



72 / 2024 / 1

Linguistica Brunensia





Linguistica Brunensia

72/2024/1



MASARYKOVA UNIVERZITA • UNIVERSITAS MASARYKIANA

BRNO 2024

Všechny zveřejněné stati prošly recenzním řízením; tematické uspořádání obsahu se opírá o systém použitý v *Bibliografii české lingvistiky* ([http: www.lib.cas.cz/casopisy/cz/Bibliografie_ceske_lingvistiky.htm](http://www.lib.cas.cz/casopisy/cz/Bibliografie_ceske_lingvistiky.htm)).

OBSAH – CONTENTS

ČLÁNKY – ARTICLES

Alexander Maxwell – Louise McMillan Error Bars for Lexicostatistical Estimates, With a Case Study Comparing the Diversity of Chinese and Romance	5
Lilla Nikolin Dukai Clitics in Old Serbian: What Does the Text of the Trojan Parable Tell Us?	23
Roman Sukač K takzvanému krácení vokálů v češtině	33
Petra Poukarová – Markéta Maturová „Protestovali před a v budově řeckého ministerstva financí.“ Krátké nahlédnutí do problematiky zeugmatu	47

RECENZE A ZPRÁVY – REVIEWS AND REPORTS

Bohumil Vykypěl Patricia de Bernardo Stempel: The Accents of Celtic: New Light on the Older and Oldest Stages	61
Bohumil Vykypěl Marta Eva Běťáková – Václav Blažek: Lexicon of Baltic Mythology	63
Bohumil Vykypěl Eduard Mühle: Slavs in the Middle Ages between Idea and Reality	65
Eduard Mühle: Slované ve středověku. Mezi ideou a skutečností	65
Arnold Suppan: 1000 Jahre Nachbarschaft. „Österreicher“, „Tschechen“ und „Sudetendeutsche“	65

Alexander Maxwell – Louise McMillan

ERROR BARS FOR LEXICOSTATISTICAL ESTIMATES, WITH A CASE STUDY COMPARING THE DIVERSITY OF CHINESE AND ROMANCE

ABSTRACT

This paper applies statistical techniques for measuring sampling error to lexicostatistics, a field in which error has often been discussed, but only rarely measured. We specifically calculate a margin of error for lexicostatistical comparisons based on Swadesh-type vocabulary lists, and use chi-squared tests to estimate a minimum threshold for when two lexicostatistical measurements will be statistically significantly different from one another. The article includes charts which mathematically unsophisticated scholars can easily use to check margins or error. We use margin of error calculations to test the claim that the relative internal diversity of Romance “languages” and Chinese “dialects” is equivalent, finding that no result is possible with extant lexicostatistical studies. We end by suggesting that lexicostatistical dendrograms depict uncertainty with “fat branches,” that is, branches whose width corresponds to statistical uncertainty.

KEYWORDS

lexicostatistics; margin of error; dendrograms; fat branches; Romance; Chinese

1 Introduction

This article proposes accounting for the uncertainty inherent in lexicostatistical comparisons with margins of error and significance thresholds. While the statistical theory we suggest is well-understood by mathematicians, we supplement our findings with reference charts, such that even scholars unfamiliar with statistical analyses can easily calculate a margin of error. Scholars wishing to indicate the

uncertainty in a single lexicostatistical measurement can look up the appropriate margin of error in the charts provided. Scholars wishing to compare two lexicostatistical measurements take the larger of the two measurements as the baseline and consult the appropriate chart to find the significance threshold. We then illustrate our technique with a case study: drawing on extant lexicostatistical studies, we test whether the relative internal diversity of Romance “languages” is comparable to the internal diversity of Chinese “dialects.” We end with suggestions for depicting statistical uncertainty in dendrograms.

Lexicostatistics is fundamentally a technique for quantifying linguistic similarity, which implies the quantification of linguistic distance. Quantitative measures of linguistic similarity/difference have many possible applications. Sheila EMBLETON (2015, 23) suggests lexicostatistical measurements can show “how closely two languages are related,” or alternatively “whether they are related at all.” Lexicostatisticians can use measurements of related varieties to construct dendrograms or “tree diagrams,” (GEISLER – LIST 2010; 2013); George STAROSTIN (2013, 126–27) once opined that “the main result of every lexicostatistical analysis is a phylogenetic tree (or network).” Lexicostatistical dendrograms in turn shed light on human prehistory, e.g. by revealing migration patterns (e.g. DYEN 1962; GRAY – ATKINSON 2003). Numerous scholars have proposed lexicostatistical techniques for sorting varieties into the categories “languages” and “dialects” (SWADESH 1954; WURM – LAYCOCK 1961; DYEN et al. 1992, 9; KORYAKOV 2017; WICHMANN 2020); or indeed into series of other collectives: families, stocks, phyla, and so forth (SWADESH 1954, 321; MCELHANON 1971, 134; CROWLEY 1992, 170). Missionaries affiliated with the Summer Institute of Linguistics (SIL) routinely gather lexicostatistical data to maximize the potential audience of Bible translations, using lexicostatistical thresholds as estimates of mutual intelligibility.

Most of these applications involve comparing several lexicostatistical measurements. The lexicostatistical similarity/distance between two varieties A and B holds little interest in isolation. Investigators are instead asking how the lexicostatistical similarity/distance between varieties A and B compares to that between varieties A and C, or between varieties C and D, and so forth. Which varieties are more closely related?

The centrality of comparison in lexicostatistical analysis raises the possibility of statistical error. If two lexicostatistical measurements are similar, how can we have confidence that a greater measurement actually implies a greater distance? Might one measurement merely appear greater as the result of a statistical fluke? Constructing dendrograms, furthermore, requires that scholars have great confidence in their data. Is such confidence justified? Understanding the many possible sources of error requires a brief summary of lexicostatistics as a technique.

The basic idea of lexicostatistics is to compare linguistic features and count similarities and differences. Any sort of linguistic feature can theoretically be com-

pared. In practice, however, most lexicostatistical studies examine word lists, because phonetic or grammatical features important in some varieties may not exist in others.

How do lexicostatisticians construct word lists? Historically, most have used the so-called “Swadesh lists,” developed in the 1950s by lexicostatistician and glottochronologist Morris Swadesh. Swadesh initially proposed a list of 225 words, but later developed shorter lists of 215 words, 200 words, and 100 words (SWADESH 1952; 1955; see also HYMES 1960, 3–5; OSWALT 1971, 421–34; MCMAHON – MCMAHON 2005, 34–44). Computational linguists Uri Tadmor, Martin Haspelmath and Bradley Taylor (TADMOR et al. 2010) subsequently introduced the “Leipzig-Jakarta list” of 100 words supposedly less resistant to borrowing. Russian sinologist Sergei Yakhontov, whose work is mostly known through the mediation of Sergei Starostin, chose a list of 35 words on the basis of “stability;” 32 of his words also appear in the Swadesh 100-word list (STAROSTIN 1995, 90–91; ZHURAVLEV 1994, 35–36). Aharun DOLGOPOL'SKY (1986, 620–628) used a list of 15 words he thought particularly resilient to borrowing. A few lexicostatisticians, finally, examine texts instead of word lists: Witold MAŃCZAK (2009), for example, compared Biblical passages.

The use of a finite word list inevitably creates sampling error, which can significantly affect lexicostatistical estimates (FELD – MAXWELL 2019). Comparing different wordlists, for example, leads to different lexicostatistical estimates. István Fodor's study of Slavic, to give a specific example, found a maximum 93% similarity between Czech and Polish using the Swadesh 100-word list and including possible synonyms, but a minimum of 88% similarity using the Swadesh 215 list and excluding possible synonyms (FODOR 1961, 303, 304). Over 22 measurements, Fodor found an average discrepancy between the two lists of 1.4%.

How do lexicostatisticians compare word lists, once created? There are many different techniques (EMBLETON 1986; HEGGARTY et al. 2011). Lexicostatisticians initially compared the percentage of shared cognates, also known as the “cognancy ratio.” Recognizing cognates is not an exact science: SERVA – PETRONI (2008) found it “often a matter of sensibility and personal knowledge” in which “subjectivity plays a relevant role.” Such subjective processes, obviously, generate statistical error. In a study of Bantu languages, for example, David OLMSTED (1957, 839–40) classified word pairs as “cognate,” “non-cognate,” or “ambiguous.” By including or excluding the ambiguous cases, Olmstead measured distances not as a single percentage, but as a range with a minimum and a maximum estimate. He found, for example, that Herero and Swahili shared somewhere between 29% and 41% of their vocabulary.

With the advent of computers, many lexicostatisticians have abandoned cognate recognition and employed instead the normalized Levenshtein distance. The Levenshtein distance, also known as the edit distance, based on an algorithm developed by Russian computer scientist Vladimir LEVENSHTAIN (1965; 1966), is the minimum number of changes to transform one line of characters into another.

It can be normalized to a percentage by dividing the edit distance by the number of characters in the longer word (MAGUIRE – McMAHON 2011, 108; WICHMANN 2020).

Applying the Levenshtein algorithm in practice still requires judgment calls. Greenhill noted, for example, that “differences in orthographies might play a crucial role in the accuracy of the Levenshtein distance.” Greenhill argued that “standardized orthography” increased the reliability of Levenshtein classifications from 41.3% “to around 65%,” suggesting that orthographic effects indeed have statistically significant consequences (GREENHILL 2011, 692). Several scholars transcribe their word lists into the international phonetic alphabet (IPA), which distinguishes over a hundred phonemes, most commonly through a computer-friendly version developed by John WELLS (1994). A team of scholars based in Leipzig, furthermore, collapse related sounds into a single symbol, using e.g. the same character for any “high and mid central vowel, rounded and unrounded (IPA ɪ, ɘ, ə, ɜ, ʉ, ɤ, ɵ)”. The Leipzig team has also done calculations based exclusively on consonants, equating e.g. the character strings “buk” and “bek” on the grounds that both share the pattern “b (vowel) k” (BROWN et al. 2008, 288, 289, 306, 307).

In short, lexicostatistics has many possible sources of error. Serious lexicostatisticians have extensively pondered the possible sources of lexicostatistical error. Scholars have also proposed many different techniques to make lexicostatistical data more robust (GUDSCHINSKY 1956; HYMES 1960; EMBLETON 1986; HEGGARTY 2010).

We suggest, however, that lexicostatisticians have not always paid sufficient attention to the *quantification* of error. SWADESH (1955, 124) admitted that lexicostatistics lacked “the accuracy of a precision instrument” and claimed for it only “considerable approximate validity.” Paul HEGGARTY (2010, 307) characterized lexicostatistics as “generally viable, if rather blunt.” Warren Maguire and April McMahon (MAGUIRE – McMAHON 2011, 16–17) judged Levenshtein calculations “rather crude,” but found them “effective for measuring the distance between varieties.” DYEN et al. (1992, 8) declared that various statistical uncertainties would “change the percentages only slightly,” while PERELTSVAIG – LEWIS (2015, 89) thought such uncertainties would “have a grave effect” that “cannot be ignored.” Lexicostatistical data have also been declared “approximately correct,” and “correct in a rough and ready sense (that is, on the average)” with “a certain rough and ready validity” (DOBSON et al. 1972, 207). We find these qualitative descriptions unsatisfactory. How approximate is approximate? How grave is “grave”? How rough is ready? Only when a source of error is quantified, we suggest, can we know whether or not it can be ignored.

In this discussion, we pay particular attention to quantifying the sampling error, as distinguished from measurement error. Measurement error concerns the comparison of two lexical items, and encompasses things like the difficulty of recognizing cognates, or the difficulty of standardizing orthography to apply the Lev-

enshtein algorithm. Sampling error arises from the selection of words to compare. One can theoretically avoid sampling error by comparing all possible items in the sample, but in practice gathering all words from two different varieties is impossible. No word list can pose as a complete lexicon, since counting the total number of words poses insurmountable difficulties of definition (KORNAI 2002). Furthermore, some words in one variety have no exact equivalents in other varieties.

Sampling error can be minimized. The normal procedure for reducing sampling error is to increase the sample size. Lexicostatisticians, however, fear that increasing the sample size would increase the measurement error, since it would increase the share of loanwords or neologisms. A longer word list that minimizes sampling error might therefore increase the total error. Lexicostatisticians have not found a solution to this dilemma.

If sampling error is inevitable, however, the question arises: how large is it? This paper provides an easy technique for estimating sampling error. We have calculated the margin of error for different thresholds of statistical reliability, and given the result in a series of charts so that scholars may visually display sampling error with error bars. Future scholars may have further ideas about how to quantify measurement error, or other sorts of error. Our primary aim, however, is to suggest that lexicostatisticians should not content themselves with efforts to minimize error. They should consider how to quantify it, and how to depict uncertainty in their results.

2 Estimating Lexicostatistical Sampling Error

We are not the first scholars attempting to quantify lexicostatistical sampling error. Taking the advice of “A. T. James of Yale University, a mathematical statistician,” the prolific lexicostatistician Isidore DYEN (1962, 42) applied “a combined Chi-squared test at the 5% level” to pairs of lexicostatistical measurements, finding that “a difference of about 10% between two percentages is necessary to produce a satisfactory result.” He warned that the 10% threshold could only “be regarded as a rule-of-thumb.” Indeed, in his subsequent work, DYEN (1975, 113) proposed “as a rule of thumb” that lexicostatistical percentages were only “significantly different if their difference is 9.5% or greater.”

Dyen’s “rule of thumb” method, however, is not a satisfactory method for estimating statistical significance in lexicostatistics. Indeed, Dyen himself noted in a footnote that his rule of thumb is not a substitute for the chi-squared test. The main problem is that the margin of error for a given percentage measurement varies: the error differs significantly depending on how close that percentage is to 50%.

We can illustrate the change in error with a pair of examples. Let us say that we have three varieties, A, B and C and that we wish to compare the margin of error at

the 95% confidence level. For this measurement, the margin of error can be calculated using the standard formula for proportions: $\pm 1.96 \times \sqrt{\hat{p}(1 - \hat{p}) / n}$, where \hat{p} is the estimated percentage, and n is the sample size. If we compare A and B using the Swadesh 100-word list and find that if 95% of the vocabulary is cognate, then \hat{p} is 95%, and n is 100, so the margin of error will be $\pm 1.96 \times \sqrt{.95(1 - .95) / 100}$, = 4.3%. But if we then compare A and C using the Swadesh 100-word list and find that 85% of vocabulary is cognate, then the margin of error will be 7.0%.

Now let us consider the process of comparing pairs of measurements, to determine whether they are significantly different from one another. For the A–B and A–C cognate measurements, we can compare them using the chi-squared test, as Dyen proposed, by treating them as a contingency table.

	Variety B	Variety C
Cognates with Variety A (%)	95	85
Non-cognate with Variety A (%)	5	15

We can apply the chi-squared test to this table using standard statistical software. The test rests on the assumption that the true proportions are the same in both columns and that any differences we see are just random variation in the observed values. In this case, the initial assumption for the test would be that Variety B and Variety C are both 90% cognate to Variety A, and that the observed 95% and 85% measurements are just due to random variation in the data. The test compares the observed values to what we would expect under that assumption. If the observed values are sufficiently different from that expectation, we conclude that the difference observed cannot actually be due to random variation and there is some underlying difference in the two measurements. In this case, if we apply the chi-squared test at the 5% significance level (which is equivalent to 95% confidence for the margin of error), we find that the observed values are unlikely to be due to chance, and that there is a statistically significant difference between how close A is to B and how close A is to C. For this case, Dyen’s rule of thumb holds: a difference of 10% between the two measurements is statistically significant.

Now consider a further pair of varieties, D and E. We compare A and D using the Swadesh 100-word list and find that 55% of the vocabulary is cognate; we then compare A and E using the Swadesh 100-word list and find that 45% of the vocabulary is cognate. Using a similar contingency table and chi-squared test as before, we find that at the 5% level there is no significant difference between how close A is to D and how close to A is to E, even though these two measurements are also 10% apart. In this case, Dyen’s rule of thumb fails.

Both of these situations can be illustrated with concrete examples by substituting A, B, C, D with some sample Indo-European varieties. Using the Swadesh

100-word list and excluding all doubtful cases, FODOR (1961, 301) found that Czech and Slovak share 95% of their vocabulary. With 95% confidence, we can say that a plausible value for the actual percentage of cognates for Czech and Slovak is any value in the range between 90.5% and 99.5%. Using the same method, FODOR (1961, 302) also found that Russian and Czech shared 85% of their vocabulary, so plausible values for the actual percentage of cognates for Russian and Czech would fall in a range between 78% and 92%. Fodor found a smaller distance between Czech and Slovak than between Czech and Russian, but applying the chi-squared test shows with 95% confidence that the smaller distance is actually statistically significant. DYEN et al. (1992, 107), meanwhile, measured the similarity between Pennsylvania Dutch and English as 55.3%, between Pennsylvania Dutch and Swedish as 45.3%. In this case, the smaller difference is not statistically significant.

It is feasible to calculate the smallest difference between A-D and A-E cognancy measurements that would be statistically different. For each value of A-D, this minimum significant distance defines a threshold value at which A-E becomes significantly different. The threshold values depend on the confidence ratio and the sample size. We have therefore prepared different charts for five different confidence ratios. We have also prepared charts for two different sample sizes, 100 and 200, facilitating measurements made using the Swadesh 200-word list, the Swadesh 100-word list, and the Leipzig-Jakarta list (which also has 100 items).

We have designed our charts for scholars without a deep understanding of statistics. For a margin of error chart, scholars take the measurement they are assessing, and read down the leftmost column to find the appropriate row. Select the chart value closest to the lexicostatistical measurement in question. An approximate margin of error is given.

For the significance threshold chart, scholars take the larger of the two measurements as the baseline and read down the leftmost column to find the appropriate row. If the higher measurement is x , any value lower than y will be significantly different from x . For entries in the significance thresholds table marked as “no significant differences,” which often occur at the bottom of the chart, the larger measurement is small and the smaller measurement therefore even smaller. In such cases, the two measurements will never be significantly different from each other.

Our charts enable scholars to find a margin of error for five different levels of statistical reliability: 5% significance, the two-sigma standard, the three-sigma standard, the four-sigma standard and, most stringently, the five-sigma standard physicists use as the standard for “discovery claims” at the Large Hadron Collider (LYONS 2013). Some linguists draw inspiration from physics; indeed, CHOMSKY (1988, 172) once expressed the hope that linguistics “can be reduced to physics.” Lexicostatisticians who take physics as their model can consult the five-sigma column of the chart, but the results are not very encouraging. Dyen, Kruskal and Black (DYEN et al. 1992, 107), using the Swadesh 200-word list, measured the distance between

standard (Stockholm) Swedish and German as 69.5%%, and the distance between standard Swedish and Danish as 87.4%. At the five-sigma confidence level (which corresponds to a confidence level of 99.99997), the lower distance measurement would have to be lower than 65.9% to be statistically significant. At the five-sigma confidence level, in short Dyen, Kruskal and Black's lexicostatistical data cannot decide whether Swedish is closer to Danish or to German. The width of the error bars illustrates just how far lexicostatistical data fall short of the accuracy standards in particle physics.

Scholars prepared to accept a lower standard of significance will naturally enjoy smaller error bars. The 95% confidence level for margin of error, i.e. a 5% significance threshold, is popular in several academic disciplines. At this lower confidence level, the figures of DYEN et al. (1992, 107) suggest that the difference between standard Swedish and German is significantly larger than that between standard Swedish and Danish Swedish. Nevertheless, even at this less rigorous confidence level, the lower distance measurement must be lower than 79.7% to be significant. The measured distances from Swedish to (Riksmål) Norwegian is 84.2%, and from Swedish to Faroese as 80.0%. So even at the relatively undemanding 95% confidence level, Dyen, Kruskal and Black's data cannot say whether standard Swedish is closer to Danish, Norwegian, or Faroese.

3 Case study: Romance and Chinese

To illustrate a more complex use of this chart, consider whether the “dialects” of Chinese are more or less internally diverse than the Romance “language family.” The terms “dialect” and “language family,” of course, anticipate a particular answer: they imply a greater diversity among Romance and a greater similarity among Chinese. To avoid pre-judging the results, we will speak only of comparing “Chinese varieties” and “Romance varieties.” So, which are more diverse, Chinese varieties or Romance varieties?

Several scholars have pondered this naïve question, typically declaring Chinese and Romance varieties to be equivalently diverse. Chomsky, for example, has twice made the comparison in order to ridicule the language-dialect dichotomy. In a 1977 conversation with French scholar Mitsou Ronat, he reasoned as follows:

“Why is ‘Chinese’ called a language and the Romance languages, different languages? The reasons are political, not linguistic. On purely linguistic grounds, there would be no reason to say that Cantonese and Mandarin are dialects of one language while Italian and French are different languages.” (CHOMSKY 1977, 195; English translation from CHOMSKY 1979, 190).

A decade later, Chomsky's *Knowledge of Language: Its Nature, Origin and Use* contained a similar *reductio ad absurdum*:

We speak of Chinese as 'a language,' although the various 'Chinese dialects' are as diverse as the several Romance languages. We speak of Dutch and German as two separate languages, although some dialects of German are very close to dialects that we call 'Dutch' and not mutually intelligible with others that we call 'German.' That any coherent account can be given of 'language' in this sense is doubtful" CHOMSKY 1986, 27).

Since Chomsky twice invoked the Chinese-Romance example, it evidently played some role in his thinking. In neither passage, however, did he adduce any actual evidence: his argument rests on simple assertion. Chomsky provided neither measurements of "diversity," nor a definition of how it might be measured. Nevertheless, we can attempt to fact-check Chomsky's assertion using extant lexicostatistical evidence.

To the best of our knowledge, no single lexicostatistical study has ever included both Chinese varieties and Romance varieties. Nevertheless, several studies have analyzed Chinese and Romance using the Swadesh lists. Indeed, for both Romance and Chinese, we have found studies using both the Swadesh 100-word list and the Swadesh 200-word list.

Yude WANG (1960, 91, 103) used the Swadesh 200-word list to calculate lexicostatistical distances for five varieties of Chinese, Mandarin, Wu, Min, Yüeh, and Hakka. Wang specifically proposed that "the dialects of Peking [Beijing 北京] (=P), Su-chou [Suzhou 苏州] (=S), Amoy [Xiamen 厦门] (=A), Canton [Guangzhou 廣州] (=C) and Moiyan [Meixian 梅州] (=M) can be considered respectively as a standard dialect" capable of meaningful comparison. Aware that cognate recognition is not an exact science, Wang provided a lower and upper bound of similarity by both excluding and including doubtful cognates. About 30 years later, XU (1991, 422) used the Swadesh 100-word list to calculate lexicostatistical distances for seven varieties of Chinese; his 21 distance measurements include 10 that correspond to Wang's. Xu did not acknowledge any doubtful cases, his figures are straightforward integer percentages. Table 1 compares Wang's lower and upper measurements with Xu's measurements between five selected Chinese varieties.

Romance languages have also attracted lexicostatistical attention. John REA (1958) used the Swadesh 100-word list to calculate 27 selected differences between eight Romance varieties: Catalan, French, Italian, Portuguese, Sardinian, Spanish, Rhaeto-Romance and Romanian. His 27 measurements do not actually cover all 28 total possibilities; he neglected to calculate a distance between Rhaeto-Romance and Catalan. Isidore Dyen, Joseph Kruskal, and Paul Black (DYEN et al. 1992, 19, 33, 103), using the Swadesh 200-word list, also included Romance varieties in a book-length study comparing no less than 84 Indo-European varieties, giving results for all 3,486 possible pairs. Rea listed simple integer percentages; Dyen, Kruskal and

Black give percentages with one digit after the decimal point. Table 2 summarizes their measurements for five selected Romance varieties: Portuguese, Spanish, French, Italian and Romanian.

Table 1. Lexicostatistical similarity between selected Chinese varieties

	WANG (lower-upper) Swadesh 200 (1960)	XU Swadesh 100 (1991)
Xiamen-Beijing	48.88 – 51.56	56
Xiamen-Suzhou	51.40 – 54.12	59
Xiamen-Guangzhou	55.31 – 56.77	63
Xiamen-Meixian	58.56 – 59.90	58
Meixian-Suzhou	63.10 – 64.43	73
Meixian-Beijing	63.78 – 65.10	69
Meixian-Guangzhou	69.70 – 70.53	79
Guangzhou-Beijing	70.16 – 70.77	74
Guangzhou-Suzhou	70.27 – 71.05	77
Beijing-Suzhou	72.73 – 73.47	73
Average distance between Chinese varieties	62.4 – 63.8	68.1

Table 2. Lexicostatistical similarity between selected Romance varieties

	DYEN et al. Swadesh 200 (1992)	REA Swadesh 100 (1958)
French-Spanish	73.4	75
French-Italian	80.3	89
French-Romanian	57.9	75
French-Portuguese	70.9	75
Italian-Portuguese	77.3	85
Italian-Spanish	78.8	82
Italian-Romanian	66	77
Spanish-Romanian	59.4	71
Spanish-Portuguese	87.7	89
Romanian-Portuguese	62.9	72
Average distance between Romance varieties	71.4	79.0

The various distances, of course, permit cherry picking: the lexicostatistical difference Xu measured between Guangzhou and Suzhou (77%) is identical to the distance Rea measured between Italian and Romanian. On the other hand, the lexicostatistical difference Xu measured between Xiamen and Beijing (56%) differs dramatically from the distance Rea measured between Spanish and Portuguese (89%). Chomsky's argument, however, adduced the internal diversity of Romance and Chinese as a whole. We suggest that the average distances given at the bottom of the charts are more relevant than any individual measurements.

Do Tables 1 and 2 support Chomsky's assertion? At first glance, the average distances appear to contradict Chomsky's claims in a way that might actually have strengthened Chomsky's argument. Chomsky assumed that the diversity between Romance varieties and between Chinese varieties was essentially the same, concluding that while consensus opinion among linguists views Romance varieties as "languages" and Chinese varieties as "dialects," linguists actually ought to classify them identically. According to the lexicostatistical measurements shown in Figures 1 and 2, however, Chinese "dialects" are actually *more* different from each other than the Romance "languages." The scholarly consensus is at even greater variance with lexicostatistical measurements than Chomsky assumed: if anything, linguists ought to speak about Romance "dialects" and Chinese "languages."

All lexicostatistical measurements, however, are subject to sampling error. Consider Rea and Xu, who both used the Swadesh 100-word list. The average of Rea's Romance measurements is 79%, and the average of Xu's Chinese measurements is 68%. Taking 79% as the higher measurement and rounding up, we see from the chart that at the five-sigma confidence threshold, the difference between Xu's measurements and Rea's measurements would be statistically significant only if Xu's average measurement were below 44.7%. Even if we satisfy ourselves with a two-sigma confidence level, the statistically significant difference appears only when Xu's average measurement is 66.5% or lower. Put another way, the significance threshold at the five-sigma confidence level is 36.3 percentage points, and the significance threshold at the two-sigma confidence level is 14.5 percentage points. The difference between Rea and Xu's measurements is below the significance threshold at either level of confidence. Thus there is no real way to say whether Chinese varieties are more or less diverse than Romance varieties.

4 Fat Branches: Depicting Error in Lexicostatistical Dendrograms

Double-digit significance thresholds for lexicostatistical differences raise questions about the validity of lexicostatistical dendrograms (tree diagrams). Lexicostatisticians who construct dendrograms have admittedly already taken steps to

highlight statistical uncertainty. In their 2003 article, for example, Gray and Atkinson expressed confidence percentages on the branches of their dendrogram (GRAY – ATKINSON 2003). Feld and Maxwell have proposed three different methods for signaling statistical insignificance on dendrograms (FELD – MAXWELL 2019, 114).

Scholars presenting dendrograms nevertheless rely too heavily on written caveats in explanatory text. NICHOLLS – GRAY (2006, 168), for example, presented dendrograms with the caveat that “inferring a single tree will be misleading,” since “there will always be uncertainty in the topology and branch lengths.” They do not, however, depict error in their dendrograms themselves. They present one of their dendrograms with an explanatory caption warning that it “does not constitute our ‘result’.” Readers of scholarly journals, however, do not always read captions: publishing a dendrogram runs an inevitable risk that readers will mistake it for a result. We are also unsure why Nicholls and Gray would choose to publish a dendrogram that they did not consider a result.

Our main observation, however, is that even error-sensitive lexicostatisticians tend to depict the branches of dendrograms as widthless lines. The inevitability of error, we suggest, calls for visual representation. We suggest that lexicostatistical error can be depicted by adding width to dendrogram branches. The greater the error, the fatter the branch. Drawing on Dyen, Kruskal, and Black’s lexicostatistical data for Indo-European languages, we have constructed some sample dendrograms with fat branches.

To illustrate the idea at its most basic, Figure 1 shows the divergence between German and Italian at three levels of confidence. Dyen, Kruskal and Black measured the cognancy ratio between German and Italian as 26.5% (DYEN et al. 1992, 106). Using the charts for a single lexicostatistical measurement, the chart depicts an error bar at the 95% confidence level, the three-sigma confidence level, and the five-sigma confidence level. The width of the branches is the width of the margin of error: the higher the confidence level, the fatter the branches.

Fat branching illustrates the advantages of a larger sample size. For the German/Italian measurement, Dyen, Kruskal and Black used the Swadesh 200-word list, and the error bar at the five-sigma confidence level is ± 15.0 . Had Dyen, Kruskal and Black calculated their measurements from the Swadesh 100-word list instead, the margin of error at five-sigma confidence would have been $\pm 21.7\%$, nearly a quarter of the total chart.

Figure 1 only depicts sampling error: it does not account for any error arising from the processes of cognate recognition, Levenshtein orthographic transcription, and so forth. Indeed, the chart may underestimate sampling error. Dyen, who from the team Dyen, Kruskal and Black took responsibility for cognate recognition, excluded lexical items in any case “when a single dialect offers two different words for one meaning and these words are members of different cognate sets.” While Dyen, Kruskal and Black assured readers that “this was not common,” they did not

quantify how common or uncommon, nor indicate which varieties were affected (DYEN et al. 1992, 20). If either German or Italian were affected, then the sample was less than 200 items and our chart slightly under-estimates the sampling error.

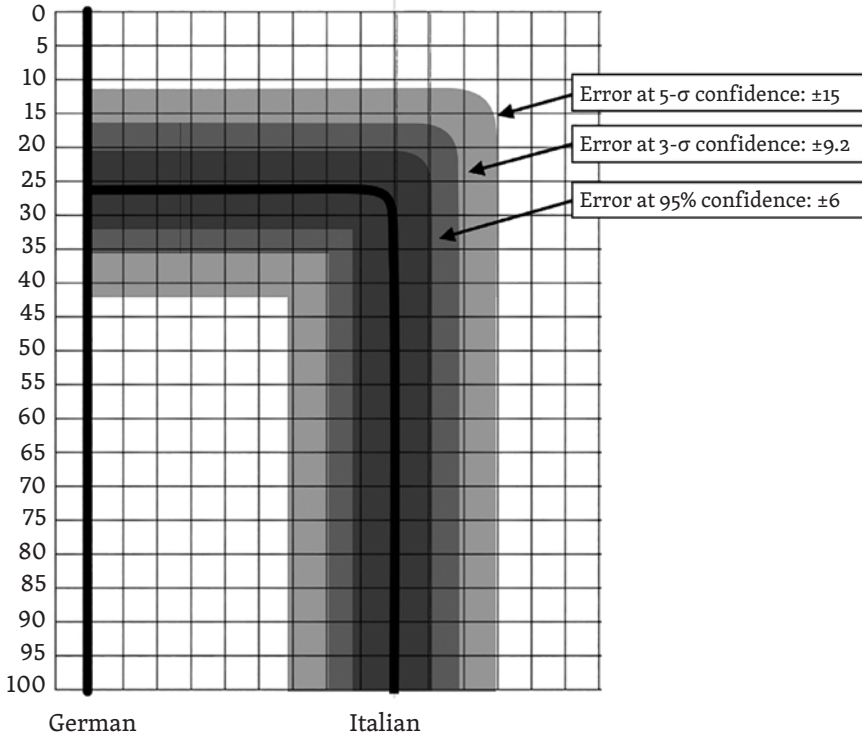


Figure 1.

Figure 2 displays five Slavic varieties as measured from Russian. Dyen, Kruskal and Black measured the cognancy ratio between Russian and Slovene, “Serbo-Croat,” Czech and Ukrainian as 61.4%, 67.5%, 74.5% and 77.9%, respectively (DYEN et al. 1992, 112). A traditional dendrogram with widthless lines is visible in white. Three different confidence levels are depicted with fat branches in three different shades of greyscale: darkest at the 95% confidence level, medium grey at the four-sigma confidence level of 99.997%, and light grey at the five-sigma confidence level of 99.9997%. At the 95% confidence level, all these varieties can just barely be distinguished. At four sigma, the fat branches begin to overlap: South Slavic can be distinguished from Northern Slavic, but the margin of error cannot separate Slovene from Serbo-Croat, nor Russian from Ukrainian, nor Russian from Czech. At five sigma, all Slavic varieties are indistinguishable.

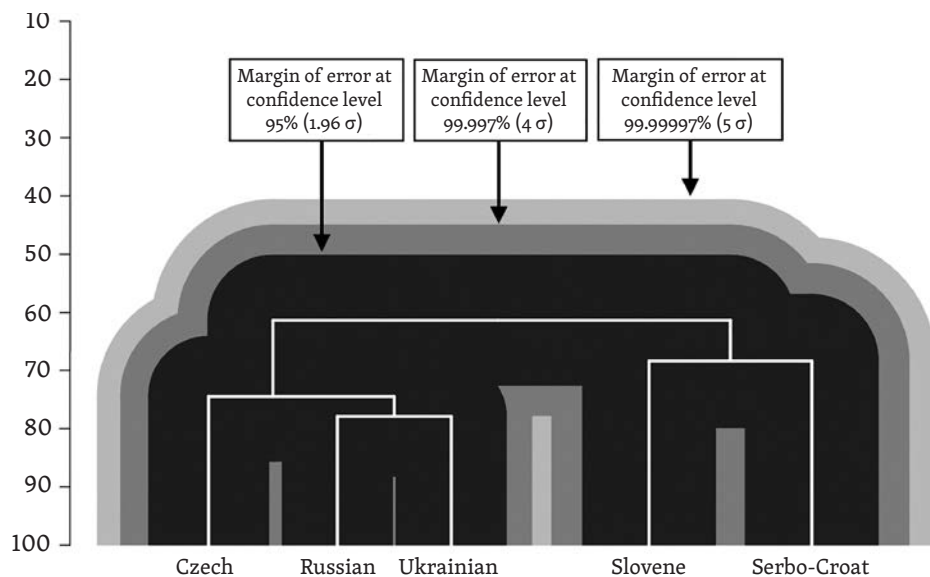


Figure 2.

Some readers may object that fat branches make the chart hard to read, since they conceal the relationship between the branches. We suggest, however, that the benefit of fat branches lies precisely in that they reveal no relationship when the margin of error is too large to draw meaningful conclusions. Fat branches indicate when lexicostatistical data are accurate enough to reveal a linguistic relationship and when they are not. The Slavic varieties are too closely related for lexicostatistical measurements to distinguish any clear relationships at the five-sigma confidence level; even at the three-sigma confidence level several Slavic varieties become indistinguishable. If the data are uncertain, then a lexicostatistical dendrogram should depict that uncertainty: presenting ambiguous results with false precision would be misleading. Fat branches prevent lexicostatisticians from becoming too enamored with imprecise data.

Other scholars may develop better techniques for the visual display of statistical uncertainty. Fat branches may have disadvantages we have not anticipated, perhaps some other technique would be better. Nevertheless, we conclude with the suggestion that dendrograms, and any other sort of diagram, should only depict clear results if the supporting data are statistically reliable. Statistically ambiguous data should produce an ambiguous diagram. Lexicostatisticians, and dialectometricians generally, should consider more carefully how to integrate error and statistical uncertainty into the visual representation of their results.

REFERENCES

- BROWN, Cecil – HOLMAN, Eric – WICHMANN, Søren – VELUPILLAI, Viveka. 2008. Automated classification of the world's languages: A description of the method and preliminary results. *Language Typology and Universals / Sprachtypologie und Universalienforschung* 61(4), pp. 285–308.
- CHOMSKY, Noam. 1977. *Dialogues avec Mitsou Ronat*. Paris: Flammarion.
- CHOMSKY, Noam. 1979. *Language and Responsibility: Based on Conversations with Mitsou Ronat*. New York: Pantheon.
- CHOMSKY, Noam. 1986. *Knowledge of Language: Its Nature, Origin and Use*. Westport: Praeger.
- CHOMSKY, Noam. 1988. *Language and Politics*. Montreal: Black Rose Books.
- CHRÉTIEN, C. Douglas. 1962. The mathematical models of glottochronology. *Language* 38, pp. 11–37.
- CROWLEY, Terry. 1992. *An Introduction to Historical Linguistics*. Oxford: Oxford University Press.
- DOBSON, Annette – KRUSKAL, Joseph – SANKOFF, David – SAVAGE, Leonard. 1972. The mathematics of glottochronology revisited. *Anthropological Linguistics* 14(6), pp. 205–212.
- DOLGOPOĽSKY, Aharon. 1986. A probabilistic hypothesis concerning the oldest relationships among the language families of Northern Eurasia. In: SHEVOROSHKIN, Vitaly – MARKEY, Thomas, eds. *Typology, Relationship and Time: A collection of Papers on Language Change and Relationship by Soviet Linguists*. Ann Arbor: Karoma, pp. 27–50.
- DYEN, Isidore. 1962. The lexicostatistical classification of the Malayopolynesian languages. *Language* 38(1), pp. 38–46.
- DYEN, Isidore. 1975. *Linguistic Subgrouping and Lexicostatistics*. The Hague: Mouton.
- DYEN, Isidore – KRUSKAL, Joseph B. – BLACK, Paul. 1992. *An Indoeuropean Classification: A Lexicostatistical Experiment*. Philadelphia: American Philosophical Society.
- EMBLETON, Sheila. 1986. *Statistics in Historical Linguistics*. Bochum: Brockmeyer.
- EMBLETON, Sheila. 2015. Historical linguistics: Numerical methods. In: WRIGHT, James, ed. *International Encyclopedia of the Social and Behavioral Sciences*. Oxford: Elsevier, pp. 23–26.
- FELD, Jan – MAXWELL, Alexander. 2019. Sampling error in lexicostatistical measurements: A Slavic case study. *Diachronica* 36(1), pp. 100–120.
- FODOR, István. 1961. The validity of glottochronology on the basis of the Slavonic languages. *Studia Slavica* 7, pp. 295–346.
- GEISLER, Hans – LIST, Johannes-Mattis. 2010. Beautiful trees on unstable ground: Notes on the data problem in lexicostatistics. In: HETTRICH, Heinrich, ed. *Die Ausbreitung des Indogermanischen: Thesen aus Sprachwissenschaft, Archäologie und Genetik*. Wiesbaden: Reichert, pp. 1–10.
- GEISLER, Hans – LIST, Johannes-Mattis. 2013. Do languages grow on trees? The tree metaphor in the history of linguistics. In: FANGERAU, Heiner – GEISLER, Hans – HALLING, Thorsten – MARTIN, William, eds. *Classification and Evolution in Biology, Linguistics, and the History of Science*. Stuttgart: Franz Steiner, pp. 111–124.
- GRAY, Russell D. – ATKINSON, Quentin D. 2003. Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, pp. 435–439.
- GREENHILL, Simon. 2011. Levenshtein distances fail to identify language relationships accurately. *Computational Linguistics* 37(4), pp. 689–698.

- GUDSCHINSKY, Sarah. 1956. The ABCs of lexicostatistics (glottochronology). *Word* 12(2), pp. 175–210.
- HEGGARTY, Paul. 2010. Beyond lexicostatistics: How to get more out of “word list” comparisons. *Diachronica* 27(2), pp. 301–324.
- HEGGARTY, Paul – McMAHON, April – McMAHON, Robert. 2011. From phonetic similarity to dialect classification: A principled approach. DELBECQUE, Nicole – AUWERA, Johan van der – GEERAERTS, Dirk, eds. *Perspectives on Variation: Sociolinguistic, Historical, Comparative*. Berlin: De Gruyter, pp. 43–92.
- HYMES, Dell. 1960. Lexicostatistics so far. *Current Anthropology* 1(1), pp. 3–44.
- KORNAI, András. 2002. How many words are there? *Glottometrics* 4, pp. 61–86.
- KORYAKOV, Yuri. 2017. Language vs. dialect: A lexicostatistic approach. *Voprosy Jazykoznanija* 6, pp. 79–101.
- LEVENSHEIN, V. I. 1965. Dvoichnye kody s ispravleniem vypadenij, vstavok i zameshhenij simbolov. *Doklady Akademii Nauk SSSR* 163(4), pp. 845–848.
- LEVENSHEIN, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady* 10(8), pp. 707–710.
- LYONS, Louis. 2013. Discovering the significance of 5 sigma. arXiv preprint, arXiv:1310.1284.
- MAGUIRE, Warren – McMAHON, April. 2011. Quantifying relations between dialects. In: MAGUIRE, Warren – McMAHON, April, eds. *Analysing Variation in English*. Cambridge: Cambridge University Press, pp. 93–120.
- MAŃCZAK, Witold. 2009. The original homeland of the Slavs. *Studia Mythologica Slavica* 12, pp. 135–145.
- McELHANON, Kenneth A. 1971. Classifying New Guinea languages. *Anthropos* 66(1–2), pp. 120–144.
- McMAHON, April – McMAHON, Robert. 2005. How do linguists classify languages? In: McMAHON, April – McMAHON, Robert, eds. *Language Classification by Numbers*. Oxford: Oxford University Press, pp. 20–49.
- NICHOLLS, Geoff – GRAY, Russell D. 2006. Quantifying uncertainty in a stochastic model of vocabulary evolution. In: FORSTER, Peter – RENFEW, Colin, eds. *Phylogenetic Methods and the Prehistory of Language*. Cambridge: McDonald Institute, pp. 161–172.
- OLMSTED, David. 1957. Three tests of glottochronological theory. *American Anthropologist* 59(5), pp. 839–842.
- OSWALT, Robert. 1971. Towards the construction of a standard lexicostatistic list. *Anthropological Linguistics* 13(9), pp. 421–434.
- PERELTSVAIG, Asya – LEWIS, Martin W. 2015. *The Indo-European Controversy: Facts and Fallacies in Historical Linguistics*. Cambridge: Cambridge University Press.
- REA, John. 1958. Concerning the validity of lexicostatistics. *International Journal of American Linguistics* 24(2), pp. 145–150.
- SERVA, Maurizio – PETRONI, Filippo. 2008. Indo-European languages tree by Levenshtein distance. *EPL (Europhysics Letters)* 81(6), 68005. URL: <https://doi.org/10.1209/0295-5075/81/68005>
- STAROSTIN, George. 2013. Lexicostatistics as a basis for language classification: Increasing the pros, reducing the cons. In: FANGERAU, Heiner – GEISLER, Hans – HALLING, Thorsten – MARTIN, William, eds. *Classification and Evolution in Biology, Linguistics and the History of Science: Concepts – Methods – Visualization*. Stuttgart: Franz Steiner, pp. 125–146.
- STAROSTIN, Sergei. 1995. *Altajskaja problema i proiskhozhenie japonskogo jazyka*. Moscow: Nauka: Glavnaja redakcija vostochnoj literatury

- SWADESH, Morris. 1952. Lexico-statistic dating of prehistoric ethnic contacts: With special reference to North American Indians and Eskimos. *Proceedings of the American philosophical society* 96(4), pp. 452–463.
- SWADESH, Morris. 1954. Perspectives and problems of Amerindian comparative linguistics. *Word* 10(2–3), pp. 306–332.
- SWADESH, Morris. 1955. Towards greater accuracy in lexicostatistic dating, *International Journal of American Linguistics* 21, pp. 121–137.
- SWADESH, Morris. 1972. What is glottochronology? In: SWADESH, Morris, *The Origin and Diversification of Language*. London: Routledge & Kegan Paul, pp. 271–284.
- TADMOR, Uri – HASPELMATH, Martin – TAYLOR, Bradley. 2010. Borrowability and the notion of basic vocabulary. *Diachronica* 27(2), pp. 226–246.
- WANG, Yude [Yu Te]. 1960. The lexicostatistic estimation of the time depths of the five main Chinese dialects, *Gengo Kenkyu: Journal of the Linguistic Society of Japan* 38, pp. 33–105.
- WELLS, John C. 1994. Computer-coding the IPA: A proposed extension of SAMPA, *Speech, Hearing and Language, Work in Progress* 8, pp. 271–289.
- WICHMANN, Søren. (2020). How to distinguish languages and dialects. *Computational Linguistics* 45(4). URL: https://doi.org/10.1162/coli_a_00366
- WURM, Stephen Adolphe – LAYCOCK, Donald C. 1961. The question of language and dialect in New Guinea. *Oceania* 32(2), pp. 128–143.
- XU, Tonquiang. (1991). *Lishi Yuyanxue*. Beijing: Shangwu Yinshuguan.
- ZHURAVLEV, Anatolij. 1994. *Leksikostatisticheskoe modelirovanie sistemy slavjanskogo jazykovogo rodstva*. Moscow: Indrik.

Alexander Maxwell

School of History, Philosophy, Political Science, and International Relations
Victoria University of Wellington
PO Box 600, Wellington 6140
New Zealand
alexander.maxwell@vuw.ac.nz

Louise McMillan

School of Mathematics and Statistics
Victoria University of Wellington
PO Box 600, Wellington 6140
New Zealand
louise.mcmillan@vuw.ac.nz



This work can be used in accordance with the Creative Commons BY-NC-ND 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). This does not apply to works or elements (such as image or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights.

Lilla Nikolin Dukai

CLITICS IN OLD SERBIAN: WHAT DOES THE TEXT OF THE TROYAN PARABLE TELL US?

ABSTRACT

The present paper examines the diachronic development of Serbian clitics. The investigation of clitics is of special interest in Slavic languages: despite the fact that these languages display free word order, the use of clitics is subject to strict rules. Clitics were present already in the Proto-Slavic language, however, in the course of the history of the language they were totally lost in Eastern Slavic and also partially in Western Slavic. The array and function of Serbian clitics has undergone substantial change: (1) similarly to Proto-Slavic, in Old Serbian manuscripts three types of enclitics were used, namely, discourse, pronominal and auxiliary clitics; and (2) clitics in present-day Serbian (PDS) tend to cluster and, within the clusters, their order is regulated by strict rules. This ordered position of clitics within the clusters has undergone changes. Moreover, there are significantly fewer clitic clusters in the investigated manuscripts than in PDS texts. The investigation of clitics is carried out in a 15th century Serbian manuscript, the Troyan Parable (Троянска притча) and in Old Serbian documents from the Middle Ages.

KEYWORDS

clitics; Old Serbian; language change

1 Introduction

The subject of this paper is the historical change or evolution of enclitics in Serbian. Research into enclitics in Slavic languages is very topical because these languages have a free word order, however, the order of enclitics in a sentence is strictly determined. The first part of the paper will describe the basic features of enclitics as a distinct grammatical category, while the second will focus on the enclitic system of modern

Serbian. In the third part, I will discuss the differences between the old and modern enclitic system, and the fourth part contains the analysis of the transitional state in the evolution of the enclitic system on the basis of a 15th century manuscript, the Trojan Parable. In the last part of the paper I will examine the differences between the enclitics in the Trojan Parable and in Serbian documents from the Middle Ages.

2 The basic characteristics of clitics

Enclitics are always unstressed language units that attach to a host next to them, i.e. the word with which they stand and with which they form a phonetic unit. It is important to distinguish enclitics from affixes. One of the main features that can determine whether a given linguistic unit is an affix or an enclitic is that clitics, unlike affixes, can stand next to any type of word¹, as well as the fact that they can be found in preceding (proclitics) or following the host (enclitics) (SPENCER – LOUÍS 2012, 21, 38). VA clitics are an exception, because they can stand only next to verbs.

One of the basic characteristics of enclitics is that they form clusters in sentences. In the cluster, each enclitic has its specific place regulated by syntactic rules (ZIMMERLING – KOSTA 2013, 180).

According to the functions they occupy in a sentence, enclitics can be assigned to different groups. An important feature of verbal enclitics is the expression of time, aspect, manner, person, and number (SPENCER – LOUÍS 2012, 34). Pronominal enclitics express person, number, gender, and case. In sentences they often function as objects and are then in a complementary distribution with explicit nouns. In such cases, they are semantically related to the verb (SPENCER – LOUÍS 2012, 43–44). However, this can be refuted in the case of clitic doubling. This phenomenon can be observed in Bulgarian and Macedonian, because these languages have VA clitics (FRANKS – KING 2000, 250–251). The third group consists of discourse enclitics which attach to the whole clause.

A very dominant syntactic feature of enclitics is that different rules apply to them than to the rest of the sentence. These rules may vary depending on the language.

Semantically, enclitics can be local or phrasal. Local enclitics attach to one word and complement its meaning, while phrasal enclitics attach to the whole clause and have a more general meaning (GYÖRFI 2020, 3).

2.1 Position of enclitics

Enclitics can occur in several places in a sentence. They usually take the second place in the sentence (2P clitics) or the position next to the verb (VA clitics). In Serbian they can occur after the first stressed word, i.e. in the so-called “Wackernagel”

¹ According to Progovac: „clitics cannot attach to the first word, unless that word is a syntactic constituent which can independently find itself in the position where clitics attach” (PROGOVAC 2005, 137).

position or after the first phrase (2W or 2D clitics). According to PAVLOVIĆ (2011, 34), the positioning of enclitics after the first stressed word or phrase can be explained by reasons of rhythm, because in that position, after the first stress in the sentence, the intonation falls and therefore the place is ideal for enclitic. If the enclitics are in the postfrontal position, i.e. at the end of the first X-bar group, they are called “Wackernagel” clitics after the Swiss linguist Wackernagel, who examined Greek enclitics in the 19th century, and who later identified this tendency in Indo-European languages. In his study of the Old Russian enclitic system, ZALIZNJAK (2008) calls these enclitics phrasal, because they refer to the whole clause. Phrasal enclitics are in many cases attach to the predicate. In other cases they can also appear after the first word of the phrase, even if it is not a predicate. For example the discourse clitic *li* behaves that way (ZALIZNJAK 2008, 24, 67). In the case when enclitics do not occur in the Wackernagel position, they are called local enclitics, and then they do not refer to the whole clause but only to the phrase in which their host is located. The distribution of local enclitics in a sentence depends on the place of the phrase to which they are related (ZALIZNJAK 2008, 25). According to Pavlović, however, these enclitics also take a postfrontal position, but only within those phrases that contain their host (PAVLOVIĆ 2011, 34).

There are enclitics that tend to move to the right in a sentence. According to Zaliznjak, the older the enclitic, the stronger it is. Thus, he divides enclitics into strong and weak enclitics, and the latter are those which can move from their positions to the right. These are pronominal enclitics in dative and accusative, and verbal auxiliaries of the verb *нечати* (ZALIZNJAK 2008, 51).

2.2 Classification of enclitics into sentence (S), noun-phrase (NP) and verb-phrase (VP) enclitics

Analyzing Old Serbian documents and letters from the 12th and 13th centuries, Pavlović divides enclitics into three groups: (1) sentence enclitics – *li* in Serbian; (2) noun-phrase (NP) enclitics – enclitics in the dative that mark possessives; and (3) verb-phrase (VP) enclitics – enclitics in the dative and accusative (and reflexive pronoun), verb enclitics (the present of the auxiliary verb *јесѣмь* in the perfective or in existential sense, the auxiliary verb *быиши* in the conditional, and the present tense form of the auxiliary verb *хѣтиши* denoting the future tense) (PAVLOVIĆ 2011, 37).

2 Enclitics in modern Serbian

In Serbian there are several types of enclitics which often form clusters in a sentence. In some syntactic positions clitics are not clustered, split is allowed. Examples include VP ellipsis, VP fronting, parentheticals (BOŠKOVIĆ 2001, 48, 50, 86; BOŠKOVIĆ 2016, 11–12). They behave unusually and, therefore, they are considered

special enclitics, meaning that the stressed form and the enclitic have different syntax (ZWIČKY 1977, 5–6). Special enclitics, unlike ordinary enclitics, are inflected and have their own paradigms (FRANKS 2009, 5). In modern Serbian, there are three types of enclitics: discourse, verbal and pronominal enclitics. The discourse clitic is *li*. Verbal enclitics are forms of the auxiliary verb *jesam* in the present tense, the auxiliary verb *biti* in the aorist, and the auxiliary verb *hteti* in the present, while pronominal enclitics occur in the dative, accusative and genitive cases, and also the reflexive pronoun *se*.

Table 1. The enclitic system in modern Serbian

Number	Person	Pronominal enclitics			Verbal enclitics		
		Dative	Accusative	Genitive	jesam	бити	хтети
Sg.	1	ми	ме	ме	сам	бих	ћу
	2	ти	те	те	си	би	ћеш
	3	му (m,n) joj (f)	га (m,n) ју, је (f)	га (m) је (f)	је	би	ће
Pl.	1	нам	нас	вас	смо	бисмо	ћемо
	2	вам	вас	вас	сте	бисте	ћете
	3	им	их	их	су	би	ће

3 Enclitics in Old Serbian

After the disintegration of the Old Slavonic language, changes occurred in the order of the enclitics in Serbian, and new forms appeared. According to Zaliznjak, the formation of new enclitics plays a great role in the order of enclitics, because older enclitics are positioned in front of younger enclitics in the sentence, looking from the left (ZALIZNJAK 2008, 27).

Table 2. The enclitic system in Old Serbian

Number	Person	Pronominal enclitics		Verbal enclitics		
		Dative	Accusative	БЫТИ	ИСТЬ	ХТѢТИ
Sg.	1	ми	ме	быхъ	съль	кю
	2	ти	те	бы	си	кѣш
	3	моу (m,n) юи (f)	га (m,n) ю (f)	бы	ѣ истъ	кѣ
	reflexive pronoun	си	се			

Number	Person	Pronominal enclitics		Verbal enclitics		
		Dative	Accusative	ВЫТИ	ЈЕСАМЪ	ХТѢТИ
Sg.						
Pl.	1	НЫ, НАМ	НЫ, НАС	ВЫ	СМО	КЕМО
	2	ВЫ, ВАМ	ВЫ, ВАС	ВЫСТЕ	СТЕ	КЕТЕ
	3	ИМЪ	ИХЪ	ВЫХОУВЫ	СОУ	КЕ

In Old Serbian, in contrast to the modern language, the enclitics occurred in the following order: (1) *li*, (2) the forms of the aorist of the auxiliary verb *biti*, (3) the pronoun enclitics in the dative and the accusative, (4) the reflexive pronoun *se*, and (5) the forms of the auxiliary verbs *jesam* and *hteti* (TOLSTAJA 1991, 202).

- (1) И ДА ИМЪ СЕ ПЛАКТА
 and that CL.DAT.3PL CL.refl.pron pay.PRS.3SG
 како ИМЪ СЪ И ПРЪВО
 how CL.DAT.3PL CL.beAUX.PRS.3PL too at.first
 ‘And so they would pay it to them, as they did the first (TOLSTAJA 1991, 202) time’

The main difference in the order is that the pronominal enclitics occur before the enclitic forms of the auxiliary verbs *jesam* and *hteti*, while in modern Serbian they follow these verbal enclitics.

4 Enclitics in the Trojan Parable

The corpus of the research was the Trojan Parable, a 15th century manuscript. The manuscript in which this novel is located is the Софијска илустрована Александрида. This manuscript is kept in the National Library of St. Cyril and Methodius in Sofia, under inventory number 771 (381).

In addition to the Trojan Parable, the book contains other texts. Namely, the manuscript begins the so-called Alexandrida, and contains other liturgical texts, while the Trojan Parable begins on page 220. The manuscript is written in two different persons’ handwriting: one scribe’s up to the Trojan Parable and another’s from the beginning of the Trojan Parable to the end of the manuscript. The language of this novel is Old Serbian, very close to the spoken language, and that is why it is important and interesting for research. The present paper investigates – as has already been mentioned – the distribution of enclitics in the Trojan Parable. The aim of the analysis is to demonstrate the differences in the order of clitic clusters. A new enclitic emerged and the old order in the cluster was disrupted. In the novel, the distribution of enclitics shows a transitional state. One can see the old order, and, in

addition to it, the new one with the presence of new enclitic forms. Examples with the old order from the novel:

- (2) СВЪКЛИ СЕ СМО СУДИ НАМЪ
undress.PTCP CL.refl.pron CL.be.AUX.PRS.1PL judge.IMP.2SG CL.DAT.1PL
'We undressed, judge us!' (227v)

In this example the reflexive pronoun *se* precedes the auxiliary clitic *smo*.

- (3) ЕРЪ ГДЕ ТЕ КЮ НАИТИ ЗНАИ
because where CL.ACC.2SG want.AUX.PST.1SG find.INF know.IMP.2SG
ЩО ТИ ХОКЮ ОУЧИНИТИ
what to.you.CL.DAT.2SG will.V.PST.1SG do.INF
'Because wherever I will find you, you should know what I will do.' (253r)

Here the enclitic in the accusative occurs in the first place and is followed by the verbal enclitic.

In these examples, the old order of enclitics can be observed, i.e. pronominal enclitics are positioned before the enclitic forms *jesam* and *hteti*.

With the appearance of new forms in the dative, a new order can already be noticed. These new forms receive a different status; thus the old order is disturbed. These enclitics are placed after verbal enclitics, as in contemporary Serbian.

- (4) КАКО^и ГА СЪ^и ВА^и ДАА
how CL.ACC.3SG CL.be.AUX.PST.1SG CL.DAT.2PL give.PTCP.SG
ЖИВА
alive.ADJ.GEN.SG
'How I gave him to you alive.' (264v)

In this example, the enclitic *ra* is still in the old place, while the new form *ва^и* occupies the new place after the auxiliary verb *съ^и*. However, the new order has not yet been established. There are examples in the manuscript in which we find new enclitics in the dative, but in the old position.

- (5) ГРЪСЦИ ОУРОВЕ СЪДА ВА^и СЪ^и
greek.ADJ.PL men.VOC.PL now CL.DAT.2PL CL.be.AUX.PST.1SG
СПОМЕНЪ^и
mention.PTCP.SG
'Greek men, I have just mentioned to you.' (264v)

In this example, the enclitic *вѣ* occurs in front of the verbal enclitic *чи*, so it behaves like the old form.

Researching old Serbian documents from the 14th and 15th centuries, Tolstaja had similar examples. She also concluded that with the appearance of new enclitic forms in the dative, fluctuations in the enclitic system can be observed, because the process of adopting new forms was not yet fully completed (TOLSTAJA 1991, 204).

5 Comparison of the enclitics in the Trojan Parable and in medieval old Serbian documents

Medieval Serbian documents provide an insight into the spoken language of that time and are therefore crucial for linguistic analysis.

In the old Serbian documents we can also see this phenomenon, namely, that when the new dative enclitics appear, the old order changes. In 13th century documents we observe the old order:

- (6) да вы е ѿтворена наша
 so CL.DAT.2PL CL.be.AUX.PST.3SG open.PTCP.SG POSS.PRON.1PL
 зѣмла
 country.NOM.SG
 ‘So our country is open to you.’ (STOJANOVIĆ 1929, 5)

- (7) дѣ их сѣ ѿтег
 so CL.DAT.3SG CL.be.AUX.PST.3PL kidnap.PTCP.SG
 ‘So they kidnapped them from him.’ (STOJANOVIĆ 1929, 32)

But in 15th century documents we can see the fluctuations, as in the Trojan Parable.

- (8) како си намь ѡбѣтоваль
 how CL.be.AUX.PST.2SG CL.DAT.1PL promise.PTCP.2SG
 ‘How you promised to us.’ (MLADENOVIĆ 2007, 67)

In this example the auxiliary clitic precedes the new form *nam* in the dative.

- (9) юрь пишеть да ю вамь рекль
 because write.PST.3SG that CL.be.AUX.PST.3SG CL.DAT.2PL tell.PTCP.3SG
 ‘Because he writes that he told you.’ (STOJANOVIĆ 1929, 407)

We can see that the new dative form *вамъ* follows the verbal enclitic.

In the Trojan Parable we can find the following number of enclitics:

Table 3. The number of enclitics in the Trojan Parable

Discourse enclitics	4
Enclitics in the dative	109
Enclitics in the accusative	138
Reflexive enclitics	86
Forms of <i>късьмь</i>	98
Forms of <i>бьтти</i>	21
Forms of <i>хтѣтти</i>	2
Clusters	28

Table 4. The number of clusters in the Trojan Parable

Clusters with the old order	15
Clusters with the new order	3
Clusters with mixed order (the new enclitic is in the new position, but other pronominal enclitics are in the old position)	1
Other clusters (without pronominal and verbal enclitics together)	13
All clusters	32

Table 5. The number of enclitics in medieval documents

Document	Discourse enclitic	Enclitics in the dative	Enclitics in the accusative	Reflexive enclitic	Forms of <i>рсьмь</i>	Forms of <i>бьтти</i>	Forms of <i>хтѣтти</i>	Clusters
Stojanović 8.	2	3	/	6	5	/	/	2
Stojanović 14.	5	24	6	7	10	/	/	7
Stojanović 32.	1	5	2	1	5	4	/	4
Stojanović 412.	/	3	/	1	1	1	/	1
Mladenović I/44.	/	19	1	2	8	/	/	2
Mladenović I/45.	/	31	3	9	19	4	/	6
All	8	85	12	26	48	9	0	22

Table 6. Clusters in medieval documents

Clusters with the old order	16
Clusters with the new order	1
Clusters with mixed order (the new enclitic is in the new position, but the old enclitics are in the old position)	/
Other clusters (without pronominal and verbal enclitics together)	9
All	26

From these data, we can see that pronominal enclitics are the most common in the analyzed manuscripts. There are many examples in the Dative that are phrasal enclitics with possessive semantics.

As observed, there are not many clusters: 28 in the Troyan Parable and 22 in the documents. This is an important feature of old Serbian, and we can see language change here, because in modern Serbian clusters have become more common and widespread.

6 Conclusion

In the Troyan Parable we can see a transitional state, where, with the appearance of new forms, the old order is disturbed. Medieval documents from the 15th century also testify to this: we have seen that in the older documents the order is still the old one, with the old forms in dative. From this we can establish that the changes in the enclitic system began in the 14th century.

This changes the ranking among enclitics and creates a new system that we can find in the modern language. As can be seen from the above examples, in this text the new order is not yet widespread, and, therefore, there is a fluctuation between the old and the new orders.

REFERENCES

- BOŠKOVIĆ, Željko. 2001. *On the Nature of the Syntax-Phonology Interface*. Oxford: Elsevier.
- BOŠKOVIĆ, Željko. 2004. Clitic placement in South Slavic. *Journal of Slavic Linguistics* 12(1), pp. 37-90.
- BOŠKOVIĆ, Željko. 2016. On second position clitics crosslinguistically. In: MARUŠIĆ, Lanko Franc - ŽAUCER, Rok, eds. *Formal Studies in Slovenian Syntax*. 23-54.
- FRANKS, Steven. 2009. Clitics in Slavic. In: KEMPGEN, Sebastian - KOSTA, Peter - BERGER, Tilman - GUTSCHMIDT, Karl, eds. *Die slavischen Sprachen. The Slavic Languages. An Interna-*

tional Handbook of Their Structure, Their History and Their Investigation. Berlin – New York: Walter de Gruyter.

FRANKS, Steven – HOLLOWAY KING, Tracy. 2000. *A Handbook of Slavic Clitics*. New York: Oxford University Press.

GYÖRFI, Beáta. 2020. A big change starts small: Pronominal clitics in 12–15th century Old Russian Chronicles. *SLAVICA* 49, pp. 8–21.

MILUTINOVIĆ, Milan – KARAANGOV, Petar. 1987. *Sofijska ilustrovana Aleksandrida*. Beograd – Sofia: Narodna biblioteka Srbije – Narodna biblioteka “Kiril i Metodij”.

MLADENOVIĆ, Aleksandar. 2007. *Povelje i pisma Despota Stefana*. Beograd: Čigoja štampa.

PAVLOVIĆ, Slobodan. 2011. Distribucija enklitika u starosrpskim poveljama i pismima XII i XIII veka. *Zbornik Matice srpske za filologiju i lingvistiku* 54(2): 31–52

PROGOVAC, Ljiljana. 2005. *A Syntax of Serbian: Clausal Architecture*. Bloomington: Slavica.

SPENCER, Andrew – LUÍS, Ana R. 2012. *Clitics: An Introduction*. Cambridge: Cambridge University Press.

STOJANOVIĆ, Ljubomir. 1929. *Stare Srpske povelje i pisma kniga I*. Beograd: Srpska Kraljevska Akademija. Electronic edition: FILIPOVIĆ, Želko – MERLOCAMPINI, Cordelio, eds. 2013. Niš – Beograd.

TOLSTAJA, Marfa Nikitična. 1991. Sistema enklitik v srpskih gramotach XIV – nachala XV v. In: *Slavistika, Indoevropaistika, Nostratika: K 60-letiju so dnja rozhdenija V.A. Dybo. Tezisy dokladov*. Moskva: Akademija Nauk SSSR, Institut slavjanovedenija i balkanistiki, pp. 201–205.

ZALIZNJAK, Andrej Anatoljevich. 2008. *Drevnerusskie enklitiki*. Moskva: Rossijskaja Akademija Nauk.

ZIMMERLING, Anton – KOSTA, Peter. 2013. Slavic clitics: A typology. *STUF* 66(2), pp. 178–214.

ZWICKY, Arnold M. 1977. *On Clitics*. Columbus: Ohio State University.

Lilla Nikolin Dukai

Faculty of Humanities and Social Sciences

University of Szeged

H-6722 Szeged, Egyetem u. 2.

Hungary

d.lilla96@gmail.com



This work can be used in accordance with the Creative Commons BY-NC-ND 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). This does not apply to works or elements (such as image or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights.

Roman Sukač

K TAKZVANÉMU KRÁCENÍ VOKÁLŮ V ČEŠTINĚ

ABSTRACT

The author deals with the data contained in the Atlas nářečí českého jazyka – krácení vokálů publication. Using the old and obsolete methodology, the authors of Atlas claim that the differences in quantity can be attributed to various post-Common Slavic prosodic processes. The author of the presented text examines the mixed bag of data in Atlas and proposes that the quantitative differences are connected to the development of quantity in paradigms.

KEYWORDS

quantity; vowels; Moravian shortening; accentology; acute

1 Úvod

Bádání v oblasti historického vývoje češtiny a jejího praslovanského základu vyvrcholilo ve dvou publikacích z druhé poloviny minulého století: *Historické mluvnici češtiny* autorů Lamprechta, Šlosara, Bauera (LAMPRECHT et al. 1986) a *Lamprechtovy Praslovanštiny* (LAMPRECHT 1987). Platí přitom, že obě práce nebyly svým rozsahem ani významem dosud překonány. Pohříchu v nich bylo méně pozornosti věnováno studiu praslovanské akcentologie, obě kanonické publikace (Lamprechtovou vinou) bohužel interpretačně setrvaly na úrovni předválečné „klasické“ akcentologie.¹

Ta, zjednodušeně řečeno, veškeré levostranné posuny přízvuku chápala jako *Hirtův zákon*, veškeré pravostranné posuny jako *de Saussurův zákon* (na základě analogie s litevštinou). Praslovanský akut měl být dlouhý a reflektovat se v češtině jako délka, cirkumflex pak jako krátkost. Různé nepravidelnosti byly vysvětlovány tzv. metatoniemi, při kterých se měly intonace přepólovat a dávat výsledky opačné od těch očekávaných. Všechny tyto mechanismy bez nějaké fonetické motivace

¹ LAMPRECHT (1987) sice v seznamu literatury uvádí jak Kortlandtovy práce, tak práce *Moskevské akcentologické školy*, a dokonce i přímo STANGA (konkrétně 2. vyd. 1965), ale jejich výklady vůbec nezohlednil.

pak byly komplikovány výsledky náhradního dloužení, kdy ztrátou koncových jerů v nominativu singuláru *o*-kmenů anebo genitivu plurálu *a*-kmenů měla (nebo také nemusela) být prodloužena kořenová slabika.

Od klíčové práce *Slavonic accentuation*, vydanou v roce 1957 Christianem Stangem (STANG¹1957; ²1965), se studium praslovanské akcentologie vyvíjelo převratně rychle.² Klasická akcentologie byla nahrazena podrobnou chronologií akcentuačních změn zakomponovaných do hláskového vývoje od baltoslovanštiny až po rozpad praslovanského systému. Metatonie byly odmítnuty jako nikdy neexistující mechanismy. De Saussurův zákon skutečně funguje, ale jen v litevštině. Naopak pro praslovanštinu bylo popsáno několik nových zákonů, přičemž postupně se začala ukazovat problematičnost mechanismů náhradního dloužení. Poznatky moderní generativní fonologie byly využity k interpretaci mechanismu interakce krátkých a dlouhých slabik a vzniku rytmicity.

Téměř nic z toho neproniklo do české diachronní lingvistiky. Na jedné straně absence rekonstrukce původní akcentuace nevádí, zvláště v etymologii, kde nám jde především o původní tvar. Na straně druhé tam, kde interpretace současného suprasegmentu vyžaduje historické vysvětlení, má neznalost současného stavu bádání v akcentologii fatální důsledky. Typickým případem je nedávná publikace NEJEDLÉHO a kol. (2019), v níž absence relevantní literatury vede k objevování mechanismů, které už byly předtím popsány (viz SUKAČ 2021, 72, pozn. 23).

Ani velké dílo české dialektologie *Český jazykový atlas* (totiž díl 5, *Hláskosloví*, dále citován jako ČJA 5) se nijak nevymyká z intencí klasické akcentologie. S oběma výše uvedenými kanonickými publikacemi v zádech jsou tak příklady kvantitativních diferencí vysvětlovány opět krácením, metatonie, náhradním dloužením, aniž by se nad těmito mechanismy a jejich důvody autoři zamýšleli (ČJA 5, 189). To, co nezapadá do daného metodologického schématu, je pak vysvětlováno tak, že „v důsledku rozkolísanosti a také proto, že se tendence k náhradnímu dloužení nemusela prosadit ve všech pozicích, kde k ní byly podmínky, a na celém území stejně, se dloužení objevuje u některých slov a tvarů jen v některých oblastech. V územním rozložení dlouhých a krátkých samohlásek jsou značné difference.“ (ibid). Zmatky vyvolává nejasnost v původu změny, např.: „Mapa zachycuje podoby s délkou a krátkostí u dvojslabičného feminina **sila** (psl **sila*). V první slabice se zachovala původní délka (za starou akutovou intonaci). Varianta **sila** zachovává pův. moravskou (a slovenskou) krátkost“ (ČJA 5, 242). Co je tedy původní, délka za starý akut, nebo je původní moravská a slovenská krátkost (nikoliv krácení!)?

Publikace *Atlas nářečí českého jazyka – krácení vokálů* (dále jen ATLAS) z roku 2020 je útlejším, ale co do mapového ztvárnění propracovanějším výtahem z ČJA 5, graficky zajímavější zpracování původních dat a mapových dokladů. Bohužel ani za patnáct

2 Přehledně SUKAČ (2020).

let od publikování ČJA 5 nejsou v této publikaci ani částečně zohledněny moderní poznatky o vývoji praslovanské akcentologie, autoři se bohužel nevyhnuli akcentologickým komentářům založených na takřka sto let staré metodologii, rovnou převzatých z „většího bratra“ ČJA 5. To bylo patrně způsobeno tím, že čtyři z pěti autorů nejsou lingvisty, ale geografy. Důvodně se tedy můžeme domnívat, že zastaralá metodologie byla přebrána *en bloc*. Nicméně v důsledku tohoto je pak zavádějící už sám podtitul *ATLASU: Krácení vokálů*, ve většině případů se zde totiž z pohledu moderních akcentuačních teorií nic nekrátí. Právě rozboru materiálu z *ATLASU* perspektivou moderní akcentologie věnujeme tuto studii.

2 Data a interpretace³

Autoři *ATLASU* jako ukázky tzv. krácení vokálů na našem území vybrali hrstku dat seřazených bez očividné souvislosti: *hrách, pláč, sáně, kláda, skála, kráva, sníh, vítr, křída, léto, jméno, péro, síla, hrouda, nést*. Proberme je detailně podle akcentologických souvislostí:

- a) *hrouda*, slk. *hruda*, s-chor. *grúda*, rus. *grúda*, **grúda*, APa (DERKSEN 2008, 192)⁴
kráva, slk. *krava*, hsrb. *kruwa*, pl. *krowa*, s-chor. *kráva*, **kóřva*, APa (DERKSEN 2008, 236)
kláda, slk. *klada*, hsrb. *klóda*, s-chor. *kláda*, **kólada*, APa (DERKSEN 2008, 228)
síla, slk. *sila*, s-chor. *síla*, **síla*, APa, (DERKSEN 2008, 451)
skála, slk. *skala*, s-chor. *skāla*, **skāla*, APa (DERKSEN 2008 neuvádí)

Všechna tato substantiva patří do praslovanského akcentuačního paradigmatu APa s akutem na kořeni. Tvary s kořenovou délkou mají varianty s krátkostí na Moravě: *hruda*, *krava*, *klada*, *sila*, *skala*. Autoři *Atlasu* vysvětlují varianty takto: *kláda/klada* – staré krácení ražených délek, *skála* – původní délka za raženou intonaci, *skala* – moravská krátkost, popř. analogické vyrovnávání podle tvarů s *-a-* (*skalám*, *skalách*); *kráva* – uchování původní délky, *krava* – staré krácení, případně vyrovnání podle tvarů s *-a-* (*kravách*, *kravami*); *síla* – původní délka za starou akutovou intonaci, *sila* – moravské krácení. Bohužel, nic z toho neodpovídá moderní akcentologii.

Především, nepatří tradovaný předpoklad, totiž že za akut musíme mít v češtině délku. Z původní akutovaných slov máme sice, mimo výše uvedených, délku v *hlína*, *hříva*, *jáma*, *lípa*, *rána*, *sláma*, *sláva*, ale už ne v *hnida*, *niva*, *piha*, *duha*, *ryba*. Naopak

3 Přehled směrů moderní akcentologie v SUKAČ (2013b; 2013c), pro českého čtenáře SUKAČ (2013a; 2015; 2020).

4 Uvádím praslovanskou rekonstrukci tak, aby kromě základního českého byl zřejmý i slovenský tvar, případně další, který pomáhá rekonstruovat původní praslovanskou akcentuaci a zařazení slova do odpovídajícího akcentuačního paradigmatu.

na Doudlebsku a Chodsku je doložen tvar *rejba* z původního *rýba*. Původní délka by měla být teritoriálně zachována v jihozápadočeském *pína*, *plíva*, *slína*, *hňída*, *jískra*, *jíkra*, *douha*, *rejba* (VORÁČ 1955), přechodném pásu česko-moravských dialektů *pína*, *plíva*, *slína*, *kúpa* (UTĚŠENÝ 1960), a východočeském *kníha*, *slína* (BACHMANN 2001). Většina českých dialektů má kořenovou délku *řípa*, na rozdíl od spisovné varianty a tvaru *řepa* na východě území.⁵

Proč tedy máme v češtině dvojí reflex původního akutového paradigmatu? Situaci navíc komplikují staročeské dublety *čěsta/ciesta* (NEDVĚDOVÁ 39), *pěna/piena* (NEDVĚDOVÁ, 40), *pleva/pléva* (NEDVĚDOVÁ 40), *ryba/rýba* (NEDVĚDOVÁ 39), *slina/slína* (NEDVĚDOVÁ 40) a doložená délka v *kníha*, *múka*, *viězka*, *níva*, *ščíka* (Klaret).⁶ Je dost obtížné vysvětlit tyto varianty různými analogiemi, obměnami a nepravidelnostmi, u nichž se neuvádí žádná motivace.

Dalším problémem je krátkost akutových tvarů na Moravě. Autoři ATLASU vycházejí samozřejmě z ČJA 5, kde se užívá „ve shodě s tradicí [...] vedle termínu moravská krátkost též termín staré krácení nebo moravské krácení“ (ČJA 5, 218). Zaměňování těchto termínů ale neznamená, že jde o tytéž jevy. *Moravská krátkost* znamená, že uvedené akutové tvary jsou od původu krátké. To je, jak uvidíme, zcela zřejmé. *Staré krácení* by znamenalo krácení akutových délek po rozpadu praslovanštiny a nástupu distinktivní kvantity v češtině. To je, jak také brzy uvidíme, zcela vyloučeno. Akut se totiž nekrátí, ale naopak akut prodlužuje. *Moravské krácení* by pak byl jev v historickém vývoji češtiny omezený pouze na Moravu. To, jak uvidíme, existuje, ale nemá nic společného s akutem. Navíc autorům očividně uniká souvislost moravských krátkých tvarů se stejně krátkými tvary slovenskými (na Slovensko pochopitelně mapa nepokračuje). Takže *moravská krátkost* nebo *krácení* by také mělo zahrnovat *slovenskou krátkost/krácení* (proto jsou také níže uváděny i slovenské tvary).

Je zřejmé, že celý problém se točí kolem termínu akut.⁷ Klasická akcentologie vycházela z předpokladu, že praslovanský akut byla dlouhá stoupavá intonace. Tím ovšem sama sebe zavedla na scesti, protože pokud by se praslovanský akut měl reflektovat jako délka, pak je tato délka zachována jen v českých dialektech *sensu stricto* a v horní lužické srbštině (délka ve slovinštině je sekundární). Všude jinde by mělo docházet k masivnímu krácení, a to z důvodů, které nejsou nijak vyloženy. To je také předpoklad, s nímž pracují autoři ČJA 5 a ATLASU. Jestliže vedle sebe na mapě postavíme tvary *kráva* – *krava*, okamžitě nás napadne, že na východě území dochází ke krácení. To je ale optický klam, protože kdyby mapa pokračovala dál na Slovensko, také tam bychom viděli krátké tvary a museli předpokládat krácení na obrovském území.

Jestliže ovšem předpokládáme, že praslovanský akut je původní **glotalizace** přecházející v krátkou stoupavou intonaci,⁸ pak tzv. Kortlandtovým pravidlem dochází

5 ČJA 5 (2005, 195).

6 Podrobně SUKAČ (2013a).

7 Podrobně a detailně SUKAČ (2015).

8 KORTLANDT (1983). Předpoklad praslovanského akutu jako praslovanské glotalizace není žádný

společně s nástupem iniciálního přízvuku k prodloužení této intonace (srov. SUKAČ 2013a; 2013c; 2015). Délka za akut je tedy sekundární. Primární je krátkost, jak ji zaznamenáváme na Moravě a ve slovenštině.⁹

b) *sáně*, slk. *sane*, strus. *sani*, stpol. *sani*. Jde o původní praslovanský *i*-kmen. Litevské *šonas* ukazuje na rekonstruovaný baltoslovanský akut **šaʔnas/-is* (DERKSEN 2008, 443), a tedy původní praslovanské akcentuační paradigma APa. Podle autorů ATLASU je „varianta *sáně* považována za starší, délka první slabiky reflektuje praslovanské ražené intonace“ (sic! Takže jich bylo více?). Těžko říct, proč autoři ATLASU vybrali z ČJA 5 právě substantivum *sáně*, jehož krátká varianta zabírá 61% území a popírá existenci tzv. moravského krácení (protože by se muselo krátit i porůznu v Čechách a na Slovensku).

Vidíme, že pokud by autoři ATLASU i ve své překonané interpretační koncepci měli zvolit nějaké pořadí substantiv, měla by původní akutová feminina logicky být u sebe. Tak tomu však není.

c) *hrách*, slk. *hrach*, pol. *groch*, hsrb. *hróch*, rus. *goróch*, s-chor *gräh*, **górchъ*, APa (DERKSEN 2008, 179)
pláč, slk. *plač*, s-chor. *pláč*, **pláčь*. Derivát, který se běžně v etymologických slovnících nevyskytuje, ale doložené tvary ukazují na praslovanské akcentuační paradigma APa.
vítr, slk. *viator*, s-chor. *vjětar*, **větrъ*, APa (DERKSEN 2008, 520).
sníh, slk. *sneh*, čak. *snih*, **sněgъ*, APC (DERKSEN 2008, 274–275)

Komentáře autorů ATLASU: *hrách* – původní délka, *hrach* – moravská krátkost; *pláč* – původní délka, *plač* – moravská krátkost; *sníh* – z původního *snieh*, jinak neobjasněno; *vítr* – ze staročeského *viatr*, jinak neobjasněno, *větr* – patrně důsledkem starého krácení, *vitr* – pravidelné krácení dlouhého *-í-*; *snih* – pravidelné krácení dlouhého *-í-*, *sněh* – neobjasněno.

V této směsi *o*-kmenů nacházíme maskulina původního akutového paradigmatu APa, ale i APC, což je paradigma s původním cirkumflexem. Je jasné, že kořenová délka tvaru *sníh* je jiného původu než u akutových paradigmat. Navíc, a to je podstatné, jde o původní nominativy singuláru zakončené na *jer*. Zde se projevuje jistá problematická zvoleného interpretačního a terminologického aparátu autorů

ad hoc výmysl, ale regulérní rekonstrukce odvíjející se od baltoslovanského akutu, podrobně a detailně SUKAČ (2015) s odkazem na další Kortlandtovy práce.

⁹ Ne každá délka v dvojslabičných *a*-kmenech je ovšem sekundární. Tvary *tráva*, *touha*, *kroupa*, *líska*, *díra* proti *žena*, *koza*, *pila*, *ruda* jsou výsledkem Dybova zákona, který obnovil předpřízvučné délky (ŠEFČÍK 2017). Nelze zde podávat detailní diskusi k moderní akcentologii. Čtenář najde podrobná vysvětlení a diskuse v pracích uvedených v seznamu literatury.

ČJA a ATLASU – jestliže tvar *snih* je výsledkem pravidelného krácení, rozhodně musí být pozdější než údajné krácení tvaru *hrách*. Navíc je z mapy jasné, že tvary *sněh* musí být původní, protože jsou zachovány na opačných stranách území, a to v západních Čechách a na Uherskobrodsku a Vsetínsku.

V distribuci délek o-kmenů z hlediska české kvantity je zdánlivě mnoho nejasností. Maskulina původního akutového akcentuačního paradigmatu APa mají většinou kořenovou délku, až na *kraj*, *pluh*, *čas*, *had*, *rak*. Tvary s délkou naopak nacházíme v jihozápadočeských dialekttech (VORÁČ 1955). Délku za původní akutová paradigmata zaznamenává i ČJA (5, 190–208), a nápadná je právě délka zde a krátkost jinde v Čechách a na Moravě: *pěna–pína*, *čas–čas*, *slina–slína*, *kniha–kníha*, *koš–kůš*, *kraj–kráj*. Na jedné straně sice existuje nápadný kontrast délky v Čechách a krátkosti na Moravě: *mák–mak*, *hrách–hrach*, *pláč–plač*, *skála–skala*, *žába–žaba*, *jáma–jama*, *kláda–klada*, *vrána–vrana*, *sláma–slama* (krátkost jen ve slezských dialekttech!), *kráva–krava* (krátkost na jihozápadě Moravy a ve slezských dialekttech!), *vítr–větr*, *bříza–březa*, *síla–sila*, *lípa–lipa*, *hříva–hřiva*, *hrouda–hruda* (ČJA 5, 218–245). Na straně druhé nemůže být žádné řeči o nějakém moravském krácení, vždyť dobrá polovina dat, která autoři zaznamenali, tomuto tvrzení odporuje. Jistě, můžeme tyto nepravdivosti vždy odbýt vysvětlením analogií, vyrovnáváním podle náhodně vybraných pádů, ale třeba podle instrumentálu se nominativ typologicky nevyrovnává (kdyžtak podle akuzativu) a problém interpretace je jen odsunut.

O-kmenová maskulina původního akcentuačního paradigmatu APb, tedy oxytona vzniklá Dybovým zákonem, se reflektují buď jako tvary s kořenovou délkou, nebo krátkostí, zvláště v případech původního kořenového -o-: *bob*, *boj*, *dvůr*, *hloh*, *kůň*, *kůl*, *koš*, *krov*, *nůž*, *půst*, *roj*, *skot*, *snop*, *stoh*, *stůl* (SUKAČ 2013a). Totéž platí o podobných o-kmenových maskulinech cirkumflexového paradigmatu APC: *bůh*, *bok*, *dům*, *hnůj*, *lůj*, *důl*, *nos*, *bol*, *bor*, *brod*, *hoj*, *hrom*, *hrob*, *chod*, *kos*, *lov*, *most*, *nos*, *pot*, *rod*, *roh*, *tok*, *tvor*, *pot*, *vosk*, *zob*, *zvon* (SUKAČ 2013a). Zdánlivě zde pozorujeme dloužení před sonorami a znělými frikativami: *bůh*, *důl*, *hnůj*, *dům*, *vůz*, ale ne před znělými explosivami a neznělými konsonanty: *drob*, *brod*, *hod*, *plod*, *plot*, *rod*, *med*, *led*, *brod*, *most*, *nos*, *rok*, *bok*, *sok*, *tok*, *nos* atd. (TIMBERLAKE 1983). Situaci dále komplikuje absence touto teorií očekávaného dloužení u tvarů *kov*, *lov*, *boj*, *bor*, *hrom*, *mor*, *roh*, *strom*, *zvon*, přičemž staročeština zaznamenává tvar *róh* stejně jako slezské dialekty *ruh*, *zvun*, *rud* (SUKAČ 2013a).

Tvary s délkou lze samozřejmě interpretovat jako tzv. náhradní dloužení po ztrátě koncového jeru. Toto dloužení by mělo probíhat ve všech původních paradigmattech APa, APb a APC (TIMBERLAKE 1983). Domníváme se, že žádné náhradní dloužení nikdy neexistovalo (podrobné zdůvodnění SUKAČ 2011).

Distribuci české délky v postpraslovanském stavu je možné jednoduše vysvětlit:

- a) využitím Kortlandtova pravidla o prodloužení krátké stoupavé intonace;
- b) teorií Bethinové (BETHIN 1993; 1998) o tom, že hláskové změny v praslovanštině

vedou k vývoji dvojslabičné fonologické domény; kombinací iniciálního přízvu-ku a distinktivní délky pak k vývoji trochejského rytmu v západoslovanských jazycích;

- c) teorií kvantitativních paradigmat navržené původně Feldsteinem (FELDSTEIN 2007) a rozpracované Sukačem (SUKAČ 2013c).

Detaily a podrobné vysvětlení najde čtenář v uvedených pracích. V principu se však jedná o toto: Původní dvojslabičné *o*-kmeny maskulin byly zakončeny na *jer* v nominativu singuláru. Původní akutové APa paradigma, např. **klin* po rozpadu praslovanštiny mělo iniciální přízvuk, tím pádem podleho Kortlandtovu pravidlu. Kombinace přízvu-ku a délky vede ke vzniku dvojslabičné, maximálně třímorové domény¹⁰ (závorky tuto doménu označují): (**kli:n.ɚ*). Následuje resylabifikace (**kli:n.ɚ*) > **kli:n*, *jer* je eliminován. Ve výsledném tvaru *klín* žádné náhradní dloužení neproběhlo, protože se nemá co nahrazovat.¹¹

Původní oxytonové paradigma APb, např. **stolʔ*, rovněž vytváří dvojslabičnou, maximálně třímorovou doménu (*sto.lʔ*). Český fonologický systém vyžaduje přízvuk na první slabice, musí tedy dojít k retrakci přízvu-ku a zároveň u tohoto typu dochází k prodloužení kořenové slabiky (*sto:lʔ*). Resylabifikací je pak eliminován koncový *jer* *(*sto:l*)ʔ. a výsledný tvar je *stól*. Opět, žádné náhradní dloužení neexistuje.

Původní cirkumflexové paradigma APC, např. **bôgʔ*, podléhá stejnému vývoji. Resylabifikace a vytvoření dvojslabičné, maximálně třímorové domény vede ke tvaru *(*bog*)ʔ, a eliminaci původního *jeru*. Cirkumflexová intonace bez ohledu na svůj charakter je v postpraslovanském období v západní slovanštině ztracena. U APb typu **stolʔ* je délka kořenového *-o-* vyvolána retrakcí přízvu-ku. U APC musíme ale počítat s jiným mechanismem. Jediné přijatelné vysvětlení se zdá být vznik morové kody a následná retrakce *mory* na slabičné jádro¹²: **bo_μg* > **bo_μg_μ* > **bo_{μμ}g* > *bo:g* neboli *bóh*. Samozřejmě, že žádné náhradní dloužení před znělým konsonantem tady rovněž neexistuje. Koncepte náhradního dloužení je jedním z dalších tradovaných mýtů, který nebere v úvahu skutečné chování slabik.

Tím jsme ovšem jen vysvětlili možné výsledky v nominativu singuláru. Proč tedy

¹⁰ Počet mor ve dvou slabikách tedy nepřekračuje číslo 3. Neboli střídá se krátká a dlouhá slabika, což je rytmický zákon. O tom, že čeština má za jistých prozodicky přesně definovaných podmínek stejný rytmický zákon jako slovenština, viz nejnoveji SUKAČ (2021).

¹¹ To je další často tradovaný omyl. Jak může ztráta oslabeného *jeru*, který už sám je od počátku redukovanou hláskou, vést k náhradnímu dloužení a vzniku plného, dvoumorového dlouhého vokálu? Nehledě na to, že se zde skutečně nemůže jednat o dloužení za ztrátu slabiky, která je navíc už redukována. Taková slabika prostě zmizí (jako v germánských jazycích), nebo zmizí proto, že dochází k resylabifikaci tvaru, což je případ nominativu singuláru *o*-kmenů a genitivu plurálu *a*-kmenů v pozdní praslovanštině.

¹² Jedině existenci morových konsonantů ve slabičné kodě se dá vysvětlit distribuce kvantity (samozřejmě v bisylabické, maximálně třímorové doméně) v tvoření dvojslabičných hypokoristik: *Kateřina* > *Ká_{μμ}řa_μ/Ka_μtka_μ* a samozřejmě také staročeskou distribuci derivátů *-anie/-ánie*, *-ař/-ář*. Oba fenomény byly popsány (BETHIN 2003; SUKAČ 2011). Principy této distribuce byly „znovuobjeveny“ v publikaci NEJEDLÉHO a kol. (2019), a to bez reference k dřívějším pracím.

nacházíme varianty typu *kraj–kráj*, *roh–ruh*, u *a*-kmenů feminin pak *kráva–krava*, *skála–skala*? Důvod je jednoduchý, a tím se dostáváme k výše uvedenému bodu c) neboli koncepci kvantitativních paradigmat.

Jakmile se ustálí přízvuk a trochejský rytmus, praslovanská akcentuační paradigmatata už více neexistují, neboť substantiva začínají vytvářet skupiny tzv. kvantitativních paradigmat. Všechny dvojslabičné *a*-kmeny a původně dvojslabičné *o*-kmeny je možné do těchto paradigmat zařadit právě podle klíčových pádů. To je to, co autorům ATLASU a ČJA 5 uniká. Sleduje se distribuce délky/krátkosti v nominativu singuláru nebo jiném pádu, ale to vůbec není důležité z hlediska celého paradigmatu. Jak jinak bychom měli vysvětlit alternaci *stůl–stolu* oproti *kůl–kůlu* (obě substantiva patří do původního APb), nebo anomální genitivu plurálu *kráva–krav*, *hlína–hlín*, *slína–slín* (všechna tři substantiva jsou původní akutová feminina paradigmatu APa).

Tyto klíčové pády jsou: nominativ singuláru a genitiv plurálu pro *a*-kmeny (typ *kráva–krav*) a nominativ singuláru a genitiv singuláru pro *o*-kmeny (typ *stůl–stolu*). Proč právě tyto pády? Jak vidíme, tyto klíčové pády jsou typické alternací koncovky vokál/nula. Z tohoto hlediska jsou irelevantními pády např. *Lpl* nebo *Ipl*, jako třeba *kravách*, *krávách*, *kravami*, *krávami*, kde dochází k alternaci kvantity v dvojslabičné doméně, ale pro zařazení slova do paradigmatu nemají vliv. Proto tedy nemůžeme z těchto tvarů pro nominativ singuláru nic vyvozovat. Instrumentál a lokál prostě nominativ neovlivňují, ačkoliv tvary mají stejný kořen. Okrajový nepřímý pád, až na výjimky, nemůže ovlivňovat pád přímý.

Rozlišujeme dvě základní kvantitativní paradigmatata: alternující a konstantní. Alternující paradigma má dvě varianty podle toho, zda je v nulovém tvaru délka nebo krátkost. Konstantní paradigma má buď konstantní délku v klíčových pádech, nebo konstantní krátkost. Distribuce dvojslabičných substantiv (*a*-kmeny a původní dvojslabičné *o*-kmeny) podle kvantity je tedy následující:

- alternující kvantitativní paradigma (tvar s nulovou koncovkou má krátkost, tvar s vokalickou koncovkou má délku a naopak, např. *jáma–jam*, *bába–bab*, *kráva–krav*; *stůl–stolu*, *bůh–boha*)
- konstantní kvantitativní paradigma (tvar s nulovou koncovkou má délku, tvar s vokalickou koncovkou má také délku, např. *bříza–bříz*, *hlína–hlín*, *hlíva–hlív*, *kůl–kůlu*;
- konstantní kvantitativní paradigma (tvary s nulovou a vokalickou koncovkou mají krátkost např. *ryba–ryb*, *pleva–plev*, *cesta–cest*, *nos–nosu*;
- alternující kvantitativní paradigma (tvar s nulovou koncovkou má délku, tvar s vokalickou koncovkou má krátkost). Toto paradigma se nevyskytuje v současné češtině (ve staročeštině ano), ale můžeme ho pozorovat ve slovenštině: *brada–brád*, *cena–cien*, *hora–hôr*.

Substantiva se do kvantitativních paradigmat zařazovala postupně. Také často do-

cháží k jejich disperzi z původních akcentuačních paradigmat do různých kvantitativních paradigmat. Proto můžeme pozorovat staročeské dublety *zima/zíma*, *strana/strána*, jihozápadočeské *kráj*, *plouh*, *rák* nebo nářeční tvary *nůž-nůža*. Neočekávané tvary tedy není třeba vysvětlovat nepravděpodobnými analogickými změnami. Také *moravská krátkost* nevyžaduje žádné nemotivované krácení, žádné totiž nikdy nebylo. Pokud odhlédneme od nominativu singuláru, na který se autoři ČJA 5 a ATLASU při *moravské krátkosti* nepochopitelně odvolávají, jedinou motivací krátkosti je buď krátkost původní, nebo je důsledkem přechodu do konstantního paradigmatu, např. *větr-větra*.

Jediné skutečné moravské krácení z tvarů uváděných ATLASEM je *sníh-snih*, ale není důsledkem žádného krácení nominativu singuláru, nýbrž přechodem do jiného kvantitativního paradigmatu. Tvar *sněh* je původní.¹³

Jak bylo výše uvedeno, prodloužení pod akutovými slabikami nastalo jen v českých dialektech a v hornolužičtině. Na zbytku západoslovanského území došlo jen k ustálení přízvuku, vzniku rytmicity a distribuci do kvantitativních paradigmat. *Moravská krátkost* je tak jen zcela zákonitý jev přechodu původních akutových paradigmat v kvantitativní paradigma konstantní s krátkostí v klíčových pádech.

Je tedy v zásadě úplně lhostejné, zda ČJA 5 nebo ATLAS zaznamená tvary *vitr*, *vitr* a *větr* z hlediska distribuce kvantity. Podstatné je, jaký je tvar genitivu singuláru, zda *vitr-větru* (alternující paradigma), *větr-větra*, *vitr-větra* (konstantní paradigma). Nyní se můžeme podívat na zbytek tvarů uváděných ATLASEM, a to opět optikou moderní akcentologie.

d) *péro*, Slk. *pero*, Čak. *però*, **però*, APb (DERKSEN 2008, 396);

Komentář autorů ATLASU převzatý z ČJA 5 (240): obě varianty *péro/pero* jsou odrazem starých intonačních poměrů.

Těžko říct, co tím autoři mysleli, protože se jedná o oxytonon akcentuačního paradigmatu APb. Vzhledem k předchozímu výkladu jsou v zásadě možné obě varianty *péro/pero*. Původní kořenové *-e-* je krátké, takže očekávaný pračeský tvar bude rovněž *pero*. Tomu ostatně odpovídá i distribuce tohoto tvaru na mapě – je zachycena v jihozápadočeských nářečích a v nářečích slezských. Většinu území pak tvoří

¹³ Jaký by byl důvod k tomu, aby z konstantního kvantitativního paradigmatu (*sněh* – Gsg. *sněhu*) slova přecházela do alternujícího kvantitativního paradigmatu (*snieh* – *sněhu*)? Čím to šlo zdůvodnit, když je takový přechod proti tendenci k unifikaci kořene a jazykové ekonomii (srov. jerové vyrovnávání *mecha* – *mcha* > *meh* – *mecha*, likvidace alternací v Lsg. *Zábřeze* > *Zábřehu* atd.)? Tendence k unifikaci kořene je opravdu jen tendence, nikoliv pravidlo. Jinak bychom v češtině, kromě kvantitativních alternací typu *vrána* – *vran* nemohli mít desítky příkladů vokalických alternací, např. *kra-ker*, *zlo-zel* nebo alternací konsonantických *kousat-koušu*. Ovšem srovnávat jerové vyrovnání (které souvisí s odstraňováním konsonantických skupin) s přechody do kvantitativních paradigmat možné není. O tom, že substantiva očividně mezi kvantitativními paradigmaty přechází, nám dokládají historická a dialektologická data. Jinak bychom se zase museli vrátit k všemocným analogiím, kde stačí vzít nějaký vhodný tvar a podle něj vysvětlit vyrovnání nominativu nebo genitivu.

sekundární tvar *péro*, očividně ve dvou kvantitativních paradigmatech *péro-per*, *pero-per* s pozdějším lokálním úžením *piro*.

e) *léto*, Slk. *leto*, S-Chor. *ljěto*, APa (DERKSEN 2008, 274-275);

Komentář autorů ATLASU: *léto/líto*, délka za starou raženou intonaci, místy krácení, *leto* – důsledek starého krácení, příp. paradigmatické vyrovnání podle podob s krátkou samohláskou, např. *let*, *letům*.

Vidíme zde opět zmatení pojmů. Jednak z komentáře vyplývá dvojí krácení – jedno, údajně staré, za akut, které ovšem, na rozdíl od očekávaného *moravského krácení*, pokrývá 70 % území. A jestliže tento výklad náhodou nevychází, na pomoc přijdou nepřímé pády s krátkým *-e-*. Pak je zde další krácení původní délky zúženého *-é-* > *-í-*: *lito*.

Z výše uvedeného je jasné, že původní tvar za akut musí být krátký, tedy *leto*, čemuž odpovídá i ona většinová distribuce na našem území. Tvar *léto/líto* je sekundární a patří do kvantitativního alternujícího paradigmatu *léto-let*, tvary *lito-let* by pak bylo odrazem pozdějšího přechodu do paradigmatu kvantitativního konstantního.

K samostatnému heslu atlasu *léta* „letokruhy“ lze dodat jen to, že z hlediska kvantitativních paradigmat není nominativ plurálu podstatný. Kvantitativně většinou odpovídá nominativu singuláru (ostatně, je to přímý pád). Pokud neodpovídá, jako je tomu v případě mapy uváděné ATLASEM, kde *líta* má jinou a geograficky širší distribuci, jde o podtyp konkrétního základního kvantitativního paradigmatu. Podstatný je tvar genitivu plurálu, ale ten, v tomto významu, bohužel ATLAS nezařazuje.

f) *jméno*, StČ. *jmě*, Slk. *meno*, S-Chor. *ime*, **jǫmę*, APc (ERHJ, 347);

Z hlediska kvantity problematický tvar, protože se jedná o *n*-kmen dotvořený do podoby *o*-kmenu. Staročeský tvar *jmě* a pozdější *jméno* ukazuje na českou (nikoliv praslovanskou) krátkost. ČJA 5, 240 i autoři ATLASU s odkazem na UTĚŠENÉHO (1960) uvádějí, že tvar *méno* vznikl v důsledku náhrady za ztrátu *j-*. Jedná se o dosti nepravděpodobné vysvětlení náhradního dloužení za ztrátu pobočné slabiky. Podobu *meno* zachycená v Čechách je očividně původní a je v kvantitativním konstantním paradigmatu, zatímco moravské *méno* je sekundární (ovšem nevzniklé z *jméno*), které přešlo do paradigmatu kvantitativního alternujícího. Tvary *mino* jsou pak terciární z *méno*, distribuované opět v konstantním paradigmatu.

g) *křída* – Není jasné, proč autoři ATLASU zvolili z ČJA 5 právě toto substantivum. Původní délka této přejímky ze středohornoněmeckého *krīde* je zachována v Čechách a na východě Moravy. Moravské *křída* se objevuje s přesahem do poměrně velké oblasti jihozápadních Čech směrem k Plzni. Překvapuje tedy, že autoři

ATLASU interpretují tento tvar jako moravskou krátkost (nikoliv ale autoři ČJA 5, tam se hovoří o zkrácení na Moravě). Atlas tedy ztotožňuje krátké *křída* s podobnými krátkými tvary původních akutových *a*-kmenů. Navíc autoři v jistém smyslu popírají sami sebe, protože podle jejich mapy východ Moravy zachovává délku. Nejde tedy o žádnou moravskou krátkost, ale jen krátkost na určitém území Moravy a značné části Čech.

Je jasné, že zkrácení je sekundární a tento tvar souvisí s přechodem do kvantitativního alternujícího paradigmatu, pokud je Gpl *kříd*.

h) *nést*, Slk. *niest'*, Rus. *neští*, S-Chor. *něsti*, **nešti*, APC (DERKSEN 2008, 350).

Komentáře autorů: na 8 % území varianta **nést**, shodná se spis. jaz., která je odrazem starého rozdílu v kvantitě. Na 29 % území obměna **nýst**, vzniklá úžením *-é* v *-í*-, na 4 % území varianta **nyst**, vzniklá pozdějším krácením, případně analogickým vyrovnáváním podle tvarů příč. min. či ind. přítentu, vyloučit nelze působení supina, na 2 % obměna **nist**, vyvinutá v důsledku slez. krácení, na 2 % varianta **něšt**, která je výsledkem starého krácení, a na 31 % varianta **nest**, jež je odrazem krátkosti v souvislosti s vývojem praslovanských přízvukových a intonačních poměrů.

K výkladům tvarů *nýst*, *ništ'* a *něšt* nelze nic namítat. Je ovšem nutné se ptát, co autoři myslí „starým rozdílem v kvantitě“ u tvaru *nést*, když kořenový vokál byl v praslovanštině stejný ve všech tvarech. A také, které praslovanské přízvukové a intonační poměry vedly k pozdějšímu kontrastu obou (z hlediska autorů očividně původních) tvarů *nést/nest* (pokud ovšem původností nemyslí pračeské/staročeské tvary *něšti-nesu*).

Jelikož v staročeštině existuje nápadný kontrast mezi dvojslabičnými infinitivy, které mají kořenovou délku, a supiny, která mají krátkost (*něšti/nest*), bylo navrženo toto řešení, které jednoduše zapadá do výše uvedeného mechanismu transformace praslovanských akcentuačních paradigmat v paradigmatu kvantitativní (ŠEFCÍK et al. 2014): Jestliže krátké stoupavé intonace byly s ustálením přízvuku v 10. století (po zániku jerů a kontrakci) prodlouženy, např. **psāti* > *psāti* u původních oxyton, a také **kráva* - *kráva* u původních akutových baryton (čili Kortlandtovo prodloužení), oba procesy vedly k nástupu rytmicity. Ta byla která byla nejzřetelnější právě u dvojslabičných tvarů. Infinitiv a supinum byly v této době stále živé tvary, ovšem rozlišené ne příliš odlišnými koncovkami *-ti* a *-t*. Kvantitativní opozice tedy vedla k převaze dvojslabičných infinitivů s dlouhým kořenovým vokálem oproti jednoslabičným tvarům supina s krátkým vokálem. Tento jev se obecně projevil u dvojslabičných tvarů (infinitiv a původní dvojslabičné supinum), neboť právě u dvojslabičných tvarů infinitivu mohlo dojít k redukci koncového *-i* a kolizi se supinem: *něst(i)* × *nest*. Vzhledem k tomu, že supinum bylo významově omezeno, docházelo k jeho náhradě infinitivem.

Tím je jasné, proč na českém území nacházíme tvary *nést/nýst*. Tvar *nest* je očividným zkrácením *nést(i)* a nemá nic společného se supinem nebo přičestím minulým, které naopak byly a jsou kvantitativně odlišné. To, že si slovesné tvary jsou podobné, neznamená, že se navzájem ovlivňují, takhle analogie přece nefunguje.

3 Závěr

Přestože tento text není recenzí, ale akcentologickým (nikoliv dialektologickým) komentářem k výsledkům prezentovaným kolektivem autorů ATLASU (s ohledem na ČJA 5), lze konstatovat následující: pokud se česká diachronní lingvistika a dialektologie neoprostí od dogmat tradovaných nejen půlstoletí, ale téměř celé století, a bude dále ignorovat práce, které se v jejich odborném prostoru vyskytují, a to se akcentologie týká především, výsledkem bude vždy jen jeden výstup. Na jedné straně barevně zajímavá publikace podpořená grantem a spoluprací s jiným odborným pracovištěm. Na straně druhé pak bohužel odborná práce mající, abychom si vypůjčili titul jednoho Kortlandtova článku: *bad theory, wrong results*.¹⁴

LITERATURA

- ATLAS: IREINOVÁ, Martina – VOŽENÍLEK, Vít – POSPÍŠIL, Marek – KONÍČEK, Jakub – VONDRÁKOVÁ, Alena. 2020. *Atlas nářečí českého jazyka – krácení vokálů*. Olomouc: Univerzita Palackého v Olomouci.
- BACHMANN, Luděk. 2001. *Nářečí na Vysokomýtsku*. Praha: Academia.
- BETHIN, Christina, Y. 1998. *Slavic Prosody. Language Change and Phonological Theory*. Cambridge: Cambridge University Press.
- BETHIN, Christina, Y. 2003. Metrical Quantity in Czech: Evidence from Hypocoristics. In: BROWNE, Wayles (ed.). *Formal Approaches to Slavic Linguistics 11: The Amherst Meeting 2002*. Ann Arbor: Michigan Slavic Materials, 63–82.
- ČJA 5 – BALHAR, Jan a kol. 2005. *Český jazykový atlas 5*. Praha: Academia.
- DERKSEN, Rick. 2008. *Etymological dictionary of the Slavic inherited lexicon*. Leiden: Brill.
- ERHJ – MATASOVIĆ, Ranko – PRONK, Tijmen – IVŠIĆ, Dubravka – RONČEVIČ BROZOVIČ, Dunja. 2016. *Etimološki rječnik hrvatskoga jezika. 1 svezak. A–Nj*. Zagreb: Institut za hrvatski jezik i jezikoslovlje.
- FELDSTEIN, Ronald F. 2011. Nominal prosodic paradigms and their synchronic reflexes in West Slavic. In: PRONK, Tijmen – DERKSEN, Rick (eds.). *Accent Matters. Papers on Balto-Slavic Accentology*. Rodopi, 97–108.
- KORTLANDT, Frederik H. H. 1983. Linguistic theory, universals, and Slavic accentuation. *Folia Linguistica Historica*, IV. 27–43.

14 Za cenné komentáře k textu děkuji Jiřímu Rejzkovi.

- KORTLANDT, Frederik H. H. 2011. West Slavic accentuation. In: SUKAČ, Roman (ed.) *From Present to Past and Back. Papers on Baltic and Slavic Accentology*. Frankfurt am Main: Peter Lang, 86–107.
- LAMPRECHT, Arnošt – ŠLOSAR, Dušan – BAUER, Jaroslav. 1986. *Historická mluvnice češtiny*. Praha: Státní pedagogické nakladatelství.
- LAMPRECHT, Arnošt. 1987. *Praslovanština*. Brno: Univerzita J. E. Purkyně v Brně.
- NEDVĚDOVÁ, Milada. [nedatováno]. *Kvantita samohlásek ve starší češtině*. Praha: Oddělení vývoje jazyka Ústavu pro jazyk český AVČR.
- NEJEDLÝ, Petr a kol. 2019. *Slovotvorný vývoj deverbálních substantiv ve staré a střední češtině*. Praha: Academia..
- STANG, Christian. 1957. *Slavonic Accentuation*. Oslo: I kommisjon hos H. Aschehoug.
- STANG, Christian. 1965. *Slavonic Accentuation*. Oslo: Universitets-forlaget.
- SUKAČ, Roman. 2011. Old Czech rhythmic law. In Standik-Holzer, Elena. *Baltische und slavische Prosodie*. (Hrsg. Stadnik-Holzer, E.), Frankfurt am M.: Peter Lang, 141–152.
- SUKAČ, Roman. 2011a. Moravian Accentual Paradigms. in Sukač, Robert (ed.) *From Present to Past and Back. Papers on Baltic and Slavic Accentology*. Peter Lang, 194–200.
- SUKAČ, Roman. 2011b. Compensatory Lengthening in West Slavic. *Zeitschrift für Slawistik*, 56/4, 417–445.
- SUKAČ, Roman. 2013a. Fish and its Fisherman. Paradigmatic and derivative length in Czech. *Zeitschrift für Slavistik*, 58/1, 72–101.
- SUKAČ, Roman. 2013b. *Introduction to Proto-Indo-European and Balto-Slavic Accentology*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- SUKAČ, Roman. 2013c. Od praslovanské akcentuace k západoslovanské prozodii (s ohledem na češtinu). *Slavia* 82/1-2, 213–219.
- SUKAČ, Roman. 2015. Praslovanský akut. *Bohemica Olomucensia*, 7/3–4, 8–21.
- SUKAČ, Roman. 2017a. Kvantita. In KARLÍK, Petr – NEKULA, Marek – PLESKALOVÁ, Jana (eds.), *CzechEncy – Nový encyklopedický slovník češtiny*. <https://www.czechency.org/slovník/KVANTITA>.
- SUKAČ, Roman. 2017b. Rytmicita. In: In KARLÍK, Petr – NEKULA, Marek – PLESKALOVÁ, Jana (eds.). *CzechEncy – Nový encyklopedický slovník češtiny*. URL: <https://www.czechency.org/slovník/RYTMICITA>.
- SUKAČ, Roman. 2020. Křížová cesta slovanské akcentologie. In: ŠLAUFOVÁ, Eva – SKWARSKA, Karolína – ČERMÁK, Václav (eds.). *Proměny české slavistiky po roce 1989*. Praha: Slovanský ústav AV ČR, v.v.i., 337–345.
- SUKAČ, Roman. 2021. Peculiarities of the national fishing: Slovak Rhythmic Law and related species. *Linguistica Brunensia*, 69/1, 59–80.
- ŠEFČÍK, Ondřej. 2017. Dybův zákon. In KARLÍK, Petr – NEKULA, Marek – PLESKALOVÁ, Jana (eds.) *CzechEncy – Nový encyklopedický slovník češtiny* [online]. https://www.czechency.org/slovník/DYBŮV_ZÁKON
- ŠEFČÍK, Ondřej – DUFKOVÁ, Kristýna – SUKAČ, Roman. 2014. K původu a fungování staročeského supina. *Linguistica Brunensia* 62/2, 5–15.
- TIMBERLAKE, Alan. 1983a. Compensatory lengthening in Slavic, 1: Conditions and dialect geography. In: WORTH, Dean S. – MARKOV, Vladimir (eds.), *From Los Angeles to Kiev*. Columbus: Slavica, 1983, 207–235.

TIMBERLAKE, Alan. 1983b. Compensatory lengthening in Slavic 2: Phonetic reconstruction.

In: FLIER, Michael S. (ed.), *American Contributions to the Ninth International Congress of Slavists*. Kiev – Columbus: Slavica, 293–319.

UTĚŠENÝ, Slavomír. 1960. *Nářečí přechodného pásu česko-moravského*. Praha: Nakladatelství ČSAV.

VORÁČ, Jaroslav. 1955. *Česká nářečí jihozápadní I*. Praha: Nakladatelství ČSAV.

Roman Sukač

Independent researcher

Czech Republic

romansukac1456@gmail.com



Toto dílo lze užít v souladu s licenčními podmínkami Creative Commons BY-NC-ND 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). Uvedené se nevztahuje na díla či prvky (např. obrazovou či fotografickou dokumentaci), které jsou v díle užity na základě smluvní licence nebo výjimky či omezení příslušných práv.

Petra Poukarová – Markéta Maturová

„PROTESTOVALI PŘED A V BUDOVĚ ŘECKÉHO MINISTERSTVA FINANČÍ.“ KRÁTKÉ NAHLÉDNUTÍ DO PROBLEMATIKY ZEUGMATU

ABSTRACT

In this paper, we focus on one of the deviations from sentence structure, the zeugma, specifically its frequency in different types of texts. Existing linguistic publications treat usually zeugma as an error to be avoided. In addition to defining the term, we also draw attention to ellipsis and condensation, which are related to zeugma. The frequency of its occurrence in texts will be verified by corpus analysis; the corpus used will be the SYN2020 corpus of written Czech, the ORTOFON_v2 corpus of spoken, informal Czech, the ORATOR_v2 corpus of formal texts, and the web corpus, Araneum Bohemicum Maximum. The analyzed examples illustrate that zeugma occurs in different types of texts and, thus, is not bound to one particular mode. The prevalence of zeugma in professional texts suggests that this device does not affect the clarity and precision of expression that are typical for them. This finding could open up a debate as to whether to consider the zeugma as a standard stylistic device by which language users condense their utterances.

KEYWORDS

ellipsis; condensation; corpus analysis; syllepsis; zeugma

1 Úvod

V předkládaném článku se zaměříme na jednu z odchylek od větné stavby, a to na zeugma neboli zanedbání dvojí vazby, a na jeho frekvenci v několika typech jazykových korpusů.

Dosavadní jazykové popisy hodnotí zeugma jako gramatickou chybu, jíž je třeba se vyvarovat (jak ostatně dokládá užití slov *odchylka* a *zanedbání* v samotné

definici). Toto již v definici implicitně obsažené negativní hodnocení vede v mluvní praxi k vytlačování mimo systém, spojenému přímo se sankcionováním u školních prací. Přitom jasné vymezení pojmu, jeho definování, exemplifikace v mluvnících spíše dokládají, že jde o jev do systému jazyka zapojený (zjednodušeně řečeno se tedy můžeme ptát, proč je gramatické chybě věnováno tolik pozornosti). Nabízí se tak otázka, zda na zeugma nepřestat pohlížet jako na chybu a naopak jej považovat za funkční prostředek kondenzace.

V následujícím výkladu se nejdříve zaměříme na vymezení pojmu zeugma, na obsahově blízké pojmy kondenzace a elipsy. Následně zeugma, respektive četnost jeho výskytu v textech ověříme v korpusové analýze; použity budou korpus psané češtiny SYN2020, korpus mluvené neformální češtiny ORTOFON_v2, korpus formálních mluvených textů ORATOR_v2 a korpus webový, Araneum Bohemicum Maximum.

Věříme, že zevrubné korpusové šetření toho, jak je zeugmatu v textech užito, v jakých typech textů se nachází a jaká je jeho četnost, může přispět k diskusi o změně pohledu na jeho status chyby a zařadit jej k prostředkům, jimiž uživatel jazyka zestručňuje svůj výklad, a navíc zabraňuje neúčelnému opakování slov.

2 Vymezení problematiky

Zeugma je v českých gramatikách shodně popisováno jako zanedbání dvojí vazby, tedy sprážením dvou a více výrazů výrazem jedním (např. PMČ 1995, 434; NESČ 2017, heslo zeugma). Toto „ovlivnění“ pak lze považovat za obousměrné: člen syntakticky závislý spráhuje dva a více výrazů řídicích, nicméně svůj tvar přebírá od blíže stojícího výrazu řídicího: „Zeugma vzniká v jazykovém projevu tím, že se u souřadně spojených řídicích členů, z nichž každý vyžaduje jinou vazbu, užití totožného závislého členu v tvaru určeném jen vazbou jednoho z těchto řídicích členů“ (HAVRÁNEK – JEDLIČKA 1981, 348). Jako příklad lze uvést spojení dvou sloves s různou valencí *řídít se* a *učít se* z příkladu našich předků (viz PMČ 1995, 434). Obecně jsou tyto konstrukce považovány gramatikami za chybné, formulačně nepřesné (ŠTÍCHA 2016, 141), nesplňující syntaktická pravidla, a je proto nutné je opravovat (možná formulace *řídít se příkladem našich předků a učít se z něj*).

Co se naopak napříč odbornými statěmi a gramatickými popisy může lišit, je šíře tohoto pojmu. Za zvláštní druh zeugmatu se dají považovat věty typu *Není možné a nechceme přijmout takový názor*. Dochází totiž ke stažení vět *Není možné přijmout* a *nechceme přijmout*, kdy infinitiv v první z nich je subjektem, v druhé objektovým doplněním. Zanedbání dvojí vazby není v morfoloické rovině zřejmé z toho důvodu, že infinitiv nevyžaduje rozdílné syntaktické členy (MČ 3, 412). J. Zeman chápe zeugma v širší a užší perspektivě: v užším pojetí se jedná o výše zmíněné zanedbání dvojí vazby, při němž závislý člen k sobě váže souřadně spojené výrazy řídicí, které

vyžadují každý jinou vazbu; naopak širší pojetí zeugmatu zahrnuje případy, kdy se jeden závislý výraz vztahuje ke dvěma výrazům dominantním, jež oba vyžadují stejnou vazbu (*Přečtl a odevzdal knihu*) (ZEMAN 2019, 189).

Některá ze zeugmat dle odborné literatury už společnost toleruje, např. oslovení *vážené dámy a pánové* nebo spojení *český jazyk a literatura* (např. ESČ 2002, 565, heslo zeugma; MČ 3, 412). Jako kodifikací přijímané (tedy nejen tolerované) uvádí M. Čechová slovesné vazby typu *Vystupujte a nastupujte do soupravy*, respektive *vystupování/výstup a nastupování/nástup do soupravy* (ČECHOVÁ 2017, 105).

2.1 Zeugma a sylepse

Odlišný od české situace je stav v mezinárodní jazykovědě. Analogicky k termínu zeugma užívá například anglickojazyčná odborná literatura pojmy¹ zeugma a sylepse, přičemž našemu pojetí odpovídá spíše sylepse (resp. syllepsis), ale ani toto srovnání není přesné (viz dále). Jako příklad sylepse můžeme uvést následující větu: *She exercises to keep healthy and I to lose weight*, kde sloveso *exercises* svým tvarem vykazuje shodu pouze se zájmenem *she*, nikoli i s *I*.²

Zeugma je pak v zahraničních pracích chápáno jako sémantická (ne)kompatibilita jednoho výrazu (nejčastěji slovesa) k více výrazům v jedné větě (viz např. SOLSKA 2008; FRETHEIM 2009), jako „zanedbání sémantických preferencí kolokability“ (NESČ 2017, heslo zeugma). V. Šmilauer uvádí záměrnou slovní hříčku: *Přistižený zloděj vzal svrchník, klobouk, roha, auto, z kapsy revolver a sobě život.*; všechny podtržené výrazy jsou valenčními doplněními slovesa *vzít*, ale zatímco v případě jmen *svrchník, klobouk, auto* a *revolver* se jedná o význam doslovný, v dokladu *sobě život* jde o význam metaforický (ŠMILAUER 1969, 96).

Hodnocení výskytu zeugmatu (resp. sylepse v cizojazyčných textech) v textu je také odlišné; v české tradici se jedná o chybu a nedostatek na úrovni syntaxe a o jeho stylistické platnosti a záměrnosti se spíše neuvažuje. To dokládá například poznámka z materiálu k literárněvědným pojmům *figury a tropy*: „zeugma: smíšení vazeb; běžná gramatická chyba spíš než stylistická figura.“ (Formální prostředky poesie: *figury a tropy*³). Naopak J. Hrabák pojem zeugma do svého výkladu básnických syntaktických figur zařazuje, pojem sylepse neužívá (HRABÁK 1977, 174). V zahraniční literatuře mají tyto pojmy status záměrných literárních figur, sloužících k vystupňování napětí, dynamičnosti sdělení, ke slovním hříčkám (viz např. KAMAGINA 2015).

1 Doklady, proč je účelné oba pojmy odlišovat, podává např. (HAMEED – HAMEED 2021). Ve slovníkových definicích se naopak někdy uvádí, že termíny jsou zaměnitelné (viz např. <https://www.collinsdictionary.com/dictionary/english/syllepsis>), nebo se užívají zaměnitelně, často specifikovány adjektivem (viz např. formulace hesla zeugma na Wikipedii: Grammatical syllepsis (sometimes also called zeugma); Zeugma (often also called syllepsis, or semantic syllepsis).

2 Příklad vybrán z <https://www.merriam-webster.com/dictionary/syllepsis> [cit. 2022-10-10]

3 Dostupné z https://is.muni.cz/el/1421/podzim2012/MED09/um/Figury_a_tropy_prehled.pdf [cit. 2022-17-11]

2.2 Zeugma a kondenzace

Jako příčina vzniku zeugmatu v textech se uvádí jazyková ekonomičnost, snaha zkracovat výpovědi (POSPÍŠILOVÁ 2015, 62), neopakovat jednotlivé výrazy, jež je adresát sdělení schopen doplnit na základě kontextu (viz např. DIBELKOVÁ 2021, 10). Dibelková ve své práci používá termín *syntaktická kondenzace*. Při ní skutečně dochází k zhuštění informací a k jazykové úspornosti (např. transformací vedlejší věty na větný člen, PMČ 1995, 752), ale motivací pro její vznik je snaha o co nejvíce informačně nasycený text, např. administrativní (PMČ 1995, 753). Domníváme se, že použití termínu kondenzace v souvislosti se zeugmatem není úplně přesné, a to ze dvou důvodů: zeugma je označováno jako chyba, která by měla být odstraněna, tedy nahrazena „plným zněním“, kondenzáty naopak ve výpovědi svůj účel mají a jsou zcela funkční; vzhledem k časté chybovosti a výskytu např. ve studentských pracích a k problému při jejich detekci v textech zřejmě ani nelze hovořit o záměrnosti zeugmatu.

2.3 Zeugma a elipsa

Zeugma lze tedy chápat jako snahu původců textu o jazykovou úspornost (viz výše). Principiálně se ale jedná o elipsu,⁴ jež je charakterizována jako vypuštění jednoho výrazu, který je adresáty dobře a snadno odvoditelný a „vázaný na jistý specifický kontext“ (PMČ 1995, 696), nebo „nevyjádření takových slov a vět, které podle daného větného schématu očekáváme“ (ŠMILAUER 1969, 90). Jako typický příklad se uvádí věta: *Petr kouří dýmku a jeho žena cigarety* (NESČ 2017, heslo elipsa). Úvahy o tom, zda se v konkrétních případech jedná o elipsu, nebo zeugma, tedy spíše rozhoduje to, zda jsou tyto příklady hodnoceny jako gramaticky nesprávné (zeugma; viz výše), nebo správné (elipsa; viz MIKULOVÁ 2011, 55). Nejednoznačnost obou termínů dokládá např. i již výše uvedené spojení *český jazyk a literatura*: P. Karlík tyto struktury popisuje jako „nezřídka tolerované jako gramatické“ (NESČ 2017, heslo zeugma), F. Daneš je považuje za elipsy, jež mluvnické připouštějí a které „někdy mají za následek nepravidelnosti ve větné stavbě“ (DANEŠ 1973, 107).

3 Metoda

Pro doložení výskytu zeugmatu v jazyce jsme využily několika korpusů, jež byly vytvořeny v Českém národním korpusu (konkrétně korpusů SYN2020, ORTOFON_v2, ORATOR_v2 a Araneum Bohemicum Maximum), a to z několika důvodů:

- a) Korpusy představují soubor autentických textů, jež dokládají, jak se jazyk skutečně užívá; jedná se tedy o vybraný vzorek jazyka. Zajímalo nás, zda je zeugma v běžném úzu časté.

4 V. Šmilauer používá spojení „násilná elipsa“.

- b) Lze předpokládat, že většina textů, které jsou součástí korpusu SYN2020, prošla jazykovou redakcí a korekturou, a jelikož se jedná o texty psané, je možné o nich uvažovat jako o textech připravených (na rozdíl od spontánních hovorů). Nabízí se tedy otázka, jak zeugma hodnotit z hlediska gramatické správnosti, když i odborníci jejich výskyt v textech tolerují a považují za účelný. (Druhá otázka je, zda tento jev vůbec v textech identifikují.)
- c) Je-li zeugma charakterizováno jako prostředek jazykové ekonomie, lze uvažovat o tom, zda jeho původ není ve spontánních mluvených textech, kde je úspornost vyjádření jedním z rysů (viz např. ZEMAN 2019, 191; spontánní hovory jsou lineární, nejsou předem připravené, ale vznikají v čase, jsou vždy ukotveny v nějaké konkrétní situaci, ve velké většině případů se jedná o komunikaci tváří v tvář, nejsou „upravitelné a opravitelné“, co jednou zaznělo, už nelze vzít zpátky).
- d) Mluvené texty částečně připravené a texty webové se nacházejí na škále mezi spontánností mluvení a připraveností psaných komunikátů (v teorii Kocha a Oesterreichera se v případě prvního typu textů jedná o texty mediálně mluvené a konceptuálně psané, u druhých zmíněných naopak o mediálně psané, ale konceptuálně mluvené; KOCH – OESTERREICHER 2012). Vyhledávat případy zeugmatu v těchto korpusech bylo motivováno především snahou pokrýt škálu typů textů, v nichž se tento jev může objevit.

Výchozím zdrojem dat byl korpus psaných textů SYN2020. Jedná se o reprezentativní referenční korpus o velikosti 100 milionů slov, tvořený texty z let 2015–2019. Korpus SYN2020 je lematizovaný (každé formě – včetně interpunkce – je přiřazena reprezentativní slovníková podoba) a morfologicky (obsahuje jednoznačnou gramatickou informaci pro každou formu obsaženou v korpusu) a syntakticky značkováný (jsou v něm závislostní vztahy mezi slovy ve větě a skladební funkce jednotlivých slov). Z hlediska reprezentativnosti⁵ je korpus SYN2020 složen z textů oborových,⁶ publicistických a beletristických, všechny jsou zastoupeny stejným poměrem (33,33 %).⁷

Dalším zdrojem dat pro vyhledávání byl ORTOFON_v2, více než milionový korpus spontánní, dialogické mluvené češtiny, jež je užívána mezi lidmi, kteří se dobře znají, v jejich přirozeném prostředí. Téma hovoru, jeho délka ani forma sdělení nebyly nijak korigovány, při sběru dat byl kladen důraz především na přirozenost, autentičnost komunikace (nahrávání nevěděli, že k nahrávání dochází, aby žádným způsobem neměnili svoje mluvní zvyklosti). Korpus je lematizovaný

5 K pojmům reprezentativnost a vyváženost viz např. CVRČEK et al. (2016).

6 Oborová literatura zahrnuje odborné texty vědecko-naučné i popularizující, učební texty a profesní literaturu.

7 Zevrubné informace o všech relevantních korpusech viz <https://wiki.korpus.cz/doku.php/cn-k:uvod>.

a morfoloicky značkováný, jeho specifikem je dvouúrovňový přepis (jednak na rovině ortografické, jednak na rovině fonetické).

Monologické mluvené texty pocházejí z více než milionového korpusu ORATOR_v2. Jedná se o korpus projevů pronesených při zvláštních příležitostech, o kterých jsou mluvčí předem informováni. Nahrávky obsahují přepisy přednášek, úvody různých společenských akcí, přípitky, uvítání hostů, vyhlášení výsledků soutěží apod. Všechny tyto situace lze hodnotit jako formální, až oficiální, faktory zásadní při jejich produkci tu jsou předem daný časově vymezený rámec (přednáška obvykle trvá 90 minut, přípitek několik desítek vteřin), možnost, nebo dokonce vhodnost předchozí přípravy, volba lexika dle velikosti a zaměření publika atd. Jedná se tedy o texty na pomezí mluvenosti a psanosti (k pojmům viz výše). Stejně jako korpus ORTOFON_v2 je i ORATOR_v2 lemmatizovaný a morfoloicky značkováný, přepis je pouze jednoúrovňový, ortografický.

Pro úplnost jsme využily i data obsažená ve více než miliardovém korpusu Araneum Bohemicum Maximum,⁸ složeném z textů internetových domén s koncovkami .cz, .sk, .com, .edu, .info, .org, .net. Korpus je lemmatizovaný a morfoloicky značkováný.

4 Zeugma v jazykových korpusech

Před samotnou kvantitativní analýzou bylo nutné rozhodnout, jaký typ zeugmatu bude vyhledáván. Vzhledem k úspěšnosti dotazu byl zvolen typ: *dvě souřadně spojené předložky sprážené jedním substantivem (před a po zápase).*⁹

Díky morfoloickému značkování, obsaženému ve všech zvolených korpusech, bylo možné vyhledávat relevantní doklady pomocí komplexnějšího dotazu, jež zde uvádíme z důvodu replikovatelnosti výzkumu:

```
[tag="R.*"][word="a|i|ani|a?nebo|či"][tag="R.*"][tag!="N.*"]*[tag="N.*"] within <s/>10
```

Uvědomujeme si, že dotaz by bylo možné ještě detailněji specifikovat, nicméně nejde nám o vyhledání úplně všech dokladů tohoto typu zeugmatu, ale o zdokumentování, zda se v jednotlivých korpusech vyskytuje a jaká je jeho forma.

8 Zevrubné informace o korpusech rodiny Araneum viz http://ucts.uniba.sk/aranea_about/index.html

9 Dalšími možnými typy zeugmatu byly: souřadně spojené slovesné tvary sprážené jedním předmětem, syntakticky závislým na druhém slovese (typ *nesledují a nezabývají se tím*) nebo přívlastek shodný spráhující dvě souřadně spojená substantiva (typ *český jazyk a literatura*).

10 Tagem R označujeme libovolnou předložku, po ní musí následovat jedna ze slučovacích spojek a, i, ani, nebo, anebo, či, následuje další libovolná předložka. Koncový tag, N, označuje substantivum, jemu předcházející pozice specifikuje, že mezi druhou předložkou a koncovým substantivem mohou být ještě další pozice (např. adjektiva ve funkci shodného přívlastku, ta mohou být rozvíta adverbii), platí ovšem podmínka, že tato pozice nebude obsazena jiným substantivem. Celá tato konstrukce se musí vyskytovat v rámci jedné věty, u mluvených korpusů v rámci jedné repliky (viz podmínka within).

4.1 Zeugma v korpusu SYN2020

V korpusu SYN2020 jsme na základě výše uvedeného dotazu vyhledaly celkem 880 dokladů, které jsme ručně vytrídily a odstranily ty, které nesplňovaly parametry zeugmatu. K nejčastějším nerelevantním příkladům patřila frazeologická spojení *pro a/i proti* nebo *před a po*, k dalším, již méně frekventovaným, se řadily případy s chybnou lemmatizací (např. zkratka *pro* sekundu, *s*, byla chybně automaticky označena jako předložka, homonymní výraz *bez* pro označení rostliny taktéž), spojení *od a do zet* apod. Konečný počet výskytů tedy byl 358 zeugmat. Ve více než dvou třetinách případů (ve 266 z 358, tedy v 74%) se jednalo o zeugma, kdy obě předložky vyžadovaly stejný pád substantiva: např.

- (1) *Předmět zájmu si lze představit jako víceméně ohraničený prostor, **do a ze kterého** vedou různé vstupy a výstupy.*¹¹
- (2) **Před a za zadní nápravou** se nacházejí lithium-iontové akumulátory.
- (3) **Pohranicím** je území nacházející se **podél nebo blíže státní hranice**.

V 92 dokladech zeugmatu vyhledaných v korpusu SYN2020 se jedná o případy, v nichž předložky vyžadují každá jiný pád:

- (4) *se měla v úmyslu vyhnout všem policejním hlídkám, **které před a v průběhu sjezdu** kontrolovaly*
- (5) *protestovali **před a v budově** řeckého ministerstva financí*

Množina typů předložek, jež tato zeugmata tvoří, je víceméně uzavřená, nejčastější kombinace jsou *do* [a/i/ani/nebo/anebo/či] z (28krát); *nad* [a/i/ani/nebo/anebo/či] *pod* (55krát); *před* [a/i/ani/nebo/anebo/či] *za* (57krát); *s* [a/i/ani/nebo/anebo/či] *bez* (61krát), tedy klasické primární předložky, obecně velmi frekventované.

Naopak nejpestřejší kombinace se objevují s předložkou *prostřednictvím* (celkem 4krát je tato předložka na prvním místě, následují tři různé předložky na pozici druhé) a *navzdory* (celkem 2krát na prvním místě, následují dvě různé předložky na pozici druhé): např. *prostřednictvím a na základě / ve jménu / za pomoci; navzdory a vně režimu*. Obě slova jsou přitom příznaková. Výraz *prostřednictvím* je dokladem sekundární předložky, prepozice *navzdory* je v SSČ charakterizována jako knižní. Lze tedy uvažovat o tom, že i synsémantické prostředky, k nimž předložky řadíme, vykazují svoji kolokabilitu (k tomu srov. následující poznámku: „O kolokabilitě se na úrovni jiné než lexikální (autosémantické) zpravidla neuvažuje, i když tomu např. u morfémů nic nebrání“; NESČ 2017, heslo kolokabilita).

¹¹ Všechny uvedené doklady z korpusu citujeme tak, jak se skutečně v korpusu nalézají, tedy někdy jako celou větu s velkým počátečním písmenem a koncovou interpunkcí, někdy jako samotnou klauzi bez koncového znaménka (délku příkladů volíme tak, aby nebyla na úkor srozumitelnosti). Jedinou změnou oproti podobě v korpusu je psaní interpunkce bez předcházející mezery.

Zajímavé jsou i doklady následujícího typu:

- (6) **Nad** i **pode mnou** se vlnila stočená klubka šroubovice
- (7) která mají údajně zabránit rozšíření požáru **v** nebo po **fasádním** systému.
- (8) Termomechanické válcování **s** a bez **samopopouštění** (QST) dovoluje vyrobit jemnozrnné oceli.

V případě sekvence *první předložka – jméno* (obojí tučně) by bylo namístež užít vokalizované podoby těchto prepozic; pro jejich výskyt v zeugmatu tato pravidelnost ale neplatí. To vede k úvahám o procesu jazykového plánování, ale také zdůrazňuje to, že nejen texty mluvené, ale i psané jsou lineární. K tomu srov. následující dvě citace:

„Výklad souvisí s povahou spontánního mluveného projevu. Vyložili jsme již, že tu běží o proces (...) význačně se lišící znakem lineárnosti a nepřerušovaného časového plynutí od projevů psaných“ (ZIMA 1954, 158);

„S plynutím textu v čase souvisí i linearita, která bývá zdůrazňována zejména pro text mluvený. Princip linearity je vlastní oběma druhům textu“ (MÜLLEROVÁ 1989, 208).

U příkladů zeugmat vyskytujících se v korpusu SYN2020 bylo dále zjišťováno, jak jsou frekvenčně zastoupena ve třech hlavních textových typech, a to beletrii, oborové literatuře a publicistice (viz výše k textům obsaženým v korpusu SYN2020). Původní předpoklad byl, že nejčastěji se zeugma objeví v beletrii nebo publicistice. Beletrie obsahuje pásmo postav a velké množství přímé řeči, zeugmatem by tak autoři mimo jiné posílili vliv mluvenosti. V publicistice by pak zeugma mohlo být výsledkem většího časového tlaku na úkor jazykové přesnosti (noviny jsou aktuální, čas na jejich přípravu je omezenější, než je tomu u redakce knihy, navíc jsou texty vázány rozsahem, a někdy je tak zapotřebí značná úspornost ve vyjadřování). Naopak jsme očekávaly, že zeugma se nebude hojně vyskytovat v oborové literatuře, která musí být (nejen) po jazykové stránce jasná, precizní, nedovolující dvojí výklad.

Zeugma se oproti původním domněnkám nejčastěji objevilo, a to s výraznou převahou, právě v oborové literatuře: celkem 242× (v populárně-naučné celkem 101×, ve vědecko-naučné pak 82×), následuje publicistika s 85 výskyty a beletrie s 31 případy. Z toho vyplývá několik závěrů:

- A) Tento typ zeugmatu je skutečně možné považovat za výsledek záměrné kondenzace, jejímž cílem je zhuštění informací a informační nasycenost, typická pro odborné vyjadřování.
- B) Zeugma v uvedených dokladech nemá vliv na srozumitelnost, jasnost, přesnost výkladu, které jsou pro texty tohoto typu stěžejní.

4.2 Zeugma v korpusech mluvené češtiny

V korpusu spontánní mluvené češtiny ORTOFON_v2 bylo nalezeno 17 dokladů, přičemž o zeugma se jednalo pouze ve dvou z nich:

- (9) *nevím jestli to měla **před nebo po štěňatech***
 (10) *nebo třeba nějaký zastávky **okolo nebo z metra***

Zbylé doklady tvořily ustálené formulace *pro a proti, kolem a kolem* nebo případy opakování, které je způsobeno předchozí nepřipraveností hovoru:

- (11) *to ono to ani nejde vidět **ani na . na ani na jedné** z těch fotek¹²*

Toto opakování se nezřídka děje i se spojovacím výrazem, opakováním mluvčí získává čas na promyšlení další formulace a přitom posluchači signalizuje, že mu ještě nepředává slovo. Užítí spojovacího výrazu mezi dvěma opakujícími se předložkami může vést k významové/postojové modifikaci výpovědi, jak to ukazuje př. číslo (12). Takto užité výrazy by tedy bylo možné hodnotit spíše jako částice než konjunkce:

- (12) *furt prostě třeba vo tom **v ani v** tý lavici *nevím už**

Korpus ORATOR_v2 obsahuje pouze 14 dokladů, jeden z nich je zeugma:

- (13) *koncentrace vodíkových iontů **nad a pod biomembránou***

Ve zbylých příkladech jde, stejně jako u korpusu ORTOFON_v2, nejčastěji o opakování, jež je výsledkem nedostatečné mluvní perspektivy či nervozity (př. (14), výskyt výplňkových slov, podtrženo), objevují se i samoopravy sdělení (př. (15), oprava tučně).

- (14) *vlastně přeci nejhorsí je jak to . že to dítě **za a v tý škole** teda jako ...*

- (15) *zda pro tyto účely byla tahleta socha . zhotovena **a na a před palácem** průmyslovým hned u vstupu na výstaviště umístěna.*

4.3 Zeugma ve webovém korpusu

Po zadání dotazu bylo ve webovém korpusu Araneum Bohemicum Maximum nalezeno celkem 41 411 výskytů; pracovaly jsme s náhodným vzorkem o velikosti 300 výskytů, po ruční kontrole jsme získaly 184 zeugmat (61 %):

- (16) *V období **kolem a po pařížské výstavě.***

Na rozdíl od korpusu psané češtiny SYN2020 obsahuje Aranea sice více typů předložek (zdaleka nejčastější kombinací je *před [a/i/ani/nebo/anebo/či] po* – 61krát), ale

12 Tečka v citovaném příkladu značí krátkou pauzu, interpunkce v přepisu není užito syntakticky.

výskyty zeugmatu se v obou korpusech obecně nijak zásadně neliší. Za pozornost by snad stály dva nalezené příklady:

(17) *Ale posluchači naprosto nic neříká před ani po ústním sdělení*

(18) *jak nejlépe pomoci mamince před a během vlastního porodu.*

V obou případech je jméno, jež spřahuje obě předložky, rozvito ještě adjektivem. Odlišnosti v obou spatřujeme v tom, jak by vypadala potenciální syntakticky bezchybná konstrukce: zatímco v č. (17) by to bylo *před ústním sdělením ani po ústním sdělení*, v příkladu (18) pak pravděpodobně *před porodem a během vlastního porodu*, chápeme-li spojení *vlastní porod* jako porod právě probíhající. Jde tu tedy o jakýsi dvouúrovňový vztah, kdy obě předložky nejsou na stejné úrovni, a jejich pořadí by tedy nebylo zaměnitelné (když nebudeme uvažovat to, že svou sémantikou předložky *před, po a během* značí časovou následnost): je možné formulovat *po ani před ústním sdělením*, ale už ne *během vlastního a před porodem*. Po stránce formální jsou doklady č. (17) a (18) tedy zeugmaty, ale po stránce významové by se v druhém z nich o spřažení dvou výrazů výrazem jedním de facto nemuselo jednat a příkládněly bychom se spíše k lipse.

5 Závěr

Tématem předkládaného článku je výskyt zeugmat v textech, jež jsme doložily analýzou korpusových dat (pracovaly jsme s psaným korpusem SYN2020, webovým Araneum Bohemicum Maximum a s korpusem mluvenými ORTOFON_v2 a ORATOR_v2). Zeugma je v české lingvistice považováno za chybu, již je nutné se vyvarovat, nicméně některé případy jsou gramatikami již tolerovány (viz např. spojení *vážené dámy a pánové*).

Korpusová analýza prokázala, že v mluvených a webových textech je výskyt zeugmatu ojedinělý, v textech psaných (přípravených) se nachází častěji. Při třídění výsledků jsme zaznamenaly nejvyšší frekvenci v textech odborných, nejnižší v beletrii.

Protože odborný text klade nároky na jednoznačnost, přesnost, objektivnost sdělení, lze na základě vysoké frekvence zeugmatu usuzovat, že jeho užití stylizací komunikátu neovlivňuje, je ho užito funkčně (plné znění, realizované například opakováním slov, může v některých případech působit neúčelně a stylisticky nevhodně), recipováno shodně se záměrem autora (odborné texty jsou dále citovány, jsou recenzovány apod.) a je možné jej považovat za prostředek jazykové kondenzace. Korpusové doklady zároveň poukazují na tendence v užívání jazykových prostředků, jejich posun z oblasti úzu do oblasti kodifikace. Domníváme se proto, že zmíněné argumenty podporují úvahu považovat zeugma za standardní stylový prostředek, a ne za syntaktický nedostatek.

Jeho zařazení do systému jazyka (tedy jeho uznání jako gramaticky správného) navíc může nastolit nové otázky, týkající se například jazykového plánování či specifictví analýzy typů předložek v této konstrukci.

6 Diskuse

Využití korpusů při analýze zeugmatu nabízí možnost prozkoumat tento jev nejen v různých módech češtiny, ale i napříč žánry, získat informaci o jejich frekvenci a rozložení napříč texty. Na druhou stranu je taková analýza vždy limitována způsoby vyhledávání relevantních dokladů. V tomto textu jsme se zaměřily na typ zeugmatu v předložkovém spojení. Podrobnější analýzy by bylo třeba ale i v případech zbylých dvou uváděných typů zeugmatu. Lze předpokládat, že souřadně spojené slovesné tvary spřažené jedním předmětem, syntakticky závislým na druhém slovese (typ *nesledují a nezabývají se tím*), nebudou překážkou ve srozumitelnosti sdělení, i když druhou stranu mince, tedy frekvenci tohoto typu a jeho rozložení v textech, neznáme. Naopak u přívlastku shodného spřahujícího dvě souřadně spojená substantiva (typ *český jazyk a literatura*) nemusí panovat na straně autora a recipienta shoda – může jít stejně dobře o zeugma (ve významu *český jazyk a česká literatura*), jako o plné znění (ve významu *český jazyk a jakákoli literatura*). Přesto jsou některé z těchto případů uznávány jako gramatické (viz výše). Proto se domníváme, že za gramatické by se měly považovat všechny tři typy zeugmatu.

LITERATURA

- CVRČEK, Václav – ČERMÁKOVÁ, Anna – KŘEN, Michal. 2016. Nová koncepce synchronních korpusů psané češtiny. *Slovo a slovesnost*, 77(2), s. 83–101.
- ČECHOVÁ, Marie. 2017. *Život s češtinou*. Praha: Academia.
- DANEŠ, František. 1973. O jednom typu elipsy. *Naše řeč*, 56(2), s. 107–108.
- DANEŠ, František – GREPL, Miroslav – HLAVSA, Zdeněk, eds. 1987. *Mluvnice češtiny: 3. Syntax*. Praha: Academia.
- DIBELKOVÁ, Monika. 2021. *Syntaktické defekty v písemných pracích středoškolských studentů* [online]. Plzeň: FPE ZČU [cit. 2021-08-11]. Plzeň. Diplomová práce. Západočeská univerzita v Plzni, Fakulta pedagogická. Katedra českého jazyka a literatury.
- FRETHEIM, Thorstein. 2009. The Borderline Area between Zeugma and Normal Usage. In: FRASER, Bruce – TURNER, Ken, eds. *Language in Life, and a Life in Language: Jacob Mey – A Festschrift*. Bingley: Emerald Group Publishing Limited, s. 97–103.
- HAMEED, May Tahseen – HAMEED, Hind Tahseen. 2021. The Employment of Zeugma and Syllepsis in Adages and Famous Quotes as a Kind of Discourse Genera. *Humanities and Social Sciences Reviews*, 9(4), s. 167–174.
- HAVRÁNEK, Bohuslav – JEDLIČKA, Alois. 1981. *Česká mluvnice*. Praha: SPN.

- HRABÁK, Josef. 1977. *Poetika*. Praha: Československý spisovatel.
- KAMAGINA, Irina Vasiljevna. 2015. Ellipsis and Similar Syntactic Phenomena in the Modern Russian Language. *RUDN Journal of Language Studies, Semiotics and Semantics*, 6(1), s. 168–174.
- KARLÍK, Petr – NEKULA, Marek – PLESKALOVÁ, Jana, eds. 2002. *Encyklopedický slovník češtiny*. Praha: NLN.
- KARLÍK, Petr – NEKULA, Marek – PLESKALOVÁ, Jana, eds. 2012–2018. *Nový encyklopedický slovník češtiny* [online]. Brno: Masarykova univerzita. Dostupný z: <https://www.czechency.org/>.
- KARLÍK, Petr – NEKULA, Marek – RUSÍNOVÁ, Zdenka, eds. 1995. *Příruční mluvnice češtiny. Syntax*. Praha: NLN.
- KOCH, Peter – OESTERREICHER, Wulf. 2012. Language of Immediacy – Language of Distance: Orality and Literacy from the Perspective of Language Theory and Linguistic History. In: LANGE, Claudia – WEBER, Beatrix – WOLF, Göran, eds. *Communicative Spaces: Variation, Contact, and Change: Papers in Honour of Ursula Schaefer*. Frankfurt am Main: Peter Lang, s. 441–474.
- MIKULOVÁ, Marie. 2011. *Významová reprezentace elipsy*. Praha: ÚFAL MFF UK.
- MÜLLEROVÁ, Olga. 1989. Ke vztahu mluveného a psaného textu (srovnání mluveného a psaného vypravování). *Slovo a slovesnost*, 50(3), s. 205–216.
- NOVOTNÝ, František. 1957. *Základní latinská mluvnice*. Praha: SPN.
- POSPÍŠILOVÁ, Lucie. 2015. *Analýza stylu*. [online]. Brno: FF MU [cit. 2021-08-11]. Brno. Bakalářská práce. Masarykova univerzita. Filozofická fakulta. Ústav českého jazyka.
- SOLSKA, Agnieszka. 2008. Accessing Multiple Meanings: The Case of Zeugma. In: MIODUSZEWSKA, Ewa – PISKORSKA, Agnieszka, eds. *Relevance Round Table I*. Warsaw: Warsaw University Press, s. 109–122.
- ŠMILAUER, Vladimír. 1969. *Novočeská skladba*. Praha: SPN.
- ŠTÍCHA, František. 2016. *O nepřesnosti našeho vyjadřování*. Praha: Academia.
- ZEMAN, Jiří. 2019. Některé specifické syntaktické konstrukce příznačné pro mluvenou češtinu. In: HOFFMANNOVÁ, Jana – HOMOLÁČ, Jiří – MRÁZKOVÁ, Kamila, eds. *Syntax mluvené češtiny*. Praha: Academia, s. 187–197.
- ZIMA, Jaroslav. 1954. K stylistice mluveného projevu. *Naše řeč*, 37(3–6), s. 154–165.

KORPUSY

- SYN2020: Korpus SYN [online; verze 1] (2020). Praha: Ústav Českého národního korpusu FF UK. Cit. 20. 11. 2022. <www.korpus.cz>.
- ORTOFON_v2: Korpus ORTOFON [online; verze 2] (2020). Praha: Ústav Českého národního korpusu FF UK. Cit. 20. 11. 2022. <www.korpus.cz>.
- ORATOR_v2: Korpus ORATOR [online; verze 2] (2020). Praha: Ústav Českého národního korpusu FF UK. Cit. 20. 11. 2022. <www.korpus.cz>.
- ARANEUM BOHEMICUM MAXIMUM: Korpus Aranea [online] (2015). Praha: Ústav Českého národního korpusu FF UK. Cit. 20. 11. 2022. <www.korpus.cz>.



Petra Poukarová

Institute of the Czech National Corpus
Charles University in Prague
nám. Jana Palacha 2, 116 38 Praha 1
Czech Republic
petrapoukarova@gmail.com

Markéta Maturová

Institute of Czech Studies
University of South Bohemia in České Budějovice
Branišovská 31a, 370 05 České Budějovice
Czech Republic
mmaturova@ff.jcu.cz



Toto dílo lze užít v souladu s licenčními podmínkami Creative Commons BY-NC-ND 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). Uvedené se nevztahuje na díla či prvky (např. obrazovou či fotografickou dokumentaci), které jsou v díle užity na základě smluvní licence nebo výjimky či omezení příslušných práv.

Patricia de Bernardo Stempel:

The Accents of Celtic: New Light on the Older and Oldest Stages.

Heidelberg: Universitätsverlag Winter. 2023. XVIII + 430 s.

ISBN 978-3-8253-9503-2

Zadání této zajímavé knihy, na niž zde chceme stručně upozornit, je jasné: chce zjistit slovní akcentuaci v nejstarších fázích keltských jazyků; protože tu není k dispozici ani svědectví textové v podobě grafému označujících přízvukování a pochopitelně už vůbec ne svědectví zvukové, je třeba se obrátit ke svědectví nepřímému, totiž k efektům přízvuku v segmentální stavbě slova. Oč je zadání jasnější, o to je ovšem složitější jeho provedení. Autorka, profilovaná keltoložka, postupuje následovně.

V úvodní kapitole stručně představuje problematiku přízvukování obecně a v keltských jazycích zvlášť, definuje korpus keltských jazykových fragmentů, na němž své téma bude zkoumat, a načrtává obsah knihy (*I Introduction*, s. 1–12). V následující kapitole blíže líčí podoby přízvukování v keltských jazycích s výhledem k hlavní otázce knihy, totiž povaze prehistorického keltského přízvuku, dosud pojednané kontroverzně (*II Patterns of accentuation in Indo-European and Celtic*, s. 13–46). Ve stručné třetí kapitole koncizně definuje podstatu nástroje svého zkoumání, tedy to, že to, zda daná slabika je přízvukovaná nebo ne, lze identifikovat podle toho, jakou má segmentální stavbu (*III Accent without accent marks and stress position indicators*, s. 47–49). V následující jádrové kapitole pak autorka jeden po druhém tyto segmentální indikátory probírá: monoftongizaci, diftongizaci, synkopu, epentezi, asimilaci a redukci vokálů, degeminaci a geminaci konsonantů, zánik konsonantů a zjednodušení konsonantických skupin, metatezi likvid (*IV The Continental Celtic stress-position indicators one by one*, s. 51–190). K této jádrové kapitole se druží kapitola pátá, v níž se probírají doklady vykazující několik ze zmíněných indikátorů synchronně (*V Multiple stress-position indicators in Continental Celtic*, s. 191–241), a šestá, v níž se probírají slova vykazující několik těchto indikátorů diachronně, a tak ukazujících na určitou jazykovou změnu (*VI Biphasic evidence of stress in Continental Celtic*, s. 243–301). Sedmá kapitola analyzuje řečené indikátory v toponymech (*VII The stress system in modern place-names of Celtic origin*, s. 303–337). Osmá kapitola pak přináší závěry, tedy především konstataci, že v nejstarších fázích keltských jazyků nebyl přízvuk na první slabice (*VIII Conclusions*, s. 339–345).

Cena této knihy je dvojí: věcná a metodologická. Věcnou stránku musejí blíže posoudit keltologové; nám se zdá být v pořádku. Metodologická cena spočívá v tom, že kniha předvádí, jak užít jeden dobrý obecný deskriptivně-explikativní nástroj. Jeho aplikace je ovšem podmíněna apriorním předpokladem, že určité změny v segmentální stavbě slov jsou indukované přízvukem. Tu si pak můžeme položit dekonstrukční otázku, zda tento předpoklad ve skutečnosti není než předsudek mluvčích moderních germánských jazyků, kteří stáli u kolébky historicko-srovnávací jazykovědy.



Bohumil Vykypěl

Czech Language Institute of the CAS, v. v. i.

Veveří 97, 602 00 Brno

Czech Republic

vykypel@ujc.cas.cz

72 / 2024 / 1

RECENZE A ZPRÁVY - REVIEWS AND REPORTS



Toto dílo lze užít v souladu s licenčními podmínkami Creative Commons BY-NC-ND 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). Uvedené se nevztahuje na díla či prvky (např. obrazovou či fotografickou dokumentaci), které jsou v díle užity na základě smluvní licence nebo výjimky či omezení příslušných práv.

Marta Eva Běťáková – Václav Blažek:

Lexicon of Baltic Mythology.

Heidelberg: Universitätsverlag Winter 2021. 290 s. Empirie und Theorie der Sprachwissenschaft, 7.

ISBN 978-3-8253-4866-3

Tato kniha, o níž zde chceme podat krátkou zprávu, je aktualizovaný anglický překlad českého originálu, který vydalo nakladatelství Libri v roce 2012. Obsah je prostý. Nejdříve se ve stručné předmluvě píše o účelu publikace, jímž je podrobně pojednat zanedbávanou oblast baltské mytologie, definuje se, co se myslí slovem baltský, a stručně se načrtávají dějiny baltských zemí a národů (*Preface*, s. 7–20). Pak se slovníkovou formou a z hlediska historicko-srovnávací jazykovědy komplexně analyzují – propriální i apelativní – výrazy baltské mytologie (s. 23–242). Knihu uzavírá obvyklý bibliografický aparát a některé poučné textové přílohy (s. 243–290).

Relativně hojné recenze už po zásluze ocenily český originál i anglický překlad a hojně diskutovaly také jednotlivosti diskuse potřebné (srov. poučné zejm. BALSYS 2015; VAITKEVIČIENĖ 2022; KESSLER 2023). I my knihu oceňujeme, a to zejména proto, že se v anglickém hávu dostane k širšímu publiku, jak si to zaslouží. Místo diskuse jednotlivostí, které nám při četbě padly do oka, přidáme jen krátkou obecnou úvahu. Lze si položit otázku, zda představa o existenci baltské mytologie není iluzivní: existoval nějaký takový systém, byť i fluidní, nebo je to výsledek moderního pořádacího rozumu, připisujícího něco, co bylo vlastní sociálním komunitám se sofistikovanou organizací (např. Římanům), mylně i komunitám bez takové organizace?

Kromě toho si nemůžeme odpustit drobnou poznámku, že Husova nebo někoho jiného reforma českého pravopisu spíše než příspěvek českých zemí Evropě a lidstvu byl nešťastný nápad, jak to ukazuje i marný zápas nakladatelství s příjmením spoluautorky: v minuskulích to ještě zvládlo, ale v majuskulích místo Ť důsledně stojí T’.

LITERATURA

- BALSYS, Rimantas. 2015. Encyclopedia of Baltic mythology in Czech, or “As some sleep, others must keep vigil...”. *Slovene = Словъне. International Journal of Slavic Studies* 4(2), pp. 247–254.
- KESSLER, Stephan. 2023. Eva Běťáková, Václav Blažek: *Lexicon of Baltic Mythology*. *Zeitschrift für Ostmitteleuropa-Forschung* 72(1), pp. 89–91.
- VAITKEVIČIENĖ, Daiva. 2022. Marta Eva Běťáková, Václav Blažek, *Lexicon of Baltic Mythology*. *Baltistica* 57(2), pp. 345–353.



Bohumil Vykypěl

Czech Language Institute of the CAS, v. v. i.

Veveří 97, 602 00 Brno

Czech Republic

vykypel@ujc.cas.cz

72 / 2024 / 1

RECENZE A ZPRÁVY - REVIEWS AND REPORTS



Toto dílo lze užít v souladu s licenčními podmínkami Creative Commons BY-NC-ND 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). Uvedené se nevztahuje na díla či prvky (např. obrazovou či fotografickou dokumentaci), které jsou v díle užity na základě smluvní licence nebo výjimky či omezení příslušných práv.

<https://doi.org/10.5817/LB2024-37773>

Eduard Mühle:

Slavs in the Middle Ages between Idea and Reality.

Leiden – Boston: Brill 2023. XVII + 610 s. East Central and Eastern Europe in the Middle Ages, 450–1450, 89.

ISBN 978–90–04–45025–7

Eduard Mühle:

Slované ve středověku. Mezi ideou a skutečností.

Praha: Argo 2023. 568 s. Každodenní život, 85.

ISBN 978–80–257–3939–6

Arnold Suppan:

1000 Jahre Nachbarschaft. „Österreicher“, „Tschechen“ und „Sudetendeutsche“.

Wien: Verlag der Österreichischen Akademie der Wissenschaften 2023. 987 s.

ISBN 978–3–7001–9256–5

Nedávno jsme v tomto časopise referovali – převážně negativně – o německé knize Eduarda Mühleho o Slovanech ve středověku (srov. VYKYPĚL 2021). S překvapením jsme zjistili, že dvě nakladatelství, označovaná obvykle za renomovaná, vydala její anglickou, resp. českou mutaci. Obsahově a koncepčně obě zůstaly stejné, a tedy obsahují i vše, co jsme kritizovali. Přirozeně jsme ani nečekali, že by autor – také pro patrnou jazykovou bariéru – na naši recenzi reagoval, ale alespoň české vydání mohlo opravit jeho nikoli řídké chyby ve slovanských slovech, manifestující řečenou jazykovou bariéru. Ale abychom přidali něco pozitivního, zmíníme, že český překlad jednoho místa, na němž se manifestují některé autorovy stereotypní představy, alespoň poskytuje námět pro budoucí bádání: autor tu v chorvatském kontextu rozvádí, že obecný atribut *slovanský* se vztahoval na selské obyvatelstvo, zatímco specifické výrazy označovaly vyšší, příp. městské vrstvy; v německém originále je mezi těmito výrazy také *magyarisch*, za něž český překlad má neadekvátně *uherský* (s. 355); historie užívání termínů *ungarisch* – *magyarisch*, příp. *uherský* – *maďarský* atd. je potřebná podrobného zpracování stejně jako její analoga jako *böhmisch* – *tschechisch* nebo dvojí polské *litewski*.

Znovu vydat svůj opus se ovšem nerozhodl jen Eduard Mühle, ale i jiný autor, o němž jsme podobně kriticky referovali jinde (srov. VYKYPĚL 2019), totiž Arnold Suppan. Svému vylíčení paralelních dějin Čechů a Rakušanů změnil podtitul: původně tam stálo „*Tschechen*“ und „*Österreicher*“ in *historischer Perspektive. Eine Synthese*; teď přibyli ještě sudetští Němci, ovšem stejně nedefinovaní jako Rakušané a Češi, a naopak zmizel příliš ambiciózní atribut *Eine Synthese*. Tu a tam autor přidal různě rozsáhlé nové pasáže, případně odkazy na literaturu. Ale celkově se jeho pojednání nese ve stejném duchu. Zůstaly silné kvalitativní i kvantitativní disproporce i všechny jednotlivosti, které jsme vytýkali. Stejný tak zůstává i náš závěr: nejzajímavější je tato kniha metahistoriograficky jako příklad pohledu na česko-německé vztahy, který bychom mohli nazvat umírněně císařsko-rakouský.

LITERATURA

VYKYPĚL, Bohumil. 2019. Arnold Suppan: 1000 Jahre Nachbarschaft. „Tschechen“ und „Österreicher“ in historischer Perspektive. *Eine Synthese. Germanoslavica* 30(2), pp. 110–112.

VYKYPĚL, Bohumil. 2021. Eduard Mühle: *Die Slawen im Mittelalter zwischen Idee und Wirklichkeit. Linguistica Brunensia* 69(2), pp. 84–87.

Bohumil Vykypěl

Czech Language Institute of the CAS, v. v. i.

Veveří 97, 602 00 Brno

Czech Republic

vykypel@ujc.cas.cz



Toto dílo lze užít v souladu s licenčními podmínkami Creative Commons BY-NC-ND 4.0 International (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). Uvedené se nevztahuje na díla či prvky (např. obrazovou či fotografickou dokumentaci), které jsou v díle užity na základě smluvní licence nebo výjimky či omezení příslušných práv.

Linguistica Brunensia

72 / 2024 / 1

(<https://journals.phil.muni.cz/linguistica-brunensia>)

Vydala Masarykova univerzita roku 2024

Redakční rada / Editorial board

Předseda & editor: prof. RNDr. Václav Blažek, CSc.

Editoři: prof. Mgr. Pavel Kosek, Ph.D., Mgr. Pavel Krejčí, Ph.D.,

PhDr. Ondřej Šefčík, Ph.D., a Mgr. Marcin Waگیel, Ph.D.

Členové rady: prof. Neil Bermel, PhD., Dr. Phil. Harald Bichlmeier, doc. PhDr. Irena Bogoczová, CSc., Bela Brogyanyi, M.A., doc. Irén Hegedüs, PhD., prof. PhDr. Petr Karlík, CSc., prof. dr. Alexander Lubotsky, PhD., prof. Dr. Stefan Michael Newerkla, Dr. Martine Robbeets, prof. PhDr. Jana Sokolová, CSc., doc. Georgij Starostin, CSc.

Výkonný editor: prof. Mgr. Pavel Kosek, Ph.D.

Odpovědný editor: Mgr. Marcin Waگیel, Ph.D.

Časopis je pokračováním SPFFBU, řady A, jazykovědné.

Vydává Masarykova univerzita, Žerotínovo nám. 9, 601 77 Brno, IČ 00216224

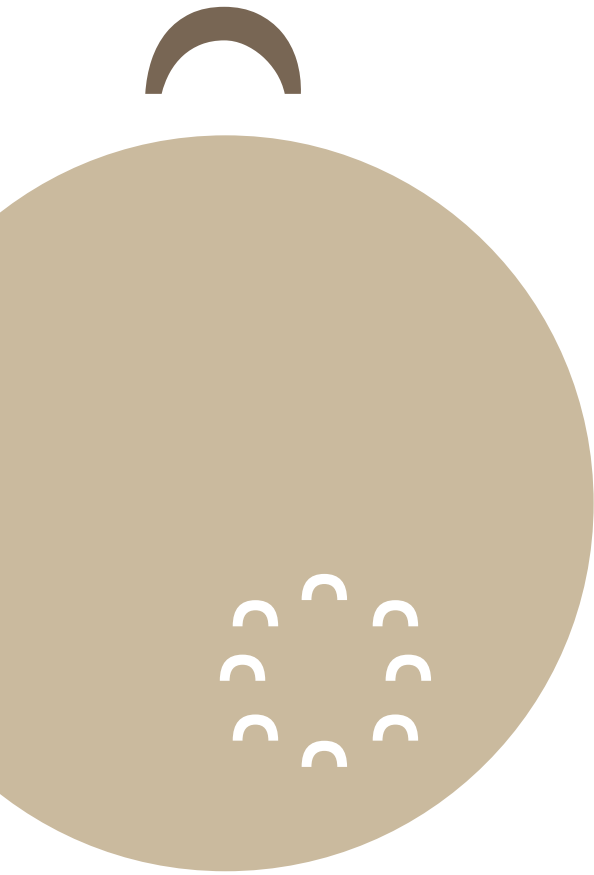
Vychází dvakrát ročně, toto číslo vychází 24. 5. 2024

Grafická úprava obálky a typografie: Pavel Křepela

Sazba: Dan Šlosar, CIT FF

Časopis je evidován MK ČR pod č. E 18712

ISSN 2336-4440 (online)



(h)

