INTERSPEECH 2021
30 August – 3 September, 2021, Brno, Czechia

Phonetic distance and surprisal in multilingual priming: Evidence from Slavic

Jacek Kudera[1], *Philip Georgis*[1], *Bernd Möbius*[1], *Tania Avgustinova*[1], *Dietrich Klakow*[1,2]

[1]Language Science and Technology, Saarland University, Saarbrücken, Germany
[2]Spoken Language Systems (LSV), Saarland University, Saarbrücken, Germany
{kudera|pgeorgis|moebius|tania|klakow}@lst.uni-saarland.de

Abstract

This study reveals the relation between surprisal, phonetic distance, and latency based on a multilingual, short-term priming framework. Four Slavic languages (Bulgarian, Czech, Polish, and Russian) are investigated across two priming conditions: associative and phonetic priming, involving true cognates and near-homophones, respectively. This research is grounded in the methodology of information theory and proposes new methods for quantifying differences between meaningful lexical primes and targets for closely related languages. It also outlines the influence of phonetic distance between cognate and non-cognate pairs of primes and targets on response times in a cross-lingual lexical decision task. The experimental results show that phonetic distance moderates response times only in Polish and Czech, whereas the surprisal-based correspondence effect is an accurate predictor of latency for all tested languages. The information-theoretic approach of quantifying feature-based alternations between Slavic cognates and near-homophones appears to be a valid method for latency moderation in the auditory modality. The outcomes of this study suggest that the surprisal-based (un)expectedness of spoken stimuli is an accurate predictor of human performance in multilingual lexical decision tasks.

Index Terms: phonetic distance, information theory, surprisal, priming, Slavic languages

1. Introduction

Priming is a general property of human cognition that refers to a behavioral response after a sequence of stimuli which are related to each other within or across modalities. A recently experienced stimulus, i.e., a prime (Token 1), influences the way one responds to a target, that is, the later stimulus (Token 2). This is true not only in lexical access tasks, but also across various sensory modalities such as auditory, visual, and olfactory. The relation between the stimuli is conventionally reflected by means of the response time (RT) measured in behavioral tasks. This relation can cause a facilitation effect, namely shorter latency, or an inhibition effect, whereby the response to a target input is delayed due to a distant, unclear, or undiscovered relation between the perceived stimuli.

In terms of human language processing, the perceptual priming effect depends on lexical, syntactic, morphological, and phonetic relations between primes and targets. The relative frequency of occurrence of stimuli, as well as the knowledge of the person experiencing the sequence, also plays a role, and hence influences the reaction to the stimuli in so-called positive or negative priming. In a cross-linguistic context, the size of these effects corresponds with phonetic similarity and, more broadly, typological relatedness between the languages of primes and targets [1, 2, 3]. Studies on priming in the auditory modality have shown that correspondence between stimuli is especially relevant on the phonetic and phonological levels of speech processing [4, 5, 6]; however, most investigations thus far have been conducted on monolingual datasets, rarely taking a cross-lingual perspective [7, 8, 9].

To address this lacuna, this paper proposes new methods for quantifying relatedness and similarity between cognates and near-homophone tokens in a multilingual setting. The information-theoretic notion of surprisal [10], which measures the (un)expectedness of an outcome, is introduced and correlated with latency measures obtained from a lexical decision task in a short-term priming framework involving four Slavic languages: Bulgarian, Czech, Polish, and Russian.

1.1. Aims and hypotheses

The primary goal of this work is to present measures of cross-lingual speech comprehension based on phonetic distance and surprisal, and to validate them in a behavioral priming task. By grounding the perceptual study in information-theoretic methodology, the following assumptions will be addressed: (i) facilitating priming is present as an effect of exposure to multilingual tokens from closely-related languages; (ii) the priming type (cognate vs. phonetic) as well as the language of stimuli influences latency in a lexical decision paradigm; (iii) behavioral reaction, measured as response time, depends not only on the similarity between successive primes and targets but also on their (un)expectedness. Thus, the following hypotheses were tested: phonetic proximity of closely-related Slavic languages shortens the latency, regardless of the language of the prime; cognates shorten response times to a greater extent than do near-homophones in a multilingual experiment; and the information-theoretic approach, by introducing a unit of (un)expectedness, outperforms canonical measures of similarity between prime and target.

1.2. Related work

Previous studies on priming effects involving stimuli from several languages have shown that these effects depend on various linguistic levels. Semantic and etymological relatedness between tokens cause a cognate facilitation effect [11, 12]. Furthermore, discussions on latencies in lexical decision tasks have supported theories of associative activation during the process of searching for unique mental representations of stimuli. The strength of the facilitation effect can be asymmetric and reflect the subjects' dominant language, which might correlate with the size of one's lexicon or, as suggested by [13], differences in conceptual representations between a dominant language and an L2. In a multilingual scenario, RT was also applied as a direct measure of similarity across related languages, under the assumption that shorter response times reflect better intelligibility of subjects' L2 [14].

http://dx.doi.org/10.21437/Interspeech.2021-1003

2. Method

The study consisted of a lexical decision task in a priming paradigm using spoken stimuli, which were either cognate pairs or semantically-distant near-homophones. In the pre-test stage of the experiment, the relation between primes and targets was measured using phonetic distance and word adaptation surprisal. The degree of phonetic similarity was measured as a feature-based phonetic distance of aligned segments and their sequence within words. In the experiment, native speakers of four Slavic languages listened to primes (Token 1) in the three non-native languages and were asked to decide whether the targets (Token 2) were truly words in their native language. The targets in all cognate and near-homophone pairs were in the participants' L1, but were interspersed with filler word pairs (in a ratio of 4:1), in which the targets were not in the participants' L1 and were both phonetically and semantically distant from the primes. Each participant was exposed to 156 prime-target pairs. The subjects were instructed that their decision should be as accurate and fast as possible. Participants were paid for completing the study and additionally motivated by a bonus payment awarded for the best performance in each language group. On-screen visual feedback was presented to the participants in the form of a real-time plot of their accuracy and amount of time taken per task. The set of fillers was discarded from further analysis. The participants were presented with randomized pairs of primes and targets, and their decisions and response times were recorded.

2.1. Stimuli

The stimuli consisted of read-speech samples of true cognates and near-homophones identified by phonetic distances in Bulgarian, Czech, Polish, and Russian. The near-homophone pairs were automatically identified from transcribed wordlists on the basis of small pairwise phonetic distances. The cognate stimuli were used in the associative priming condition, whereas the near-homophone stimuli were used in the phonetic priming condition. The tokens, both primes and targets, controlled for frequency, were extracted from recordings of female native speakers of each language in self-paced reading sessions of token lists. The readings were recorded in an acoustically controlled environment with a 48 kHz sampling rate to uncompressed format. Two recording sessions were conducted for each native speaker.

2.2. Phonetic distance

The cross-lingual phonetic distance between primes and targets was calculated as the weighted sum of three component scores: dissimilarity of consonantal segments (0.5), dissimilarity of vocalic segments (0.3), and difference in syllable structure (0.2). Segments in word pairs were first aligned automatically using the Needleman-Wunsch algorithm [15], with alignment costs based on segment pairs' differences in distinctive features and sonority. The cosine similarity of distinctive feature vectors was then taken for each pair of aligned segments, with gap sequences (a segment aligned to nothing) and alignments of vowels to non-glide consonants receiving a similarity score of zero.

The overall similarity of consonant segments was calculated as the mean of cosine similarities of aligned consonants, and the similarity of vocalic segments was likewise calculated as the mean of the cosine similarities of the aligned vowels. Each of these similarity values was then subtracted from the maximum similarity of 1.0 to yield a measure of dissimilarity.

The difference in syllable structure was quantified as the length-normalized Levenshtein distance of each word pair encoded as sequences of 'C' (consonant) and 'V' (vowel). For example, the Polish word *chłodny* /xwɔdnɨ/ 'cold' was rendered as CCVCCV, whereas its Bulgarian counterpart *хладен* /xładɛn/ was rendered as CCVCVC.

Hence, the distance between inter-lingual homophones exhibiting identical features and syllable structure was equal to zero, and the higher the phonetic distance, the more phonetically distinctive the primes and targets. A maximum phonetic distance of 1.0 would only be reached in the case of zero aligned segments and no overlap in syllable structure (e.g., a word consisting of a single consonant and a word consisting of a single vowel).

2.3. Word Adaptation Surprisal

The (un)expectedness of the phonetic form of the targets (Y) given their primes (X) was quantified by means of surprisal, measured in bits, given in Equation (1).

$$surprisal(Y|X) = -log_2 P(Y|X) \tag{1}$$

The surprisal measures were calculated for pairs of stimuli using a method based on word adaptation surprisal (WAS), which has been presented as a model of inter-comprehension among Slavic languages in orthographic texts [16]. As an adaptation to the spoken modality, word adaptation surprisal was calculated on the phonetic level using IPA transcriptions of stimuli. Probabilities of inter-lingual phonetic correspondences were extracted from 1030 automatically aligned pairs of true cognates in Bulgarian, Czech, Polish, and Russian. These correspondence probabilities were then used to calculate the length-normalized phonetic word-adaptation surprisal between prime and target word pairs used in this study, according to Equation (2):

$$WAS = \frac{1}{n}\sum_{i=1}^{n} -log_2 P(L1_i|L2_i) \tag{2}$$

where: $L1_i$ = i^{th} phone in native (decoder) language and $L2_i$ = i^{th} phone in foreign (stimulus) language. Hence, surprisal quantifies the informativity of cross-linguistic phoneme correspondences for aligned pairs of primes and targets.

2.4. Participants

In total, 200 participants (50 native speakers of each tested language) recruited via an online crowd-sourcing platform took part in the experiment. On the basis of a pre-test questionnaire, participants with diagnosed hearing disorders or a background in Slavistics were excluded from the analysis (less than 1% of participants). The analyzed dataset was then supplemented with additional experimental sessions to ensure equally balanced groups of participants for each language.

2.5. Analysis

Only correct answers, not including responses to fillers, were included in the analysis. The RT range (150 ms $<$ RT $<$ 1500 ms) was altered in accordance with studies measuring monolingual priming effects [17]. The typical floor was set to filter out accidental responses, whereas the ceiling defined the threshold of the short-term priming effect. Each prime–target pair was repeated three times over the course of the experimental session, and a participant's RT for a particular prime–target pair

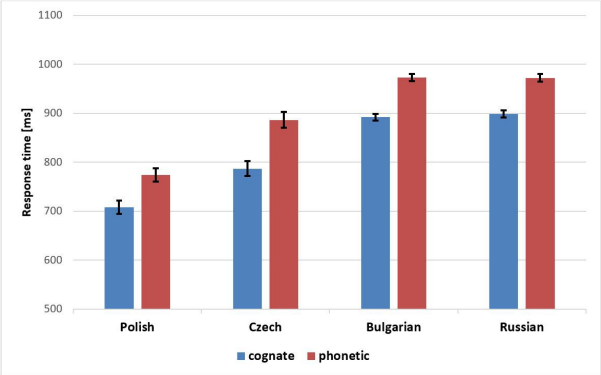

Figure 1: *Response times in cognate and phonetic priming according to participants' L1*

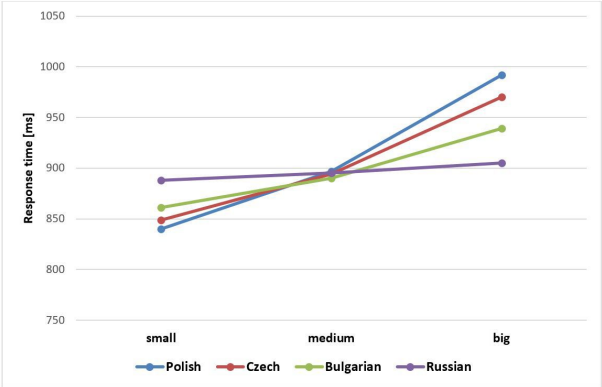

Figure 2: *Response times across three levels of phonetic distance*

was calculated as the mean of their three RTs for this pair. Basic descriptive statistics were calculated with the Kolmogorov-Smirnov test. A three-way ANOVA was applied for independent samples, and a Pearson correlation and moderation analysis was carried out to validate phonetic distance and surprisal – the methods proposed to quantify the cross-lingual correspondences.

3. Results

The Kolmogorov-Smirnov test showed that the distribution of mean results in both the pre-test and the post-test was significantly different from the normal distribution. However, both skewness and kurtosis were smaller than the absolute value of 2, so a parametric test could be performed. The descriptive statistics are presented in Table 1.

3.1. Priming type and language effect

In order to verify the differences in response times depending on priming type (cognate vs. phonetic) across languages, a three-way ANOVA for independent samples was performed. A statistically significant main effect for priming type was found $F(1, 4344) = 95.43; p < .001; \eta^2 = .02$. Hence, on the basis of the gathered data, the cognate facilitation effect was confirmed and response time was higher in the phonetic priming type.

The language of primes did not appear to affect the results. The Token 1 language main effect was not statistically significant, $F(3, 4344) = 0.59; p < .619; \eta^2 = 0$. Therefore, no post-hoc analysis was performed.

The four groups of subjects did not differ significantly in their measured RTs. Furthermore, no statistically significant interaction effect between priming type and Token 2 language was found: $F(3, 4344) = 0.49; p = .690; \eta^2 = 0$. On the other hand, a statistically significant Token 2 language effect was found: $F(1, 4344) = 122.68; p < .001; \eta^2 = .08$. A post-hoc analysis by means of Sidak tests was then performed. Response times in Polish were significantly lower than in the other three languages ($p < .001$). Response times in Czech were also significantly lower than in Bulgarian and Russian ($p < .001$). Bulgarian and Russian response times were not significantly different from one another. Response times across the four language groups are plotted in Figure 1.

Furthermore, a statistically significant interaction effect between priming type and Token 1 language type was found; $F(3, 4344) = 3.82; p = .01; \eta^2 = .003$. Therefore, a simple effect analysis was obligatorily performed. A statistically significant effect of the language of Token 1 was found, but only in the phonetic priming condition, $F(3, 4344) = 11.60; p < .001; \eta^2 = .008$. In total, three statistically significant differences in the post-hoc analysis were discovered. In the cognate priming condition, the Token 1 language effect was not statistically significant, $F(3, 4344)S = 2.13; p = .094; \eta^2 = .001$. On the other hand, a significant priming type effect in all languages was discovered - Polish: $F(3, 4344) = 67; p < .001; \eta^2 = .015$; Czech: $F(3, 4344) = 19.29; p < .001; \eta^2 = .004$; Bulgarian: $F(3, 4344) = 16.45; p < .001; \eta^2 = .004$ and Russian: $F(3, 4344) = 11.40; p = .001; \eta^2 = .003$. In line with the cognate facilitation effect, the RT values were higher in the phonetic priming type.

3.2. Phonetic distance and surprisal

In the following step, a Pearson correlation analysis was performed to find out whether phonetic distance and surprisal measures influence response times. A statistically significant correlation was discovered in both analyses. The phonetic distance measure was positively correlated with latency ($r = .190; p < .001$). As expected, larger phonetic distances were associated with longer response times. The measure of (un)expectedness via phonetic word adaptation surprisal was also positively correlated with RTs ($r = .219; p < .001$). Again, the higher the surprisal values – in other words, the higher the degree of unexpectedness – the more time subjects needed for their responses.

3.3. Moderation analysis

Subsequently, a moderation analysis was conducted to validate the quantification measures in the behavioral test. The phonetic distance and surprisal scores were clustered into three groups to verify their influence on the RT variable. It appeared that the latency moderation effect of phonetic distance is strong for Polish ($B = 946.34, SE = 90.42, t = 10.47, p < .001$) and Czech ($B = 750.38, SE = 91.03, t = 8.24, p < .001$), whereas no significant effect was found for Bulgarian or Russian. The results are plotted in Figure 2.

The information theory based measure of surprisal significantly moderates the RT in the data obtained from this study. It outperforms the phonetic distance measure and applies to all four tested languages: Polish ($B = 33.58, SE = 3.18, t =$

Table 1: *Lexical decision task results. Descriptive statistics*

	M	Me	SD	$Sk.$	$Kurt.$	$Min.$	$Max.$	D	p
Response time	902.1	888.68	227.15	0.14	0.07	151.63	1499	.03	$< .001$
Phonetic distance	0.08	0.07	0.06	0.36	-1.30	0	0.20	.16	$< .001$
Surprisal	2.34	1.64	1.90	1.25	1.04	0.05	8.91	.18	$< .001$

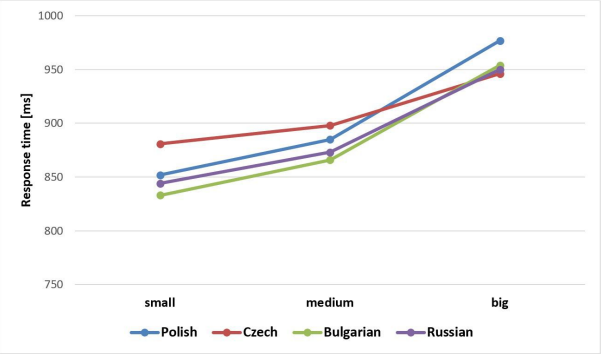

Figure 3: *Response times across three levels of surprisal*

$10.55, p < .001$), Czech ($B = 17.54, SE = 2.79, t = 6.28, p < .001$), Bulgarian ($B = 32.53, SE = 5.61, t = 5.79, p < .001$), and Russian ($B = 28.24, SE = 3.96, t = 7.14, p < .001$). The results are plotted in Figure 3.

4. Discussion

In this short-term priming study, two methods of quantifying phonetic similarities between cognates and near-homophones were tested in a multilingual lexical decision task experiment: one based on the degree of similarity in phonological features among corresponding phones; whereas the other, grounded in information theory, was based on the (un)expectedness of stimuli, measured in bits. These two methods were applied in order to quantify the relations among Bulgarian, Czech, Polish, and Russian targets and primes, with latency measures from the behavioral lexical decision task taken as a validation reference. The outcome of this study sheds light on the cognitive processing of linguistic relatedness and phonetic similarity of spoken stimuli in closely-related Slavic languages.

The first hypothesis, regarding the cognate facilitation effect, was supported in the multilingual priming experiment. It seems that the proximity of the Slavic languages is a crucial factor that enables an immediate recognition of cognates and thus promotes the facilitation effect. The discrepancies in the results among the different language groups, however, suggest asymmetric intelligibility effects. This finding could be further tested on a collection of more distantly related languages with more divergent phonemic inventories.

The second hypothesis was supported with respect to the effect of priming type. In line with expectations, it appears that cognates sharing a semantic field in the Slavic languages facilitate responses to a significantly greater extent than unrelated word pairs identified by their phonetic similarity. This outcome suggests that in lexical access tasks, a primary role should be directed towards the semantic relatedness rather than to similarity based on surface representations.

Thirdly, the information theoretic approach for presenting multilingual lexical relatedness was justified. With regard to the application of phonetic distance and (un)expectedness, the new proposals, and especially the information theoretic notion of surprisal, seem to be valid moderators of human response times after exposure to meaningful stimuli in a closely-related language. As a parallel to the positive priming effect, the influence of (un)expectedness of targets was found: the higher the surprisal, the longer the latency. This suggests that quantification of the cognate facilitation effect should be supplemented with a surprisal component, which corresponds to the (un)expectedness of stimuli.

5. Conclusions

The experimental results confirmed the perception of phonetic proximity of four closely-related languages from the Slavic branch of the Indo-European family, i.e., Bulgarian (South Slavic), Czech and Polish (West Slavic), and Russian (East Slavic).

It was reported that phonetic word adaptation surprisal and cross-lingual phonetic distance between primes and targets are moderators of latency in an auditory lexical decision task. The facilitating effect of cognate tokens clustered in the associative priming condition relates to the phenomenon of Slavic receptive multilingualism. This finding provides an argument for the primacy of associative correspondences and subjects' ability to identify the semantic relatedness of stimuli from another Slavic language. The participants' ability to immediately recognize the associative link between a non-native Slavic prime and a target in their Slavic L1 contributes to a strong inter-comprehension effect, even among individuals without any formal education or training in linguistics or Slavistics. Another outcome of this study turns the focus from lexical and phonetic similarity between tokens to the contextual (un)expectedness of stimuli.

The phase of method validation, conducted on four languages, provided evidence for context-based word adaptation surprisal outperforming the classical measures based on similarity between primes and targets in a short-term priming paradigm. Whereas the latency scores obtained from this study are moderated by the (un)expectedness of tokens and explain the effect in all four languages, the phonetic distance seems to moderate RTs only in Czech and Polish tokens. The results suggest that context-based methods for establishing the relation between two meaningful words in closely related languages are better predictors of human performance than metrics established exclusively on the basis of the similarity of stimuli.

6. Acknowledgements

Research funded by the Deutsche Forschungsgemeinschaft (German Research Foundation), Project ID 232722074 – SFB 1102.

7. References

[1] S. Joordens and S. Becker, "The long and short of semantic priming effects in lexical decision." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 23, no. 5, p. 1083, 1997.

[2] J. L. Nicol, "Syntactic priming," *Language and cognitive processes*, vol. 11, no. 6, pp. 675–680, 1996.

[3] P. A. Luce, S. D. Goldinger, E. T. Auer, and M. S. Vitevitch, "Phonetic priming, neighborhood activation, and parsyn," *Perception & psychophysics*, vol. 62, no. 3, pp. 615–625, 2000.

[4] L. M. Slowiaczek and D. B. Pisoni, "Effects of phonological similarity on priming in auditory lexical decision," *Memory & cognition*, vol. 14, no. 3, pp. 230–237, 1986.

[5] J. Sereno and A. Jongman, "Phonetic priming effects in auditory word recognition," *Working Papers of the Cornell Phonetics Laboratory*, vol. 7, pp. 151–177, 1992.

[6] K. Gor, "Phonological priming and the role of phonology in nonnative word recognition," *Bilingualism*, vol. 21, no. 3, p. 437, 2018.

[7] E. Lalor and K. Kirsner, "Cross-lingual transfer effects between english and italian cognates and noncognates," *International Journal of Bilingualism*, vol. 4, no. 3, pp. 385–398, 2000.

[8] W. Duyck, "Translation and associative priming with cross-lingual pseudohomophones: evidence for nonselective phonological activation in bilinguals." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 31, no. 6, p. 1340, 2005.

[9] A. Dijkstra, B. Hilberink-Schulpen, and W. J. Van Heuven, "Repetition and masked form priming within and between languages using word and nonword neighbors," 2010.

[10] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[11] A. Costa, A. Caramazza, and N. Sebastian-Galles, "The cognate facilitation effect: implications for models of lexical access." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 26, no. 5, p. 1283, 2000.

[12] D. Peeters, T. Dijkstra, and J. Grainger, "The representation and processing of identical cognates by late bilinguals: Rt and erp effects," *Journal of Memory and Language*, vol. 68, no. 4, pp. 315–332, 2013.

[13] T. Dijkstra and J. F. Kroll, "Bilingual visual word recognition and lexical access," *Handbook of bilingualism: Psycholinguistic approaches*, vol. 178, p. 201, 2005.

[14] C. Gooskens, "Experimental methods for measuring intelligibility of closely related language varieties," *The Oxford handbook of sociolinguistics*, vol. 195, p. 213, 2013.

[15] S. B. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of molecular biology*, vol. 48, no. 3, pp. 443–453, 1970.

[16] I. Stenger, T. Avgustinova, and R. Marti, "Levenshtein distance and word adaptation surprisal as methods of measuring mutual intelligibility in reading comprehension of slavic languages," in *Computational Linguistics and Intellectual Technologies: International Conference 'Dialogue 2017'Proceedings*, vol. 16, 2017, pp. 304–317.

[17] M. Perea and E. Rosa, "The effects of associative and semantic priming in the lexical decision task," *Psychological research*, vol. 66, no. 3, pp. 180–194, 2002.