Grazer Linguistische Studien 83 (Frühjahr 2015); S. 133-138

Distribution of vocalic quantity in Czech[1]

Aleš Bičan
Institute of the Czech Language, Academy of Sciences of the Czech Republic

Abstract. In the phonological literature three claims are made about vocalic quantity in Czech: (1) it is not subject to any positional restriction, (2) long vowels are not possible before certain consonant clusters, and (3) within a word any syllable may contain a long vowel. By confronting them with the data from the Phonological Lexical Corpus of Czech (containing 257,962 phonological words), it is shown that the first hypothesis must be rejected in its strong form, that there is no support for the second one, and that the third one must be rejected altogether.

Keywords. phonotactics, syllable, consonant cluster, phonological corpus

1. Introduction

The goal of this paper is to test hypotheses assumed by various linguists about the distribution of short and long vowels in Czech (e.g. Trnka 1982: 187–94, Horálek 1986: 128–9, Ibrahim et al. 2013: 14):

- (H1) Vocalic quantity is not subject to any positional restriction.
- (H2) Long vowels are not possible before certain consonant clusters.
- (H3) Irrespective of the number of syllables in a word, any syllable may contain a long vowel.

It will be shown that there is either no support for the claims or that there is evidence for the opposite. The data will be based on the Czech Phonological Lexical Corpus (http://www.ujc.cas.cz/phword) containing as much as 257,962 phonological words as types. The corpus is a phonologically transcribed database of the Czech vocabulary as recorded in the major dictionaries, and inflected forms are thus not included.

For the rest of this paper, two categories will be distinguished: short vowels (henceforth S; 80.3% of all vowels in the whole corpus), and long vowels (henceforth L; 19.7% of all vowels); the latter includes three diphthongal vowels. The syllabic sonants /r/ and /l/ will not be considered except for section 4; they constitute only 1.3% of all syllabic nuclei. For a detailed overview of the Czech phonology, see Bičan (2013). For a discussion of the role of vocalic quantity (henceforth VQ) in derivational and other morphological processes, see Bethin (2003), Scheer (2004), and Sukač (2013).

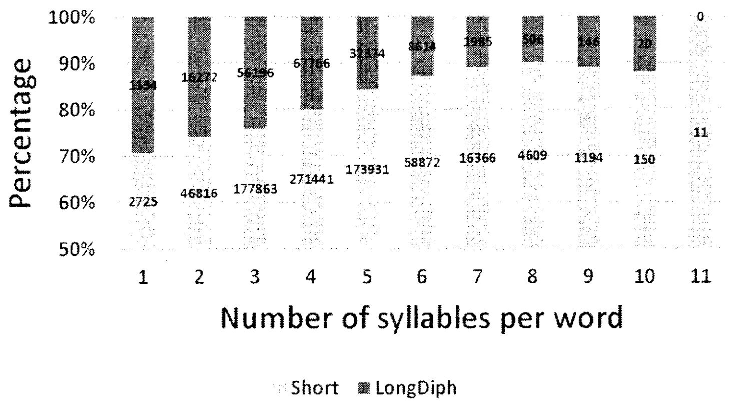
[1] The paper was supported by the grant 13-15361P *Issues in the Phonology of the Word in Czech* (Grant Agency of the Czech Republic).

2. Distribution of VQ in Word Positions and Environments

To test whether VQ is subject to some positional restriction in Czech, we have examined the distribution of S and L in various word positions and environments. The percentual proportion between them is not greatly different in word-initial and word-medial syllables (initially, S: 87.99% × L: 12.01%; medially, S: 85.51% × L: 14.49%), but it differs considerably in word-final syllables (S: 63.9% × L: 36.1%). Thus, it makes sense to distinguish between word-initial and word-medial position on the one hand, and word-final position on the other. In the former the occurrence of S and L before single consonants (S: 86.72% × L: 13.28%) is comparable to their occurrence before consonant clusters (S: 85.76% × L: 14.24%). However, in word-final position there is a considerable difference. L are obviously disfavored before clusters (S: 95.79 × L: 4.21%), whereas before a single consonant the proportion between S and L is 74.31% × 25.69%. Lastly, L are highly preferred at the absolute end (S: 57.37% × L: 42.63%). An explanation for this fact is easy to find: L are forms of many suffixes; for instance, most adjectives end in L.

Examining the distribution of VQ in words of different numbers of syllables, we have found out that the longer a word is, the more common are L in comparison to S. This is illustrated in fig. 1. The rising tendency is apparent, although the amount of L slightly decreases in words with 9 and 10 syllables. This may be due to the limited number of such words (they comprise only 0.16% of all words). Let us also mention that in medial position the frequency of L decreases with the increase of syllables per word, whereas in final position their frequency increases with the increase of syllables per word.[2]

Fig. 1: Distribution of VQ according to the number of syllables per word (in token frequencies and percent)



[2] Due to the limited space we cannot provide actual percentages to illustrate this and other facts later on, but they are to be found at this web site: http://www.ujc.cas.cz/phword.

3. Distribution of VQ before Consonant Clusters

The hypothesis H2 is based on Trnka's (1982) paper and especially on his statement (p. 188): "[in Czech] long vowels do not occur before homomorphemic consonant clusters other than *st, sť, sk, zd, zď, tr, tř, dr, tř, rt, rť, mň*". Leaving aside the question how to determine what a homomorphemic cluster is and whether this information is relevant in phonology at all, we see one obvious problem with his claim: Unless we generalize the non-occurrence of L with some clearly defined rule or constraint, it cannot be ascertained that their absence before these clusters is not just a matter of accidence. Trnka does not propose any such generalization, and indeed we see no way how the exceptional clusters could be united in a well-defined set.

Still, our corpus confirms that the distribution of L is severely limited before consonant clusters (henceforth just clusters). There are 1,646 different word-medial clusters in total in the Czech lexicon. 2.25% of them are not preceded by S, whereas as much as 72.17% of them are not preceded by L. The ratio between S and L is 85.76% × 14.24%, but this is comparable to the ratio before a single consonant (i.e. 86.72% × 13.28%, see above). Similarly, the occurrence of L is greatly limited before final clusters. There are 138 different word-final clusters in total. S are not found before 3.62% of them, whereas L are not found before 77.54%. The ratio between S and L is 95.79% × 4.21%, and as we have seen, this is quite different from the ratio before final single consonants (74.31% × 25.69%, see above). All of these facts suggest that there might be certain restrictions governing the distribution of VQ before clusters.

However, examining closely the occurrence of S and L, we have not found any obvious restriction on L. The occurrence does not seem to be influenced by the number of consonants in a cluster. One may expect that the more consonants there are in a cluster, the less common L will be. It is not the case. Table 1 exemplifies this for medial clusters, but the same mixed results have been found for final clusters. The last row states how many instances of the given clusters are found. As we see, L are most common before four-consonant clusters followed by two-consonant medial clusters. As to the six-consonant clusters, only two are attested, so the dominance of S cannot be decisive here. In the case of final clusters, L are more common before three-consonant ones than before two-consonant ones.

Table 1: Distribution of VQ before medial clusters according to the number of consonants (C2 = two consonants etc.)

	C2	C3	C4	C5	C6	Total
S	85.53%	87.04%	82.29%	91.88%	100%	85.76%
L	14.47%	12.96%	17.71%	8.12%	0%	14.24%
Tokens	197,405	54,371	7,295	382	2	259,455

Similarly, the distribution of VQ does not seem to be influenced by the quality of the consonants in a cluster. An obvious thing to look at is the difference between clusters consisting of obstruents and those containing at least one sonant (henceforth O = obstruent, R = sonant). Both S and L are found before any type of clusters (with some exceptions to be mentioned presently), so we cannot claim that certain types disallow either class of vowels. Moreover, there does not seem to be any obvious preference for any of them.

In the case of two-consonant medial clusters, the hierarchy of the preference for L is RR > OO > OR > RO, which means that L are most common before clusters of two sonants. In final position, however, these are the clusters before which L are least common, the hierarchy being OR > OO > RO > RR.

L are also found before most three-consonant clusters. In medial position they are not found before type RRO. However, such clusters are attested only 58 times and may be regarded as rare given the fact that there are 54,371 total occurrences of three-consonant medial clusters. The hierarchy of preference for L is ROO > OOO > RRR > OOR > ORR > ORO > ROR, i.e. L are most common before type ROO. Three-consonant final clusters are not very numerous; there are only 322 instances in our corpus. Some types are not attested at all (ORO, ORR, RRR), whereas others are preceded by S only. One such type is ROO before which L are found just once (out of 178 occurrences). As we have just seen, this type is the most preferred one for L in medial position. Other types before which L are not found are ROR and RRO, but they are attested only 9 times and 11 times, respectively. L are clearly preferred before type OOO where they are even more common than S.

Czech has no final clusters of four and more consonants. In the case of medial clusters, there are 16 logically possible four-consonant types, but three are not attested at all (ORRO, RRRO, RRRR). 8 types are attested less than 10 times—always for S (OORO, OORR, OROO, OROR, RORO, RORR). Type ORRR is attested 11 times, type RROO 54 times, and type RROR 17 times––always for S. These types may be regarded as rare given the fact that there are 7,295 total occurrences of four-consonant medial clusters. The four remaining types are: OOOR, ROOR with more than 700 occurrences and OOOO, ROOO with more than 2,600 occurrences. L are found before all of them, the hierarchy being ROOO > OOOO > OOOR > ROOR.

Czech has also medial clusters of five consonants, but they are very rare. There are 56 such clusters with 382 occurrences, but only four are attested more than 10times. Although L do not precede 49 of them, it is hard to make any generalization due to their overall limited occurrence. Lastly, there are only two six-consonant clusters, both attested only after S.

As we see, the occurrence of L is quite variable before medial and final clusters, and there is no obvious correlation between VQ and the quality of consonants within these clusters. Although we have focused only on whether a consonant is an obstruent or a sonant, we have also made a finer classification sorting consonants according to their manner of articulation (stops, fricatives, nasals, liquids) as well as according to their place of articulation (labials, alveolars, palatals, velars). Even in this case no correlation between VQ and the quality of the consonants could be discerned. Finally, let us add that L are more common in open syllables (the ratio being S: 79.06% × L: 20.94%) than in closed syllables (S: 84.11% × 15.98%). The difference is most likely a consequence of the low occurrence of L before final clusters.

4. Co-occurrence of S and L within Words

In this section we will address the hypothesis H3 which states that Czech words may contain any number of L, that is, in extreme they may contain as many L as they have syllables. This statement is found not only in phonological descriptions of Czech, but also in treatises on Czech versification. In confrontation with the data from the Czech Phonological Lexical Corpus, this statement cannot be held, though.

Table 2 shows the percentual occurrence of L within Czech words according to the number of syllables per word. The upper row corresponds to the number of syllables (i.e. S2 = words of two syllables), whereas the first column corresponds to the number of L per word. Here L0 means that a word does not contain any L, i.e. it contains either S or at least one syllabic sonant /r/ or /l/.

Table 2: Occurrence of L in a word according to the number of syllables (values in rows L0–L5+ are in percent, those in the last row in token instances)

	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11
L0	*53.71*	41.94	38.64	*41.59*	*42.62*	*44.98*	*42.97*	34.23	*29.41*	*100*
L1	42.31	*45.73*	*45.06*	41.07	40.82	37.19	37.97	*38.26*	*29.41*	0
L2	3.98	11.86	15.07	15.40	14.44	15.29	16.25	22.82	35.29	0
L3	–	0.47	1.20	1.87	1.98	2.47	2.66	4.70	5.88	0
L4	–	–	0.02	0.06	0.13	0.08	0.16	0	0	0
L5+	–	–	–	0	0	0	0	0	0	0
Tok.	32,373	79,300	85,908	41,641	11,308	2,630	640	149	17	1

Several conclusions can be drawn from the table. To begin with, Czech clearly prefers words with no L and words with just one L; in total, 84.33% of words fit this pattern. Actually, examining quantity patterns of Czech words where every syllable contains either a short vowel, a long vowel, a diphthongal vowel or a syllabic sonant (for example, SLS is a pattern of zapálit 'to ignite'), it has become obvious that words with just short vowels are preferred irrespective of the number of syllables. What also follows from the table is the fact that the number of attested words decreases with the increase of L within these words.[3] Finally, and most importantly, there are no words containing 5 or more L irrespective of the number of syllables.

This fact invalidates the hypothesis H3, for it is obvious that Czech words cannot contain any number of L: their amount is limited to four and even words with four L are extremely rare. Our corpus has only 59 instances of such words (0.02%). Words with three L are not very common, either; they comprise merely 0.98% (2,500 instances). In contrast, words with two L are relatively common (12.69%, 32,233 instances). Finally, the overall percentage of words with one L is 43.97% (111,677 instances), and this is almost the same as the percentage of words with no L at all (43.33%, 107,498 instances).

[3] Column S10 is exceptional, but this must be a consequence of the very small number of such words. There are only 17 words containing 10 syllables in our corpus.

5. Conclusion

Three claims H1, H2 and H3 made about VQ in Czech have been tested against the lexical corpus of 257,962 words:

(H1) VQ is not subject to any positional restriction. The hypothesis must be rejected in its strong form: No absolute positional restriction has been found, but VQ is influenced by word position (L are more common in final position and at the absolute end) and by word length (the longer a word is, the less common L are).

(H2) L are not possible before certain consonant clusters. Although L are non-occurrent before a lot of clusters, no obvious evidence has been found as a support for this claim. The occurrence of L does not seem to be influenced by the number of the following consonants or by the quality of the consonants in terms of their place and manner of articulation. Hence, the absence of L before clusters must be contributed to the overall limited occurrence of L rather than to the clusters themselves.

(H3) Irrespective of the number of syllables in a word, any syllable can contain L. This hypothesis must be rejected: Words with 4 L are very rare, and words with 5 or more L are non-occurrent at all. Although the Phonological Lexical Corpus includes only uninflected words, it does not seem likely that derivation and/or inflection would produce words with 5 L in them. There are only 59 words with 4 L in our corpus all of which end in L. These words are either neuter nouns ending in *-í* (e.g. *zaříkávání* 'incantation') or adjectives (e.g. *králíkářský* 'rabbit breeder (adj.)'). When words of these classes are inflected, the rules of the Czech inflection does not offer any possibility to append to them another syllable with L. Still, words with 5 L may arise through compounding, but this does not seem to be likely, either.

It is obvious that although two equivalent paradigmatic classes of vowels, S and L, are postulated for Czech, they are hardly equivalent syntagmatically due to the differences in their distribution.

References

Bethin, Christina (2003). Metrical Quantity in Czech: Evidence from Hypocoristics. In W. Browne et al. (eds.), *Formal Approaches to Slavic Linguistics 11: The Amherst Meeting*. Ann Arbor: Michigan Slavic Publications, pp. 63–82.

Bičan, Aleš (2013). *Phonotactics of Czech*. Frankfurt am Main: Peter Lang.

Horálek, Karel (1986). Fonologie spisovné češtiny. In Jan Petr (ed.), *Mluvnice češtiny 1*. Praha: Academia, pp. 122–156.

Ibrahim, Robert et al. (2013). *Úvod do teorie verše*. Praha: Akropolis.

Scheer, Tobias (2004). O samohláskové délce při derivaci v češtině. In Zdeňka Hladká & Petr Karlík (eds.), *Čeština – univerzália a specifika 5*. Praha: Nakladatelství Lidové noviny, pp. 224–239.

Sukač, Roman (2013). Fish and its Fisherman: Paradigmatic and Derivative Length in Czech. *Zeitschrift für Slawistik*, 58, 72–101.

Trnka, Bohumil (1982). The Distribution of Vowel Length and its Frequency in Czech. In Bohumil Trnka, *Selected Papers in Structural Linguistics*. Berlin et al.: Mouton, pp. 187–194.

Grazer Linguistische Studien 83 (Frühjahr 2015); S. 139-148

Serbian verbs front to be expressive[*]

Sabina Halupka-Rešetar
University of Novi Sad

Neda Todorović
University of Connecticut

Abstract. The squib shows that the Serbian Aorist, an aspectual tense, is peculiar in that whenever it occurs in sentence initial position it necessarily triggers an expressive meaning. We argue that this expressive interpretation is associated with a discourse-oriented projection to which Aorist raises. Aorist, we propose, may come with a focus feature, in which case this feature needs to be licensed by an operator in the Foc head. Raising the Aorist to the Foc head results in expressive interpretation. In wh-exclamatives, however, it is the wh-phrase which introduces the focus feature and brings in the expressive component, rendering the position of Aorist irrelevant.

Keywords. Serbian, Aorist, expressive component, initial position, Focus

1. Introduction

In addition to periphrastic past tense forms (1), Serbian makes use of an aspectual tense, Aorist, which describes punctual, completed actions (2a). Although it is somewhat archaic, the Aorist is still used in vivid narration. But what is peculiar about the Aorist is that when the verb occurs in sentence-initial position, it carries an expressive component, in addition to the descriptive one. In (2b), the speaker also expresses his/her attitude (great surprise in this case). This fronting of the Aorist comes with a particular intonation, which is characteristic of exclamatives. Crucially, however, this does not hold when the Aorist is not fronted, as in (2a), i.e. the sentence is not an exclamative and there is no expressive component.

(1) Jovan je udario Mariju.
Jovan is hit Marija
'Jovan has hit Marija.'

(2) a. Jovan udari Mariju.
Jovan hit-aor. Marija
'Jovan has hit Marija.'

[*] The first author acknowledges funding from project nr. 178002 of the Ministry of Education, Science and Technology of the Republic of Serbia, entitled Languages and Cultures in Time and Space. The authors wish to thank the audiences at SinFonIJA7, as well as Ljiljana Progovac, Željko Bošković, Jonathan Bobaljik, and Stefan Kaufmann for their valuable comments and suggestions to improve the quality of the paper.