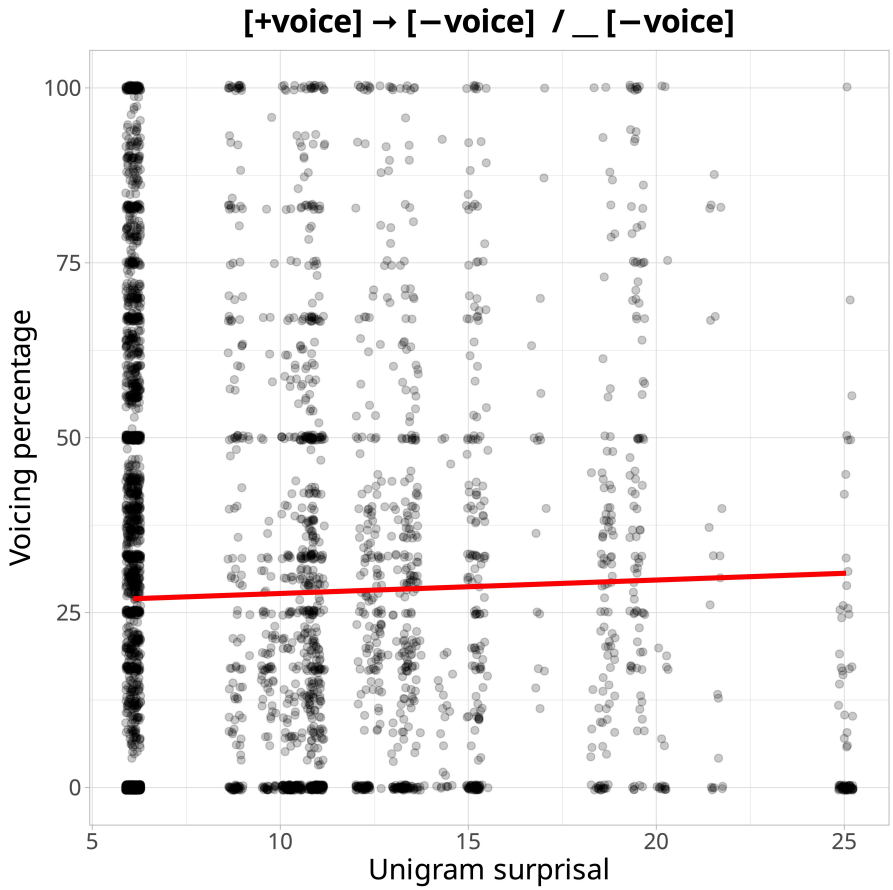

Figure 2 – Percentage of voicing in underlyingly voiced obstruents preceding voiceless obstruents. The red regression line shows that devoicing decreases (i.e., more voicing is retained) as surprisal increases.

The present study provides valuable insights into voicing assimilation in Bulgarian, but several limitations should be acknowledged. The research is confined to a specific set of phonological contexts and speech samples, which may not fully represent the variability found in other contexts or varieties of Bulgarian. Additionally, the use of controlled speech data might not fully capture the variability observed in spontaneous speech, where assimilation processes could differ. This study also does not explore the influence of word class, morpheme type, word stress, or accentuation patterns, all of which could significantly affect the realisation of voicing assimilation. Such lexical, grammatical, and prosodic factors remain important avenues for future research. Finally, no consideration has been given to potential sociolinguistic or cognitive effects on the assimilation processes examined. A more interdisciplinary approach addressing these dimensions could enhance the understanding of voicing assimilation in Bulgarian.

We can confirm that Bulgarian obstruents do assimilate to the voicing of a following obstruent and that this assimilation can be construed as categorical, insofar as obstruents in voicing environments are likely to be consistently perceived as voiced, and those in devoicing contexts as voiceless. Categorical assimilation alone, however, cannot capture the full complexity of the phenomena involved, as it co-occurs with gradient variation in voicing that is correlated with surprisal: assimilation is stronger in low-surprisal words, while in high-surprisal words speakers attempt to maintain the underlying [±voice] specification of an obstruent to a measurably higher degree. The findings presented in this article add to a growing body of research that demonstrates that processes once thought of as entirely categorical in fact exhibit gradient variation in fine phonetic detail, as speakers are aware of statistical patterns in language use, and respond to the predictability of linguistic items to maintain a balance between phonetic encoding and information density.

Acknowledgements

This research was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): Project ID 232722074, SFB 1102.

References

[1] PIERREHUMBERT, J. B.: *Exemplar dynamics: Word frequency, lenition and contrast*. In J. BYBEE and P. HOPPER (eds.), *Frequency and the Emergence of Linguistic Structure*. Benjamins, Amsterdam, 2001.

[2] JAEGER, T. F.: *Redundancy and reduction: Speakers manage syntactic information density*. *Cognitive Psychology*, 61(1), pp. 23–62, 2010. doi:10.1016/j.cogpsych.2010.02.002.

[3] AYLETT, M. and 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(1), pp. 31–56, 2004. doi:10.1177/00238309040470010201.

[4] AYLETT, M. and A. TURK: *Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei*. *The Journal of the Acoustical Society of America*, 119(5), pp. 3048–3058, 2006. doi:10.1121/1.2188331.

[5] MALISZ, Z., E. BRANDT, B. MÖBIUS, Y. M. OH, and B. ANDREEVA: *Dimensions of segmental variability: Interaction of prosody and surprisal in six languages*. *Frontiers in Communication*, 3, 2018. doi:10.3389/fcomm.2018.00025.

[6] IBRAHIM, O., I. YUEN, M. VAN OS, B. ANDREEVA, and B. MÖBIUS: *The combined effects of contextual predictability and noise on the acoustic realisation of German syllables*. *The Journal of the Acoustical Society of America*, 152(2), pp. 911–920, 2022. doi:10.1121/10.0013413.

[7] COHEN PRIVA, U.: *Informativity affects consonant duration and deletion rates*. *Laboratory Phonology*, 6(2), pp. 243–278, 2015. doi:10.1515/lp-2015-0008.

[8] HALE, J.: *A probabilistic Earley parser as a psycholinguistic model*. In *Proc. 2nd Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies*, pp. 1–8. Association for Computational Linguistics, Pittsburgh, PA, USA, 2001. doi:10.3115/1073336.1073357.

[9] LEVY, R.: *Expectation-based syntactic comprehension*. *Cognition*, 106(3), pp. 1126–1177, 2008. doi:10.1016/j.cognition.2007.05.006.

[10] LEVELT, W. J. and L. WHEELDON: *Do speakers have access to a mental syllabary? Cognition*, 50(1), pp. 239–269, 1994. doi:10.1016/0010-0277(94)90030-2.

[11] LEVELT, W. J. M.: *Producing spoken language: A blueprint of the speaker*. In *The neurocognition of language*, pp. 82–122. Oxford University Press, Oxford, 2000. doi:10.1093/acprof:oso/9780198507932.003.0004.

[12] CARREIRAS, M. and M. PEREA: *Naming pseudowords in spanish: Effects of syllable frequency*. *Brain and Language*, 90(1), pp. 393–400, 2004. doi:10.1016/j.bandl.2003.12.003.

[13] CHOLIN, J., W. J. LEVELT, and N. O. SCHILLER: *Effects of syllable frequency in speech production. Cognition*, 99(2), pp. 205–235, 2006. doi:10.1016/j.cognition.2005.01.009.

[14] LOSIEWICZ, B. L.: *The effect of frequency on linguistic morphology*. Ph.D. thesis, University of Texas at Austin, 1992.

[15] BYBEE, J. and J. SCHEIBMAN: *The effect of usage on degrees of constituency: The reduction of* don't *in English. Linguistics*, 37(4), pp. 575–596, 1999. doi:10.1515/ling.37.4.575.

[16] JURAFSKY, D., A. BELL, M. GREGORY, and W. D. RAYMOND: *Probabilistic relations between words: Evidence from reduction in lexical production*. In J. L. BYBEE and P. J. HOPPER (eds.), *Frequency and the Emergence of Linguistic Structure*, pp. 229–254. John Benjamins, Amsterdam, 2001. doi:10.1075/tsl.45.13jur.

[17] WHITESIDE, S. P. and R. VARLEY: *Dual-route phonetic encoding: Some acoustic evidence*. In *Proc. 5th International Conference on Spoken Language Processing*, pp. 3155–3158. Sydney, 1998. doi:10.21437/ICSLP.1998-812.

[18] CROOT, K. and K. RASTLE: *Is there a syllabary containing stored articulatory plans for speech production in English?* In *Proc. 10th Australian International Conference on Speech Science and Technology*, pp. 376–381. Sydney, 2004.

[19] BENNER, U., I. FLECHSIG, G. DOGIL, and B. MÖBIUS: *Coarticulatory resistance in a mental syllabary*. In *Proc. 16th ICPhS*, pp. 485–488. Saarbrücken, 2007.

[20] TILKOV, D., T. BOJADŽIEV, E. GEORGIEVA, J. PENČEV, and V. STANKOV: *Gramatika na săvremennija bălgarski knižoven ezik* [Grammar of contemporary Standard Bulgarian], vol. 1: *Fonetika* [Phonetics]. Bulgarian Academy of Science Press, Sofia, 1982.

[21] HATEVA, N., P. MITANKIN, and S. MIHOV: *BulPhonC: Bulgarian speech corpus for the development of ASR technology*. In *Proc. 10th International Conference on Language Resources & Evaluation (LREC) 2016*, pp. 771–774. 2016.

[22] MITANKIN, P., S. MIHOV, and T. TINCHEV: *Large vocabulary continuous speech recognition for Bulgarian*. In *Proc. Recent Advances in Natural Language Processing (RANLP) 2009*, pp. 246–250. 2009.

[23] BOERSMA, P. and D. WEENINK: *Praat: Doing phonetics by computer [Computer program], v. 6.4.17*. `http://www.praat.org/`, 2024.

[24] R CORE TEAM: *R: A language and environment for statistical computing, v. 4.4.1*. `https://www.R-project.org/`, 2024.