Gutiérrez Rubio, E., Falkowska, M., Kislova, E., Stępień, M. (Hrsg.) 2015. Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV) 18. München – Berlin – Washington/D.C., 26-36.
(Die Welt der Slaven. Sammelbände/Сборники. 57.)

A CORPUS-BASED ANALYSIS OF THE CZECH SYLLABLE[1]

Abstract: This paper provides a quantitative analysis of the syllable in contemporary Czech through a corpus of 146,703 syllables contained in Czech words recorded in *Slovník spisovné češtiny*. It examines the syllable structure of words belonging to different word classes, showing that there are only minor differences between them where syllable structure is concerned, although pronouns and uninflectable words do have a less complex structure. It is also demonstrated that combinations of syllables within words show certain tendencies. In multisyllabic words, Czech prefers all the syllabic nuclei to contain short rather than non-short vowels. Additionally, words with one closed syllable are preferred to words with more than one closed syllable or with none. Finally, the distribution of heavy vs. light syllables is examined.

Keywords: corpus, syllable, phonotactics, word classes, phonological word.

1. Introduction

This paper presents a quantitative, corpus-based analysis of the syllable in Modern Standard Czech. The syllable is understood here as a self-contained phonotactic unit of the phoneme distribution (Mulder 1989, Bičan 2013). First, we survey Czech syllable structure, focusing on the frequency of syllable types. In particular, we examine the syllable structure of words belonging to different word classes. The second part of the paper considers the co-occurrence of syllables within words, with emphasis on the distribution of vocalic quantity, open vs. closed syllables and light vs. heavy syllables.

The syllable structure of Czech has already been quantitatively analyzed (Kučera, Monroe 1968 and Ludvíková 1972a, 1972b, 1976, 1978, also Ludvíková in Těšitelová et al. 1985), but these analyses were limited by the techniques then available. Kučera's database contained 42,217 syllables, and that of Ludvíková 5,000. Both databases were created from actual Czech texts, and thus included inflected forms. We have chosen another approach. Since the phonological structure of Czech vocabulary as a whole has not so far been quantitatively analyzed, our paper seeks to fill this gap. Phonological analyses of the complete lexicons of various languages have already been carried out (Rousset 2004); our findings may thus contribute to a potential typological comparison of languages with regard to syllable structure.

Our analysis is based on the vocabulary recorded in the dictionary *Slovník spisovné češtiny pro školu a veřejnost* (2003; hereinafter the SSČ), which is part of the much larger Czech Phonological Lexical Corpus.[2] The SSČ contains 49,365 entries, counted in tokens (i.e. meaningful character strings in computerized lexical analysis; recurring tokens are counted as single *types*). It includes only uninflected forms; that is, nouns and adjectives are in the nominative singular, and verbs in the infinitive. Every entry has been provided with a phonological transcription of its assumed pronunciation according to the rules of Czech orthoepy (see the VSČ). The phonological theory behind the transcription, as well as the following analysis, is that of functional phonology in the tradition of Nikolai Trubetzkoy (1939), André Martinet (2011) and Jan Mulder (1989); its details are not immediately relevant to our paper, and readers are directed to Bičan (2013) for an overview.

[1] This paper was supported by grant 13-15361P *Issues in the Phonology of the Word in Czech* (Grant Agency of the Czech Republic). I would like to thank to David Jones for correcting my English.

[2] See <http://www.ujc.cas.cz/phword>. We have chosen not to consider the whole lexical corpus here, because it includes many archaic, dialectal and occasional words which are not commonly used in contemporary Czech. The whole corpus contains 276,246 entries consisting of 950,949 syllables.

The phonological transcription takes into consideration both the segmental and the suprasegmental structure of words. Entries are transcribed as sequences of phonemes divided into syllables, phonological words and larger units. A phonological word is a unit defined by accentual properties (it is not larger than a stress group) and/or by the occurrence of boundary signals, the most important of which is a glottal stop (Bičan 2014a). The glottal stop is not a realization of any phoneme in Czech, but occurs before vowels at a grammatical boundary. The boundary may lie either between grammatical words (e.g. *k* [ʔ]*autu* 'toward the car') or between a prefix and the word base (e.g. *na*[ʔ]*učit* 'to teach (perfective)', cf. *učit* 'to teach (imperfective)') or between parts of a compound (e.g. *indo*[ʔ]*evropský* 'Indo-European'). Consequently, as in many other languages (Dixon, Aikhenvald 2003), phonological words, which are defined phonologically, and grammatical words, which are defined grammatically, are not identical units. The vocabulary of the SSČ corresponds to 45,978 phonological words (counted as types, consisting of 146,703 syllables). These syllables are the data for our analysis.

2. Czech syllable structure

Let us begin by summarizing the known facts about the syllable in Czech (for details, see Bičan 2013). Every syllable contains a nuclear element which is either a vowel or a sonant /r/ or /l/. The vowels always function as syllable nuclei, whereas the sonants acquire this function only in two situations: first, when standing between two consonants (cf. *prst* 'finger', *vlk* 'wolf'); second, word-finally after a consonant (cf. *vítr* 'wind', *mysl* 'mind'). Everywhere else (that is, word-initially, as in *rty* 'lips' or next to a vowel, as in *rám* 'frame') the sonants are not nuclear. It should be noted that the domain where the nuclearity of sonants is determined, and which is also the starting point of our analysis, is a phonological word (Bičan 2014b). Across phonological word boundaries the sonants /r/ and /l/ are not nuclear, even if standing between two consonants (cf. *utřít rty* 'to wipe one's lips').

In a syllable a nuclear phoneme may stand on its own (cf. *a* 'and') or be preceded and/or followed by one or more non-nuclear consonants. As many as five may occur syllable-initially (cf. *on* 'he', *jen* 'only', *prase* 'pig', *strach* 'fear', *vzpřímený* 'erect', *vzkvět* [fskvjɛt] 'prosperity') and as many as three may occur syllable-finally (*pes* 'dog', *kost* 'bone', *zábst* 'to freeze'). This gives 20 logically possible syllable types which are listed in table 1 (V = nuclear phoneme, C = non-nuclear phoneme; for example, *strach* 'fear' corresponds to CCCVC).

		End			
		-Ø	*-C*	*-CC*	*-CCC*
Beginning	*Ø-*	V	VC	VCC	VCCC
	C-	CV	CVC	CVCC	CVCCC
	CC-	CCV	CCVC	CCVCC	CCVCCC
	CCC-	CCCV	CCCVC	CCCVCC	CCCVCCC
	CCCC-	CCCCV	CCCCVC	CCCCVCC	*CCCCVCCC
	CCCCC-	CCCCCV	CCCCCVC	*CCCCCVCC	*CCCCCVCCC

Table 1: Logically possible syllable types in Czech (V = nuclear phoneme, C = non-nuclear phoneme)

Though logically possible, not all the types listed in table 1 are actually attested. Even though a syllable may begin with five as many as consonants, such syllables are rare, and not all of the variants are possible in Czech. As we have argued elsewhere (Bičan

2013), Czech disallows syllables containing more than six consonants, irrespective of whether they are accumulated at the beginning or at the end. The disallowed types are asterisked in table 1. To put it another way, Czech does not allow syllable types CCCCVCCC, CCCCCVCC and CCCCCVCCC. Indeed, such types have not been found in the whole lexical corpus. The remaining syllable types beginning with five consonants, CCCCCV and CCCCCVC, are attested only three times in the whole corpus, and neither of them occurs in any word recorded in the SSČ. Consequently, syllables with five initial consonants will not be considered in the rest of this paper.

Table 2 provides percentages for the types attested in the SSČ. It is obvious that Czech prefers open syllables (69.99%) to closed ones (30.01%). Similarly, syllables beginning with a single consonant (69.44%) are much more common than those beginning with no consonant or those beginning with two or more consonants (30.56%). The intersection of these two preferences is the most common syllable type, CV. The occurrence hierarchy of the five most common types is CV > CVC > CCV > CCVC > CVCC. Our findings here agree with previous quantitative analyses based on actual texts, although the exact percentage differs a bit (see Těšitelová et al. 1985). This means that preferences for given syllable types are largely the same, regardless of whether it is the Czech lexicon or Czech texts, where inflected forms abound.

All words		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	3.02%	1.41%	0.09%	0.0007%	4.52%
	C-	48.05%	18.17%	3.13%	0.08%	69.44%
	CC-	17.11%	5.51%	0.96%	0.03%	23.61%
	CCC-	1.76%	0.43%	0.16%	0.005%	2.36%
	CCCC-	0.05%	0.02%	0.001%	0%	0.07%
Total		69.99%	25.54%	4.34%	0.12%	100%

Table 2: Percentages of syllable types in the whole vocabulary (146,703 syllables in total)

Although the Czech syllable has already been analyzed, one question has not been considered yet, namely whether the syllable structure is identical in words belonging to various word classes. Intuitively, we might have expected that, for example, the structure of inflectable words would differ from the structure of uninflectable words, but no analysis has confirmed or disproved that. Since the vocabulary included in our corpus can be sorted according to word classes (of which Czech grammars recognize 10), it is easy for us to examine this aspect. Percentages of the various syllable types found in the individual word classes are given in tables 3-12.

Nouns		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	2.99%	1.49%	0.10%	0.002%	4.58%
	C-	47.83%	17.90%	5.31%	0.06%	71.11%
	CC-	14.65%	5.81%	1.72%	0.01%	22.18%
	CCC-	1.21%	0.52%	0.33%	0.002%	2.07%
	CCCC-	0.03%	0.02%	0.002%	0%	0.06%
Total		66.71%	25.75%	7.46%	0.07%	100%

Table 3: Percentages of syllable types in nouns (59,756 syllables in total)

Adjectives		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	2.54%	1.19%	0.09%	0.07%	3.89%
	C-	57.68%	10.14%	2.39%	0%	70.22%
	CC-	19.64%	2.87%	0.48%	0.01%	22.99%
	CCC-	2.50%	0.30%	0.04%	0%	2.85%
	CCCC-	0.05%	0.01%	0%	0%	0.06%
Total		82.41%	14.51%	3%	0.08%	100%

Table 4: Percentages of syllable types in adjectives (39,913 syllables in total)

Pronouns		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	1.57%	0.31%	0%	0%	1.89%
	C-	50.63%	18.55%	2.83%	0%	72.01%
	CC-	21.07%	4.72%	0.31%	0%	26.10%
	CCC-	0%	0%	0%	0%	0%
	CCCC-	0%	0%	0%	0%	0%
Total		73.27%	23.58%	3.14%	0%	100%

Table 5: Percentages of syllable types in pronouns (318 syllables in total)

Numerals		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	3.78%	0%	0%	0%	3.78%
	C-	59.70%	10.08%	4.79%	1.76%	76.32%
	CC-	13.10%	2.02%	1.26%	0.50%	16.88%
	CCC-	2.52%	0%	0%	0%	2.52%
	CCCC-	0.50%	0%	0%	0%	0.50%
Total		79.60%	12.09%	6.05%	2.27%	100%

Table 6: Percentages of syllable types in numerals (397 syllables in total)

Verbs		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	3.76%	1.58%	0.05%	0%	5.39%
	C-	36.45%	30.41%	0.42%	0.12%	67.39%
	CC-	15.77%	8.73%	0.40%	0.09%	24.98%
	CCC-	1.65%	0.44%	0.05%	0.02%	2.15%
	CCCC-	0.07%	0.02%	0%	0%	0.09%
Total		57.69%	41.17%	0.91%	0.23%	100%

Table 7: Percentages of syllable types in verbs (33,770 syllables in total)

Adverbs		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	2.71%	1.34%	0.11%	0.07%	4.22%
	C-	49.27%	12.05%	2.34%	0%	63.67%
	CC-	24.76%	4.04%	0.37%	0.03%	29.20%
	CCC-	2.38%	0.39%	0.08%	0%	2.85%
	CCCC-	0.04%	0.01%	0.01%	0%	0.06%
Total		79.17%	17.82%	2.91%	0.10%	100%

Table 8: Percentages of syllable types in adverbs (12,203 syllables in total)

Prepositions		End				Total
		-Ø	-C	-CC	-CCC	
Beginning	Ø-	4.79%	1.37%	0%	0%	6.16%
	C-	54.11%	12.33%	3.42%	0%	69.86%
	CC-	15.75%	5.48%	1.37%	0%	22.60%
	CCC-	0.68%	0.68%	0%	0%	1.37%
	CCCC-	0%	0%	0%	0%	0%
Total		75.34%	19.86%	4.79%	0%	100%

Table 9: Percentages of syllable types in prepositions (146 syllables in total)

Conjunctions		End				Total
		-Ø	-C	-CC	-CCC	
Beginning	Ø-	3.50%	2.10%	2.10%	0%	7.69%
	C-	48.95%	22.38%	2.80%	0%	74.13%
	CC-	13.99%	3.50%	0%	0%	17.48%
	CCC-	0.70%	0%	0%	0%	0.70%
	CCCC-	0%	0%	0%	0%	0%
Total		67.13%	27.97%	4.90%	0%	100%

Table 10: Percentages of syllable types in conjunctions (143 syllables in total)

Particles		End				Total
		-Ø	-C	-CC	-CCC	
Beginning	Ø-	5.32%	3.19%	0%	0%	8.51%
	C-	47.87%	28.72%	5.32%	0%	81.91%
	CC-	7.45%	1.06%	0%	0%	8.51%
	CCC-	0%	1.06%	0%	0%	1.06%
	CCCC-	0%	0%	0%	0%	0%
Total		60.64%	34.04%	5.32%	0%	100%

Table 11: Percentages of syllable types in particles (94 syllables in total)[3]

Interjections		End				Total
		-Ø	-C	-CC	-CCC	
Beginning	Ø-	3.39%	2.12%	0%	0%	5.51%
	C-	53.81%	13.98%	1.69%	0%	69.49%
	CC-	17.37%	6.36%	0.42%	0%	24.15%
	CCC-	0%	0.42%	0.42%	0%	0.85%
	CCCC-	0%	0%	0%	0%	0%
Total		74.58	22.88%	2.54%	0%	100%

Table 12: Percentages of syllable types in interjections (236 syllables in total)

Several conclusions can be drawn from the tables. To begin with, all word classes prefer open syllables to closed ones. The open syllables are most common in adjectives (82.41%) and numerals (79.60%) and least common in particles (60.64%) and verbs (57.69%). Similarly, syllables beginning with a single consonant are most frequent in all word classes, though there are again individual differences: they are most common in numerals (76.32%) and particles (81.91%) and least common in verbs (67.39%) and adverbs (63.67%).

[3] In Czech, particles express, among others, the modality of a sentence (*ať* as in *ať jste šťastní* 'may you be happy'), the attitude of the speaker (*klidně* as in *klidně mluv* 'you can speak, don't be afraid to speak') or the validity of a sentence (*ano* 'yes').

Secondly, all word classes agree in preferring the same syllable types, namely CV, CCV, CVC, CVCC, CCVC and V. The following are the individual hierarchies (the overall hierarchy for the SSČ being CV > CVC > CCV > CCVC > CVCC; see above):

Nouns: CV > CVC > CCV > CCVC > CVCC
Adjectives: CV > CCV > CVC > CVCC > V
Pronouns: CV > CCV > CVC > CCVC > CVCC
Numerals: CV > CCV > CVC > CVCC > V
Verbs: CV > CVC > CCV > CCVC > V
Adverbs: CV > CCV > CVC > CCVC > V
Prepositions: CV > CCV > CVC > CCVC > V
Conjunctions: CV > CVC > CCV > CCVC, V
Particles: CV > CVC > CCV > CVCC, V
Interjections: CV > CCV > CVC > CCVC > V

The individual word classes thus only differ from each other in the ranking of the six most preferred types. CV is always the most common type. CCV is the second most common, and CVC is the third, except for nouns, verbs, conjunctions and particles, for which CVC is the second and CCV is the third. CCVC is the fourth, except for adjectives, numerals and particles, for which the fourth is CVCC. Finally, the fifth most common type is either CVCC (nouns and pronouns) or V (adjectives, numerals, verbs, adverbs, prepositions and interjections). For conjunctions and particles, type V shares the fourth position with CCVC and CVCC, respectively. For nouns and pronouns, type V is the sixth.

The tables clearly show that not all syllable types are attested for every word class. Uninflectable words (prepositions, conjunctions, particles and interjections) have a simpler syllable structure in comparison to inflectable words. They do not allow syllables beginning with four consonants or ending in three consonants. Nor does either of them have syllables with three initial and two final consonants (i.e. type CCCVCC). However, it is pronouns that have the simplest syllable structure: in these words all syllables begin and end with one or two consonants (or with none).

It is obvious that there is some similarity here between pronouns and uninflectable words, both categories having a less complex syllable structure than any others. Even the overall length of these words is similar. This follows from the data in the left-hand part of table 13, where the average number of syllables per word for the individual word classes is provided. Pronouns and uninflectable words contain on average 2 syllables, and are one syllable shorter than the other word classes. Interjections are the shortest, while the longest are adjectives.

	Average number of syllables per word	*Percentage of words containing the given number of syllables*				
		1	*2*	*3*	*4*	*5*
Nouns	3.02	5.80%	27.88%	36.67%	20.70%	6.77%
Adjectives	3.68	0.44%	9.78%	36.46%	34.77%	13.58%
Pronouns	2.19	24.83%	39.31%	27.59%	8.28%	0%
Numerals	3.18	5.60%	28.00%	27.20%	24.80%	11.20%
Verbs	3.04	2.01%	24.54%	46.31%	22.74%	3.51%
Adverbs	3.21	2.16%	21.07%	40.75%	27.55%	6.87%
Prepositions	2.15	32.35%	38.24%	16.18%	8.82%	4.41%
Conjunctions	2.13	20.90%	46.27%	31.34%	1.49%	0%
Particles	2.09	33.33%	35.56%	20.00%	11.11%	0%
Interjections	1.74	49.27%	34.56%	11.03%	3.68%	1.47%
Total	3.19	3.44%	22.33%	39.18%	25.01%	7.57%

Table 13: Average number of syllables per word by class, with percentages of those containing 1-5 syllables

Table 13 also provides the percentages of words containing 1 to 5 syllables for each word class. The shaded cells have the highest percentages. Pronouns as well as uninflectable words prefer two syllables.[4] Interjections are exceptional, for almost a half of them are monosyllabic. The other word classes prefer three syllables, except for numerals, among which the most common are two-syllable words, though three-syllable words are almost as common. Finally, there are no pronouns, conjunctions or particles with five or more syllables.

3. Co-occurrence of syllables in Czech words

Having discussed the structure of isolated syllables, let us now turn our attention to the co-occurrence of syllables in Czech words. Although words may consist of just one syllable (e.g. *ten* 'that'), only 3.44% of them are monosyllabic in our corpus (see table 13). The longest word contains ten syllables, but such words are very rare (only three are attested in our corpus, e.g. *devadesátikilometrový* 'ninety-kilometer'). However, we will not be interested in word length any longer, focusing instead on the co-occurrence of syllables of different structures. In this paper we have limited ourselves to words containing only two, three, four or five syllables. This sample is nevertheless sufficient, because such words comprise approximately 94.09% of all the words in our corpus, and the proportion of words with more syllables than five is only 2.47%. Moreover, due to limitations on the length of this paper, we will no longer consider word classes.

Since every syllable in Czech contains either a vowel, short, long or diphthongal, or a nuclear sonant, this defines the four types of syllabic nucleus. The short vowels are by far the most frequent nuclei (72.62%). Next come the long vowels (19.1%), then the diphthongal vowels (2%), while the least common are the nuclear sonants (1.28%).

Now the first thing to consider is the distribution of the four types of nucleus. Every word may be viewed as having a certain pattern formed by its syllabic nuclei–let us call it a quantity pattern. Thus the word *polévka* 'soup' has the quantity pattern SLS, where S stands for a short vowel and L for a long vowel (likewise the letter D will stand for a diphthongal vowel and the letter R for a nuclear sonant). Table 14 provides hierarchies of the five most common quantity patterns. In all cases the most common type are words with only short vowels. For example, almost one half of words with two syllables are like this. The same preference has been found with longer words: the most frequent are words containing only short vowels. On the other hand, the more long vowels and diphthongs a quantity pattern contains, the rarer it tends to be (see Bičan forthcoming).

	Percentages for words with the given number of syllables			
	2	*3*	*4*	*5*
Pattern rankings in descending order	SS (48.39%)	SSS (40.20%)	SSSS (39.08%)	SSSSS (36.42%)
	SL (20.95%)	SSL (20.42%)	SSSL (24.02%)	SSSSL (30.62%)
	LS (16.14%)	SLS (13.22%)	SSLS (8.40%)	SSSLL (6.61%)
	SD (3.59%)	LSS (6.41%)	SLSL (5.08%)	SSSLS (5.40%)
	DS (2.86%)	LSL (4.99%)	SSLL (4.54%)	SSLSL (4.68%)
Total words	10,269	18,013	11,499	3,482

Table 14: The five most common quantity patterns for words with 2-5 syllables

[4] The fact that words in these categories contain two syllables *on average* does not of course mean that two-syllable words are necessarily the most common in them (cf. the situation at numerals, where the average syllable count is clearly not the commonest), although two-syllable words are indeed the most common in Czech as a whole.

	Number of syllables in the word			
	2	*3*	*4*	*5*
No closed syllable	22.05%	25.08%	29.52%	29.32%
1 closed syllable	62.72%	55.36%	48.28%	45.72%
2 closed syllables	15.23%	18.40%	19.67%	20.36%
3 closed syllables	–	1.17%	2.75%	4.19%
4 closed syllables	–	–	0.07%	0.37%
5 closed syllables	–	–	–	0.03%

Table 15: Ratios between the number of syllables in a word and the number of them that are closed

Another aspect to consider is the distribution of open and closed syllables. As has already been mentioned, Czech prefers open syllables (69.99%) to closed ones (30.01%), i.e. Czech words contain more than twice as many open syllables as closed ones. Hence, the occurrence of closed syllables is expected to be limited. This is confirmed by table 15, which shows percentages for given numbers of closed syllables.

In table 15 the shading highlights the most common number of closed syllables (as indicated by the highest percentage in each column). The table reveals several things. The most striking of these is the fact that the occurrence of closed syllables follows the same tendencies irrespective of the total number of syllables per word. The most common type of Czech word has one (and only one) closed syllable. Next come words containing no closed syllable. Then those containing two, three, four and five closed syllables, with the hierarchy always being the same. The next thing to note is that the percentages of words with all syllables closed are in inverse proportion to the numbers of syllables. Finally, words with three or more closed syllables are quite infrequent. Less than half a percent of all Czech words with fewer than six syllables have more than three of them closed.

Another possible way to look at words from the syllabic perspective is to make a distinction between light and heavy syllables. Traditionally, a light syllable is an open syllable containing a short vowel or, we might add, a nuclear sonant, whereas a heavy syllable is one containing a long or diphthongal vowel, or one that is closed (Hyman 1985). This distinction is not only made in phonological descriptions of various languages, but also in versification theory (cf. Ibrahim et al. 2013).

Table 16 shows percentages for given numbers of heavy syllables. The most common types are again highlighted by shading. Although there is no uniform relationship to the number of syllables, it is still apparent that Czech prefers words with 1 and 2 heavy syllables. In fact it emerges from data not shown here that 83.25% of all words are of this kind. In contrast, words with no heavy syllable (8.37%) and words with three and more heavy syllables (8.38%) are not very common. The low frequency of words with three or more heavy syllables was expected, because a heavy syllable is the one containing a long or diphthongal vowel and/or one that is closed and it has already been demonstrated that the occurrence of long and diphthongal vowels is limited in Czech, with the frequency of words decreasing in inverse proportion to the number of closed syllables (see above).

	Number of syllables in the word			
	2	*3*	*4*	*5*
No heavy syllable	10.00%	8.14%	7.06%	8.59%
1 heavy syllable	56.74%	43.49%	39.55%	33.69%
2 heavy syllables	33.26%	40.23%	29.90%	37.71%
3 heavy syllables	–	8.14%	12.04%	16.40%
4 heavy syllables	–	–	1.45%	3.25%
5 heavy syllables	–	–	–	0.37%

Table 16: Ratios between the number of syllables in a word and the number of them that are heavy

4. Conclusion

Having considered various aspects of the syllable structure in Czech from a quantitative perspective, we have found clear preferences for certain syllable types. It is well known that Czech allows syllables beginning and ending with consonantal clusters. Many and various clusters are attested; in our corpus there are 337 syllable-initial clusters and 73 syllable-final clusters.[5] However, our data has shown that only 26.04% of the syllables contain an initial cluster. 4.44% begin with a vowel, and the remaining 68.44% begin with a single consonant. The last syllable type is clearly favored. Similarly, only 4.44% of syllables end in a consonant cluster. 25.54% end in a single consonant, and 69.99% end in a vowel. Again, one syllable type is clearly preferred. The most common syllable type is CV. Similar preferences have been found in analyses based on actual Czech texts (Těšitelová 1985). This allows us to conclude that the described distributions of syllable types reveal a consistent property of Czech, i.e. it holds for the lexicon as well as for actual texts which, unlike the lexicon, contain inflected forms of words.

The same results have been found across the various word classes, although the actual percentages differ. The greatest proportion of initial clusters is in adverbs (32.11%), and the smallest in particles (9.57%). The greatest proportion of final clusters has been found in numerals (8.32%), and the smallest in verbs (1.14%). It is likewise clear that the preferred syllable type in Czech is CV, regardless of word class. This should not be surprising, because the same preference has been observed in other languages too (Rousset 2004). Claims have been made that CV is, perhaps not surprisingly, a universally occurring syllable structure (e.g. Malmberg 1967), but to what extent it is a preferred syllable structure, on the universal scale, has yet to be ascertained.

The tendency for Czech to prefer certain syllable types can also be seen in another set of data. As mentioned, the goal of this paper is to describe the syllabic structure of Modern Czech. Thus, only the vocabulary from the SSČ has been included in the analysis, as it is assumed that this vocabulary covers the most commonly used Czech words. However, it may be worthwhile to consider less commonly used words, and archaisms, too. Accordingly, we present in table 17 the data for nouns recorded only in the PSJČ, and not in other dictionaries of Czech, namely the SSČ and the SSJČ. The PSJČ is the oldest dictionary included in the Czech Phonological Lexical Corpus, recording many archaic words, poetic variants or dialectisms. We have limited ourselves to nouns, because they constitute the largest group of words. If one compares table 17 with table 3 (nouns recorded in the SSČ), it is obvious that, first, the very same syllable types are preferred, and second, the percentages of the various types are very similar.

[5] Many more clusters are possible; see Bičan (2013).

Nouns		*End*				*Total*
		-Ø	*-C*	*-CC*	*-CCC*	
Beginning	*Ø-*	2.58%	1.3%	0.07%	0%	3.94%
	C-	49.31%	17.26%	5.69%	0.06%	72.33%
	CC-	15.04%	5.53%	1.09%	0.01%	21.67%
	CCC-	1.34%	0.45%	0.19%	0.001%	1.98%
	CCCC-	0.05%	0.03%	0%	0%	0.08%
Total		68.32%	24.57%	7.04%	0.07%	100%

Table 17: Percentages of syllable types in nouns recorded only in the PSJČ (109,125 syllables in total)

However, our analysis has also shown that certain word classes do not make use of all the syllable types attested. Pronouns do not have syllables beginning with three or four consonants, or syllables ending in three consonants. In fact, pronouns exhibit the simplest syllable structure. Likewise, uninflectable words (prepositions, conjunctions, particles and interjections) have a simpler structure. They too do not make use of syllables beginning with four consonants or ending in three consonants. Considering all these facts, we can conclude that so-called function (as opposed to content) words tend to have a simpler syllable structure (Czech pronouns as well as uninflectable words fall into this group because they usually perform a syntactical function rather than convey any lexical content). It is hardly a coincidence that function words are phonologically simple in Czech, and a similar situation may exist in other languages. In any case, the results confirm what is intuitively known to native speakers of Czech.

The fact that open syllables are favored in Czech is also reflected in the distribution of syllables within phonological words. One might expect that words with all open syllables would be preferred, but this has not been confirmed. Instead, Czech favors words with one closed syllable (the others being open), and this preference is apparent regadless of the number of syllables. 54.44% of words are of this kind, whereas only 24.80% contain only open syllables. Next come words with two closed syllables (18.14%), and the overall proportion of words with more closed syllables than that is very low (1.53%).[6] Thus, even if Czech favors open syllables, three fourths of Czech words contain at least one closed syllable.

Finally, Czech is fond of syllables whose nucleus is a short vowel. 77.62% of the syllables are of this kind. This preference for short-vowel syllables is obvious, because 42.29% of all words contain only short vowels. Of course, the proportion of words with at least one non-short vowel is larger (i.e. 57.71%), but if we consider that the short vowels are just one of four possible nuclei in Czech (the others being the long and diphthongal vowels and the nuclear sonants), the proportion is enormous.

References

Bičan, A. 2013. *Phonotactics of Czech*. Frankfurt am Main.

Bičan, A. 2014a. K pojmu fonologické slovo v češtině. In Boček, V., Vykypěl, B. (eds.). *Sophia Slavica*. Brno, 13-23.

Bičan, A. 2014b. On the Czech Nuclear /r/ and /l/. In Witkoś, J., Jaworski, S. (eds.). *New Insights into Slavic Linguistics*. Frankfurt am Main, 21-34.

Bičan, A. forthcoming. Distribution of Vocalic Quantity in Czech. *Grazer Linguistische Studien*. Draft available here: <http://www.ujc.cas.cz/phword>.

Czech Phonological Lexical Corpus: <http://www.ujc.cas.cz/phword>.

Dixon, R. M. W., Aikhenvald, A. Y. 2003. Word: A Typological Framework. In: Dixon, R. M. W., Aikhenvald, A. Y. (eds.). *Word: A Cross-linguistic Typology*. Cambridge, 1-41.

[6] As explained above, we have considered only words with 2, 3, 4 and 5 syllables when counting these percentages, not only for closed syllables but also for heavy syllables, but these polysyllabic words constitute 94.09% of all the words included in our corpus.

Hyman, L. 1985. *A Theory of Phonological Weight*. Dordrecht.
Ibrahim, R., Plechač, P., Říha, J. 2013. *Úvod do teorie verše*. Praha.
Kučera, H., Monroe, G. K. 1968. *A Comparative Quantitative Phonology of Russian, Czech, and German*. New York.
Ludvíková, M. 1972a. Quantitative Syllable Analysis of Words in Czech. *Prague Studies in Mathematical Linguistics* 3, 27-34.
Ludvíková, M. 1972b. Some Quantitative Aspects of the Czech Syllables. *Prague Studies in Mathematical Linguistics* 4, 141-54.
Ludvíková, M. 1976. On Some Statistical Differences in Two Spoken Texts on the Syllabic Level. *Prague Studies in Mathematical Linguistics* 5, 91-104.
Ludvíková, M. 1978. On the Occurrence of Syllables in Different Word Positions. *Prague Studies in Mathematical Linguistics* 6, 39-45.
Malmberg, B. 1967. *Structural Linguistics and Human Communication*. Berlin, Heidelberg, New York.
Martinet, A. 2011. *Éléments de linguistique générale*. Paris.
Mulder, J. 1989. *Foundations of Axiomatic Linguistics*. Berlin.
PSJČ = *Příruční slovník jazyka českého I–VIII*. Praha, 1935-1957.
Rousset, I. 2004. *Structures syllabiques et lexicales des langues du monde*. Ph.D. thesis. Université Stendhal. <https://tel.archives-ouvertes.fr/tel-00250154>
SSČ = *Slovník spisovné češtiny pro školu a veřejnost*. Praha, 2003.
SSJČ = *Slovník spisovného jazyka českého I–IV*. Praha, 1960-1971.
Těšitelová, M. et al. 1985. *Kvantitativní charakteristiky současné češtiny*. Praha.
Trubetzkoy, N. 1939. *Grundzüge der Phonologie*. = *Travaux du Cercle Linguistique de Prague* 7.
VSČ = Hála, B. et al. 1968. *Výslovnost spisovné češtiny*. Praha.

Ústav pro jazyk český, Akademie věd České republiky
Veveří 97
Brno – Czech Republic

Aleš Bičan
(bican@phil.muni.cz)

Gutiérrez Rubio, E., Falkowska, M., Kislova, E., Stępień, M. (Hrsg.) 2015. Beiträge der Europäischen Slavistischen Linguistik (POLYSLAV) 18. München – Berlin – Washington/D.C., 37-44.
(Die Welt der Slaven. Sammelbände/Сборники. 57.)

DIALEKTY SŁOWEŃSKIE W SŁOWNIKU LINDEGO – PRÓBA CHARAKTERYSTYKI

Slovene dialects in Linde's dictionary of Polish – an attempt at description

Abstract: The paper presents some basic information on the Slovene language data in Linde's dictionary of Polish. The language spoken on the territory of what is now Slovenia, which was not codified at that time, was represented in the dictionary as two dialects: Vindish and Carniolic. For research purposes a digital version of the dictionary is used which can be searched through with a search engine that allows dictionaries to be browsed as if they were corpora.

Keywords: Linde's dictionary, Slovene, lexicography.

1. Wprowadzenie

Niniejszy artykuł jest wstępną analizą słoweńskojęzycznych danych zamieszczonych w *Słowniku języka polskiego* Samuela Bogumiła Lindego, będącym pierwszym ogólnym słownikiem języka polskiego. Pierwsze wydanie słownika pochodzi z lat 1807-1814 (Linde 1807-1814), natomiast drugie ukazało się w latach 1854-1861 r. (Linde 1854-1861) i to na nim opieram swoje badania.

Autor dążył do udokumentowania całości słownictwa polskiego i jednocześnie przedstawienia materiału leksykalnego. Ze względu na etymologiczne i panslawistyczne zainteresowania autora słownik stał się jednocześnie wielojęzyczny (choć nie przekładowy) – por. ilustracja 1. Zawiera bowiem przykłady z wielu języków i dialektów, które autor uznał za niezbędne do pełnego przedstawienia znaczeń opisywanych przez siebie leksemów (Matuszczyk 2007:13-14). Wśród tych licznych fragmentów obcojęzycznych pojawiają się nader często przykłady z dialektów słoweńskich.

MAGIEL, glu, *m.*, *et* gli, *ż.*, *Boh.* mandl; *Vind.* likaunik, likalu, rolovíhè, povalavíhe, povalilu; *Ross.* катокъ; §. a) narzędzie do gładzenia płótna lub chust pranych, wałkownica, bie Mangel, Manbel, Rolle, die Wäſche zu glätten. Jaworowe drzewo do maglów płóciennych bardzo zdatne. *Ład. H. N.* 54. Magle do chust, warsztat bywa dębowy, wałki dębowe, brzozowe, gdzie zaś płótna bielą, jaworowe. *Kluk. Rośl.* 2, 161 *et* 21. Magiel w fabrykach płóciennych składa się z siedmiu ruchomych wałków, pionowo nad sobą ułożonych, w dwóch belkach osadzonych, nie więcej od siebie odległych, jak tylko, aby płótno między niemi przesunąć się mogło. *Przedz.* 96, cf. kalendra. — §. b) *transl.* Daremne były jej bryże, jej magle. *Pot. Syl.* 78, t. j. maglowanie, gładzenie twarzy.

Ilustracja 1: Przykładowy artykuł hasłowy

Nie poprzestawał przy tym na samym odnotowaniu leksyki – wszystkie informacje ilustrował materiałem źródłowym, podając jednocześnie odesłania bibliograficzne. Powstał w ten sposób obszerny, sześciotomowy słownik o ciekawym alfabetyczno-gniazdowym układzie artykułów hasłowych. Dzieło to, choć z różnych względów krytykowane przez niektórych naukowców tego okresu i późniejszych, stało się

DIE WELT DER SLAVEN
SAMMELBÄNDE · СБОРНИКИ

Herausgegeben von
Peter Rehder (München) · Barbara Sonnenhauser (Zürich)
Igor Smirnov (Konstanz) · Schamma Schahadat (Tübingen)

Band 57

2015

Harrassowitz Verlag · Wiesbaden

Beiträge zum 18. Arbeitstreffen der Europäischen Slavistischen Linguistik (Polyslav)

3.-5. September 2014, Budapest

Herausgegeben von
Enrique Gutiérrez Rubio, Marta Falkowska,
Ekaterina Kislova und Marzena Stępień

2015

Harrassowitz Verlag · Wiesbaden