sciendo
DOI 10.2478/jazcas-2023-0055

SYLLABIC CONSONANTS IN HISTORICAL CZECH AND HOW TO IDENTIFY THEM

MARKÉTA ZIKOVÁ – MARTIN BŘEZINA
– RADEK ČECH – PAVEL KOSEK
Department of Czech Language, Faculty of Arts, Masaryk University,
Brno, Czech Republic

ZIKOVÁ, Markéta – BŘEZINA, Martin – ČECH, Radek – KOSEK, Pavel: Syllabic Consonants in Historical Czech and How to Identify Them. Journal of Linguistics, 2023, Vol. 74, No 1, pp. 391 – 400.

Abstract: The paper provides fine-grained evidence concerning the development of syllabic consonants /r l/ in Czech, that is only sketched in the existing literature. The evidence is based on an automatic parser that identifies potential syllable-projecting segments according to sonority. The parser was applied to six verse texts from the 14th–16th centuries, which show a strong tendency towards octosyllabicity. The data provided by the parser newly reveal that the shift from non-syllabic to syllabic /r l/ is position-dependent: word-medial non-syllabic strings C(r/l)C change more rapidly than non-syllabic word-final ones C(r/l)#. This finding is in line with a cross-linguistic observation that non-syllabic C(r/l)C are marked, hence they are regularly syllabified prior to less marked C(r/l)#.

Keywords: syllabic consonants, historical Czech, syllable markedness, Sonority Sequencing Principle

1 INTRODUCTION

The paper investigates syllabification algorithm throughout the history of Czech. We focus on sonorants /r l/ (henceforth R) whose syllable status varied considerably during the 14th–16th centuries, as reported in the literature (Gebauer 1963; Komárek 1982; Lamprecht et al. 1986). We examine three contexts shown in Tab. 1 in which R is C(onsonant)-adjacent, but not vowel-adjacent. According to the literature, there is an asymmetry between word-initial #RC on the one side and word-medial CRC and word-final CR# on the other: the former are syllabified uniformly in historical Czech, the latter variously. Word-initial #RC never project syllables; e.g. *lhal* ‘he lied’ and *rval* ‘he tore’ are both monosyllabic, and the number of syllables (σ) corresponds to the number of vowels. By contrast, word-final CR# underwent a change towards syllable-projecting structures. For example, final R of *nesl* ‘he carried’ and *Petr* ‘Peter’ received the same syllabic status as a preceding vowel, i.e., they became syllabic consonants; originally monosyllabic words *nesl* and *Petr* thus turned into bisyllabic ones. Finally, word-medial CRC vary between

syllabic and non-syllabic, depending on particular lexical items; cf. monosyllabic *slza* 'tear' or *krvi* 'blood.gen.pl' on the one hand (where only the vowels project syllables) and bisyllabic *vlna* 'wool' or *brzo* 'soon' on the other (where both the vowels and R are syllable-projecting). This lexical contrast has however been eliminated: all word-medial R eventually became syllabic.

#RC	lh**a**$_{\sigma}$l, rv**a**$_{\sigma}$l	=	lh**a**$_{\sigma}$l, rv**a**$_{\sigma}$l	'he lied, tore'
CR#	n**e**$_{\sigma}$sl, P**e**$_{\sigma}$tr	->	n**e**$_{\sigma}$s**l**$_{\sigma}$, P**e**$_{\sigma}$t**r**$_{\sigma}$	'he carried, Peter'
CRC	slz**a**$_{\sigma}$, krv**i**$_{\sigma}$	->	s**l**$_{\sigma}$z**a**$_{\sigma}$, k**r**$_{\sigma}$v**i**$_{\sigma}$	'tear, blood.gen.pl'
CRC	v**l**$_{\sigma}$n**a**$_{\sigma}$, b**r**$_{\sigma}$z**o**$_{\sigma}$	=	v**l**$_{\sigma}$n**a**$_{\sigma}$, b**r**$_{\sigma}$z**o**$_{\sigma}$	'wool, soon'

Tab. 1. Syllable structure of /r l/ in historical Czech (14th–16th century)

The aim of this paper is to verify the above-mentioned claims, made by the historical grammars. Syllabic consonants are quite easy to be detected in the contemporary language: we can simply ask native speakers how they syllabify words like *lhal, nesl* or *slza*. This method cannot, of course, be applied in investigating historical Czech because we rely on written records. Moreover, the investigation is complicated by the fact that the R-syllabicity is not marked consistently by any graphic means in the historical texts. Thus, we work with the idea that the syllabic structure of R can be seen well in syllable-based poetry.

The idea of examining poetry to learn about syllable structure is not new, it has been put forward already in the literature mentioned above, and more recently for example, in Scheer and Ziková (2017). However, to our knowledge, there is no empirical study that thoroughly examines behavior of non-vowel-adjacent R in historical verses, which was one of the motivations for this pilot study.

The paper is organized as follows. In Section 2, we introduce the Sonority Sequencing Principle governing syllabification according to sonority of phonological segments. Section 3 describes implementation of this principle into an automatic sonority parser that identifies potential syllable-projecting segments according to sonority. The parser enables us to pick up all the instances of potential syllabic R in the contexts CRC and CR#. In Section 4, we show and discuss the results we got by applying the parser to six syllabic verse texts from the 14th to the 16th century. In Section 5, the results of our research are discussed.

2 SONORITY SEQUENCING PRINCIPLE

In derivational approaches to phonology, syllabification algorithm is governed by the Sonority Sequencing Principle (Selkirk 1984; Clements 1990), according to which the syllable structure is derived in terms of sonority.

As for sonority, two major categories of segments are identified, i.e., vowels and consonants, the latter being further subcategorized into obstruents and sonorants.

These three sonority categories form a hierarchy, shown in Fig. 1, in which sonorants (R) are between the more sonorous vowels (V) and the less sonorous obstruents (T).

\+ V >> R >> T −

Fig. 1. Sonority hierarchy: V>>R>>T

The Sonority Sequencing Principle (SSP) postulates that the sonority of a syllable decreases from the nucleus towards the margins, i.e., to onset and coda. Thus, according to this principle, syllable nuclei are the sonority peaks of words. And since it is the nucleus that constitutes a syllable, the number of syllables of a word is equal to the number of (syllabic) nuclei in the word, and, transitively to the number of sonority peaks. These default sonority-syllable correspondences are summarized in Fig. 2.

number of syllables = number of nuclei = number of sonority peaks

Fig. 2. Sonority-syllable correspondences

From the perspective above, the diachronic evolution in which sonority peaks based on sonorants /r l/ (henceforth R-peaks) are syllabified as nuclei, can be understood as a path towards optimal syllabification. As illustrated in Fig. 3, the newly created bisyllabic structures *nesl* 'he carried' and *slza* 'tear' are fully in accordance with the sonority-syllable correspondences predicted by the SSP. Each of the sample words has two relative sonority peaks (shaded): one is created by a vowel and the other by a sonorant. Both the V-peak and R-peak project syllable nuclei (N_σ) regularly, yielding thus optimal bisyllabic structures.

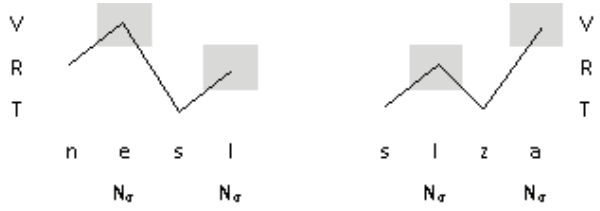

Fig. 3. Optimal syllabification: syllabic CR# and CRC

The R-peaks in CRC and CR#, which pattern with V-peaks in terms of syllabification, contrast with R-peaks in #RC. In a mono-syllabic word *lhal* 'he lied', for example, only the V-peak (occupied by *a*) projects the nucleus, but not the initial R-peak. Furthermore, a comparison of the monosyllabic *lhal* (on the left in Fig. 4) and the bisyllabic *udal* 'he provided' (on the right) reveals that V-peaks always project nuclei, even word-initially.

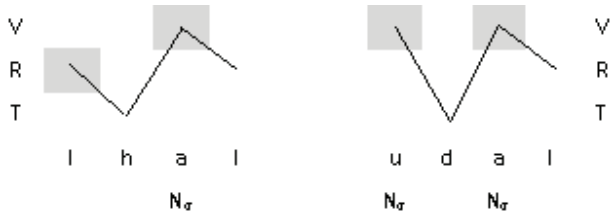

Fig. 4. Non-syllabic #RC: R-peak ≠ Nσ vs syllabic #VC: V-peak = Nσ

To sum up, the idea we pursue is that syllable structure of words is read off from their sonority profiles: the universal principle is that sonority peaks are syllabified as nuclei. While V-peaks are nuclei by default, R-peaks are syllabified upon language-specific parameters. To put it simply, all languages feature syllabic vowels, but only a subset of them have syllabic consonants.

In the certain stage of history of Czech, Czech had the parameter on syllabic consonants set positively: R-peaks do project syllable nuclei. However, the projection of a syllabic nucleus by an R is dependent on R's position (word-initial R-peaks never project syllables), and, also, the situation changes over time (in case of final and medial R-peaks). In the next section, we test these assumptions on a relatively large historic data sample.

3 AUTOMATIC SONORITY PARSER

To see the sonority peaks and their contribution to the syllable structure, we need to annotate words (and syllables) with respect to their sonority. For this purpose, we created an automatic sonority parser.[1] The parser has two main ingredients: the sonority hierarchy and the inventory of phonological segments classified according to the sonority scale.

The hierarchy embodied in the parser categorizes sounds into 7 sonority levels. As shown in Tab. 2, this fine-grained hierarchy, based on Parker (2011), identifies three subclasses of sonorants; the liquids /r l/ (that are our main concern in this paper), are in the 'middle', surrounded by the less sonorous nasals and the more sonorous glides. Thus, /r l/ are sonorants of level 4 in Tab. 2, and hence we refer to them as R_4 from now on.

[1] The parser is available at https://github.com/cechradek/analysis_of_syllables_in_old_czech/blob/main/02SLABIKA_sonoritni_profily.py.

[2] Line-by-line glosses: 'tear, elm'; 'to provide'; 'apple'; 'wool, brother'; 'elm, he carried'; 'wool, tear'; 'brother'.

son-level	phon-class	phon-subclass	example[2]
7	V	low and mid vowels	slz**a**, jil**e**m
6	V	high vowels	**u**dat**i**
5	R	glides	**j**ablko
4	R	liquids	v**l**na, b**r**at**r**
3	R	nasals	jile**m**, **n**esl
2	T	fricatives	**v**lna, **sl**z**a**
1	T	plosives	**b**rat**r**

Tab. 2. 7-point sonority hierarchy

The second ingredient the parser considers is a sonority-annotated set of segments that were part of the phonological system of the 14th–16th century. For convenience, the segmental inventory, compiled from the historical grammars (see Section 1 for the references), is shown in Tab. 3 (IPA annotated), and it is supplemented by the corresponding graphemes.[3] (We should add that the table displays only those graphemes that are recorded in the edited texts analyzed in this paper.)[4]

son-level	segments	graphemes
7	/a aː e eː o oː/	a á e ě é o ó
6	/i iː u uː/[4]	i y í ý u ú ů
5	/j/	j
4	/r l/	r r’ l l’
3	/m n ɲ/	m n ň
2	/f v s z ʃ ʒ ɾ x ɦ/	f v s z š ž ř ch h
1	/p b t d t͡s t͡ʃ c ɟ k g/	p b t d c č ť ď k g

Tab. 3. Segmental inventory of 14th–16th century

In Fig. 5, the outputs provided by the parser are illustrated. The diagram displays the sonority profile of a word *milosrdenstvie* ‘mercy’: the segments (on the horizontal axis) were mapped onto the 7-point sonority hierarchy (on the vertical axis).

[3] The segmental inventory underwent several changes in the 14th–16th century, which however did not result in reordering of segments with respect to the sonority levels. The consonant inventory was simplified to the extent that palatalized consonants merged with their plain counterparts, e.g. /nʲ/ merged with either /n/ or /ɲ/. Since the input and the output of this diachronic change (called depalatalization in the historical literature) are always on the same sonority level, only the output segments are involved in the parser. The similar strategy was used for the vocalic part of the parser: the reported historical merger of high vowels (of any length) /i y/ is represented by the output /i/.

[4] In addition to monophthongs, high vowels were involved in rising diphthongs /uo/ and /ie/, the latter having been a reflex of a Common Slavic long vowel, called *jať*; for details see Kosek and Ziková (2022). In terms of syllable structure, these diphthongs count as a single nucleus, similarly to monophthongs. That is exactly what is predicted by the fine-grained sonority scale: since high vowels sit lower in the sonority hierarchy than mid vowels, only the latter project the V-peaks in the diphthongs /uo/ and /ie/. See also Fig. 5.

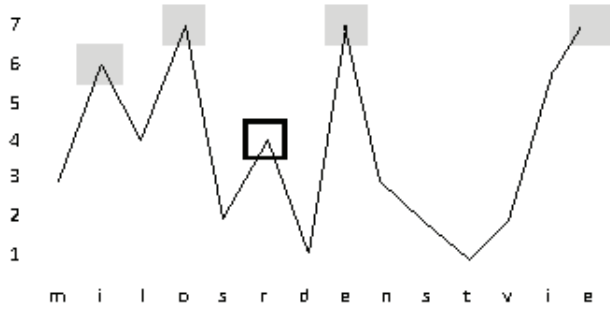

Fig. 5. Sonority profile of *milostrdenstvie* 'mercy'

For the given word (***milo****srd****e****ntsv****ie***) shown in the diagram, there are 5 vowels that correspond to 4 sonority peaks (shaded); all of them are either V_6-peaks or V_7-peaks. The difference in the number of the vocalic segments and the sonority peaks is due to the word-final bi-vocalic string *ie*: *e* is more sonorous than *i,* and therefore projects a single V_7-peak. This is in line with the fact that *ie* is one of the two rising diphthongs in OCz; see also footnote 4. In addition to the vocalic peaks, *milosrdenstvie* contains a consonantal peak (squared): it is a R_4-peak, projected by /r/. (Notice that there is yet another instance of the R_4 in the given word, i.e., /l/. Since it appears intervocalically, in this position it does not project a sonority peak.)

Diagrams generated by the parser (as the one above), are inputs for the analysis of syllable structure. As mentioned previously, the analysis pursues the idea that V-peaks are syllable nuclei by default, but the syllabification of consonantal peaks is parameterized. The parameter is assumed to be set so that the consonantal peaks of type R_4 can be syllabified as nuclei in a certain point in the history of Czech. This means that the word *milosrdenstvie* either could have five or four syllables in the examined historical period, depending on whether the /r/ was syllabified the same way as the V-peaks – or not.

To conclude, the sonority parser automatically identifies words with R_4-peaks that could project syllable nuclei in various stages in the history of Czech. To identify syllable-projecting R_4-peaks, i.e., syllabic instances of /r l/, we have been tracing the behavior the words with R_4-peaks show in the syllable-counting verses. The method used, as well as the results are described in the next section.

4 DATA FROM THE SONORITY PARSER

In the 14th–16th century, much of the poetry is syllable-counting; thus, verses have a regular number of syllables – and, therefore, a regular number of nuclei. In particular, the most common verse is octosyllable (Jakobson 1932). We thus follow this line of thinking: the appearance of a word like *milosrdenstvie* in the 8-peak verse

indicates that its R_4-peak /r/ counts as a verse unit, i.e., projects the syllable nucleus – to keep the octosyllabic rhythm.

We applied this method to a corpus of 15,837 verses extracted from six texts from the 14th–16th century; see the details listed in Tab. 4. The right-most column displays the total number of verses.[5]

source text	century	number of verses
Kunhutina modlitba [KunM] 'Kunhuta's Prayer'	14th	154
Alexandreis – Budějovický fragment [AlexB]	14th	342
Život Svaté Kateřiny [SvKat] 'The Life of Saint Catherine'	14th	3,518
Alexandreis – Svatovítský fragment [AlexSv]	15th	2,462
Hádání Prahy s Kutnou Horou [Had] 'Disputation between Praha and Kutná Hora'	15th	2,989
Instrukcí Šimona Lomnického z Budče [Lom] 'Instructions by Simon Lomnicky'	16th	6,372

Tab. 4. Analyzed texts

The texts were run through the parser, and the data provided by the parser confirm a strong tendency towards octosyllable. As shown in Tab. 5, the proportion of 8-peak verses, identified by the parser, does not decline below 85%.

source	total number of verses	proportion of 8-peak verses
KunM	154	93%
AlexB	342	95%
SvKat	3,518	89%
AlexSv	2,462	85%
Had	2,989	95%
Lom	6,372	97%

Tab. 5. Proportion of 8-peak verses

Given their octosyllabic nature, the selected texts are thus a good ground for verifying the assumptions of the historical grammars. In particular, we assume that: 1. #RC are never syllabic, 2. CR# turn from non-syllabic to syllabic, 3. CRC vary between syllabic and non-syllabic parsing.

[5] The analyzed corpus is built on the following critical editions: *Kunhutina modlitba* [KunM]; available at https://vokabular.ujc.cas.cz/moduly/edicni/seznam-edic/datace-asc/strana [14/03/ 2023]; Vážný, V. (1963). *Alexandreida* [AlexB; AlexSv]. Praha: Nakladatelství československé akademie věd; Hrabák, J., and Vážný, V. (1959). *Dvě legendy z doby Karlovy* [SvKat]. Praha: Nakladatelství Československé akademie věd; Daňhelka, J. (1952). *Husitské skladby Budyšínského rukopisu* [Had]. Praha: Orbis; Heřmanská, K. (2016), *Instrukcí aneb Krátké naučení hospodáři mladému Šimona Lomnického z Budče (edice a literárně historický rozbor)* [Lom]. Master Thesis, UK Praha.

The absolute non-syllabicity of #RC is clearly confirmed: we recorded 40 words with initial R_4-peaks and neither of them is involved in the 8-peak verse as a syllabic nucleus.

As for CR#, they predominantly occur in 9-peak verses: 190 vs 34 in 8-peak verses. Since 9-peak verses do exist in the analyzed texts, we might simply conclude that 9-peak verses with final R_4-peaks (found in words like *mysl* 'mind', *mistr* 'master' or *spadl* 'he fell') violate the octosyllabic rhythm. Under this assumption, thus, the 9-peak verses are indeed 9-syllabic, and the word-final syllabic /r l/ are projected as an extra syllable on top of the octosyllable. The second possible approach is to take 9-peak verses as an indicator of non-syllabicity of R_4-peaks. Under this approach, 9-peak verses follow the octosyllabic rhythm regularly, because the final R_4-peaks are not syllabified as nuclei. From this perspective, thus, the attested words like *mysl, mistr* and *spadl* are monosyllabic, and only the V-peaks project nuclei.

We favour the latter approach, thus, 9-peak verses contain non-syllabic R_4-peaks; there are two reasons for that. First, the analyzed texts have a strong tendency to – indeed – be octosyllabic (proven by the proportions in the Tab. 5 above). Moreover, 9-peak verses tend to contain R_4-peaks: in our corpus, more than 60% of 9-peak verses include R_4-peaks.

In sum, the considerable difference between syllabic and non-syllabic CR# (34 vs. 190) suggests that the postulated diachronic process resulting in syllabicity of CR# proceeded relatively slowly in the examined period of 14th–16th century and that most of the final R_4-peaks remained non-syllabified.[6]

Following the same logic, we classify CRC as syllabic or non-syllabic according to their distribution in 8-peak and 9-peak verses. In this case, the distributional discrepancy between the two classes is not as sharp as the one in CR#. This, however, is expected by the historical grammars: many instances of the syllabic CRC were inherited from Proto-Czech, hence syllabic CRC could appear even in the texts from the very beginning of the 14th century This expectation is validated by the data: one of the two earliest texts in our corpus, i.e., *Kunhutina modlitba*, contains 5 instances of syllabic CRC in 8-peak verses (***krm****iteľu* [v.16] 'feeder.voc.sg', ***prv****niemu* [v.67] 'first.dat.sg', *čt****vrt****ému* [v.70] 'forth.dat.sg', ***mrt****vých* [v.80] 'dead.gen.pl', ***krm****ě* [v.87] 'food.nom.sg').

Similarly to CR#, CRC are supposed to gravitate to syllabicity in the course of time – and that should be manifested in two things. First, we expect the original syllabic CRC to retain their syllabic status, hence appearing in 8-peak verses. This holds for all the four syllabic CRC-roots bolded above. They are involved in various word types and tokens that are distributed quite evenly throughout our corpus, and they predominantly appear in 8-peak verses.

[6] It is not surprising that 33 out of 34 instances of syllabic CR# are found in the texts from 15th and 16th century, but not in earlier texts from the 14th century. Though, even in the later texts, the non-syllabic CR# are still prevalent.

The second hypothesis is that non-syllabic CRC-roots should turn into syllabic roots. This diachronic change is well-documented by the root *krv* 'blood', recorded quite widely in our corpus. The data in Tab. 6 show an obvious shift in the distribution towards 8-peak verses, in which, as we assume, the root *krv* is syllabic.

source	8-peak verses	9-peak verses
KunM	0	1
AlexB	0	2
SvKat	1	7
AlexSv	2	7
Had	10	0
Lom	4	0

Tab. 6. Tokens with the root *krv*

To conclude, the data confirm the hypotheses concerning the development of syllabic R: word-initial #RC are not syllabic at all, word-medial CRC and word-final CR# are both non-syllabic or syllabic. Through the examined period of the 14th–16th century, there is a tendency towards R-syllabicity, which is, however, stronger in the word-medial context than in the word-final context.

5 CONCLUSIONS

In this paper, we investigated development of syllabic consonants in Czech. We followed a view that syllabic consonants – in accordance with vowels – are sonority peaks that project syllable nuclei. On this backdrop, we were tracking the sonority peaks formed by sonorants /r l/ and observed their behavior in syllable-counting verses: if they contribute to the regular octosyllabic rhythm, they are proven to be syllabic.

We created a sonority parser that automatically identifies /r l/ as sonority peaks in three contexts: #RC, CR#, and CRC. The parser was applied to 6 verse texts from 14th–16th century (all of the texts have a strong tendency towards the octosyllable). The data provided by the parser were analyzed according to the following criteria: 1. the octosyllable corresponds to 8 sonority peaks; 2. if one of the 8 peaks is /r/ or /l/, they are syllabic consonants; 3. if the octosyllabic verse has 9 sonority peaks and one of them is /r/ or /l/, then the /r/ or /l/ are not syllabic consonants. The synoptic picture of the data we obtained is shown in Tab. 7.

	non-syllabic	syllabic	tendency to R-syllabicity
CRC	158	968	high
CR#	190	34	low
#RC	40	0	none

Tab. 7. Syllabicity of R-peaks

The results of our pilot study reveal a distinction between CRC and CR# contexts and this observation, to our knowledge, has not yet been documented: CR# show a weaker tendency towards syllabicity than CRC. In other words, the change from non-syllabic to syllabic R proceeded more rapidly in word-medial position than in word-final position.

The revealed dynamic of the change of syllabic structure is in line with a cross-linguistic observation made by Cyran (2010): non-syllabic CRC are more marked typologically than non-syllabic CR#. In other words, the more marked non-syllabic CRC turn to regular syllabic structures more rapidly than the less marked non-syllabic CR#.

ACKNOWLEDGEMENTS

The research has been supported by the Czech Science Foundation (grant No. 23-04719S *Development of Syllabic Sonorants in Czech*).

References

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston – M. Beckman (eds.): Papers in laboratory phonology 1: between the grammar and physics of speech. Cambridge: Cambridge University Press, pages 283–333.

Cyran, E. (2010). Complexity scales and licensing in phonology. Berlin, New York: De Gruyter Mouton.

Gebauer, J. (1963). Historická mluvnice jazyka českého I. Hláskosloví. 2nd ed. Praha: ČSAV.

Jakobson, R. (1932). Verš staročeský. In Československá vlastivěda 3 – Jazyk. Praha: Sfinx, pages 429–459.

Komárek, M. (1982). Nástin fonologického vývoje českého jazyka. Praha: SPN.

Kosek, P., and Ziková, M. (2022). Czech Vowel Fronting (Česká Přehláska). In M. L. Greenberg (ed.): Encyclopedia of Slavic Languages and Linguistics Online. Leiden: Brill. Accessible at: https://referenceworks.brillonline.com/browse/encyclopedia-of-slavic-languages-and-linguistics-online.

Lamprecht, A., Šlosar, D., and Bauer, J. (1986). Historická mluvnice češtiny. Praha: SPN.

Parker, S. (2011). Sonority. In M. Oostendorp et al. (eds.): The Blackwell Companion to Phonology. Vol. II: Suprasegmental and Prosodic Phonology. Oxford: Blackwell, pages 1160–1184.

Scheer, T., and Ziková, M. (2017). Branching onsets in Old Czech. In O. Mueller-Reichau – M. Guhl (eds.): Aspects of Slavic Linguistics: Formal Grammar, Lexicon and Communication. Berlin: De Gruyter, pages 285–309.

Selkirk, E. (1984). On the Major Class Features and Syllable Theory. In M. Aronoff – R. Oehrle (eds.): Language Sound Structure. Cambridge, Mass.: MIT Press, pages 107–136.