10th International Conference on Speech Prosody 2020
25-28 May 2020, Tokyo, Japan

Accent-Groups vs. Stress-Groups in Czech Clear and Conversational Speech

Jan Volín, Radek Skarnitzl

[1]Institute of Phonetics, Charles University, Czech Republic
jan.volin@ff.cuni.cz, radek.skarnitzl@ff.cuni.cz

Abstract

Descriptions of Czech prosody have operated with the term stress-group (foot) for more than a century, probably under an undeclared influence of poetry analyses, but also due to the focus on speech styles that require high clarity. In conjunction with this, however, it has been repeatedly evidenced since the 1970s that the Czech lexically stressed syllable does not exhibit any clear acoustic manifestations, even though division of the speech continuum into units smaller than intonation phrases is undeniable. The present study addresses the clash between the formerly established rules of stress-grouping and the reality of the current Czech language.

Analyses were carried out of multiple samples of casual conversational speech and the speech of news readers where high clarity is an imperative. Although no significant departure from the current modal prosodic norms was perceptible through routine observations, analyses of the two speaking styles suggested differing strategies to prosodic parsing. Most noticeably, conversational speech produced longer stress-groups, but at the same time shorter intonation phrases than clear speech. Also, substantial numbers of stress-groups were found that contradicted the traditional model of stress-grouping in Czech. Potential changes in terminology are discussed.

Index Terms: speech styles, stress-group, accent-group, melodic cues, clear speech, conversational speech, Czech.

1. Introduction

Lexical stress in Czech, a West Slavic language spoken by approximately 10 million people, is fixed to the first syllable of the prosodic word. Stress placement is independent of vowel length and vowel quality (i.e., any of the 13 Czech vowels – short, long, or diphthongal – may appear in stressed, as well as unstressed syllables), and there is no relation of lexical stress to the words' morphological structure. The word *nejstálejší* [ˈnɛjstaːlɛjʃiː] (*the most durable*) serves as an example: the first stressed syllable *nej–* is a prefix, has a phonologically short vowel while the word has two long ones.

The definition given above appears to be straightforward but, upon closer examination, many questions remain with respect to stress placement in Czech. The controversy hinges on the definition of the prosodic word, or stress-group. As many other languages, Czech tends to avoid the so-called stress clash; in other words, speakers usually avoid two consecutive stressed syllables. That means that monosyllabic words are likely to join the neighbouring word as enclitics, and form a stress-group with it, as shown in (1).

Dřív než jsem ho poprosil, Pavel mi to podal. (1)
[ˈdr̝iːf nɛʃ sɛm ɦo ˈpoprosɪl | ˈpavɛl mɪ to ˈpodal ‖]
Before I him asked Paul me it gave

In Czech, the so-called "genuine" monosyllabic prepositions become head of the stress-group: they take the stress from the following word, as shown in (2).

Od pondělí do pátku pracuje na poště. (2)
[ˈʔot poɲɟɛliː ˈdo paːtku | ˈprat͡sujɛ ˈna poʃcɛ ‖]
From Monday till Friday (he) works at post office

The rule about placing stress on monosyllabic prepositions is not strictly adhered to in Czech. If the following word is long and the resulting stress-group, with stress on the preposition, would have been too long, speakers are likely to move stress to the following word [1], [2], as exemplified in (3). Other factors also seem to contribute to the stress shift from the preposition to the following word [2].

Petr přišel na zajímavou přednášku. (3)
[ˈpɛtr̩ ˈpr̝̊ɪʃɛl na ˈzajiːmavo͡u ˈpr̝̊ɛdnaːʃku ‖]
Peter came to interesting lecture

The above-mentioned examples illustrate regularities in spoken Czech and corroborate the hypothesized significance of rhythmical configurations in speech. Speakers appear to adjust the length of adjacent stress-groups in the larger phonetic context, so as to achieve a more regular rhythmical configuration, or a greater eurhythmy [3, p. 273].

Examples (1)–(3) all address monosyllabic words and how they join with neighbouring words into prosodic words. Informal observations indicate that similar "clustering" into stress-groups is not limited to monosyllabic words. However, such processes have not been captured in phonetic books or scientific papers dealing with Czech, until it was proposed in a recent handbook [4]. Traditionally cited sources [5], [6] and [7] provide sets of rules to handle monosyllables, but claim that once a word has two or more syllables, it will retain its lexical stress and form an independent stress-group, albeit with a possible clitic adhering to it. These authoritative sources are not necessarily wrong – they are most probably based on observations of formal, well-prepared monologues, possibly even public reciting of poetry. One thing is clear though: they are not based on replicable analysis of a sample of speech recordings. (Only [5] actually specifies the analysed material, but that comprises written texts divided into stress-groups according to theoretical presumptions by the author and students.)

The important outcome of traditional descriptions of Czech stress-grouping is that they rule out other than monosyllabic clitics; in other words, disyllabic and longer words cannot lose their stress and aggregate into a stress-group as stressless. Thus, they would predict division as in (4).

Petrovi bylo nějak divně, tak jsme ho odvezli k lékaři. (4)
[ˈpɛtrovɪ ˈbɪlo ˈɲɛjag ˈɟɪvɲɛ | ˈtak smɛ ho ˈʔodvezlɪ ˈk leːkar̝ɪ]
Peter felt somehow strange so we him brought to doctor

10.21437/SpeechProsody.2020-142

Our observation of current Czech, on the other hand, would allow even for the stress-grouping given in (5).

Petrovi bylo nějak divně, tak jsme ho odvezli k lékaři. (5)
[ˈpɛtrovɪ bɪlo ɲɛjag ˈɟɪvɲɛ | ˈtak smɛ ho ʔodvezlɪ ˈk leːkar̝ɪ]
Peter felt somehow strange so we him brought to doctor

Thus, against the traditional division 3+2+2+2 | 3+3+3 as in (4), we suggest 7+2 | 6+3 in (5) as a possible outcome. (The numbers correspond to counts of syllables in each stress-group.)

It is clear that the actual clustering will depend on a number of factors. Apart from contextual and co-textual requirements, i.e, the specific unique semantics of the given utterance, eurhythmy often plays a role. In [4], many examples are listed where it is easily imaginable for stretches of words to aggregate into longer groups. However, like in previous studies, a rigorous analysis of authentic speech recordings is absent. Therefore, the objective of the present study is to analyse a sizeable sample of natural speech production to either confirm or modify the existing models of stress-grouping in Czech.

2. Method

2.1. Material

Two speaking styles were represented in the analysed sample: a plain conversational one and the so-called clear speech represented by radio news reading. For the latter we retrieved authentic recordings of news-bulletins from a national broadcaster (Czech Radio) read by various voices over the past few years. The speaking style of professional news readers is supposed to be as clear as possible for the news to have any public acceptance, and current Czech Radio staff is to a large extent a guarantor of model standard Czech speech production without any noticeable mannerisms, colloquialisms or salient idiosyncrasies. All our news readers were active, experienced middle-aged employees of the broadcaster.

Conversational speech was recorded at the Institute of Phonetics in Prague. Undergraduate students were asked to bring in their friends with whom they first read out the transcripts of the news bulletins and then they discussed individual items of the news and personalities of politicians appearing therein. The recording took place in a quiet, comfortably furnished office with a short natural reverberation. High-quality lavalier microphone Sennheiser E604 was plugged directly to a portable recorder set to uncompressed 48 kHz 16-bit mode. Although for the purposes of our study the high quality of the recording was not as crucial as it would be, for instance, in the case of spectral characteristics investigation, care was taken to achieve highly intelligible disturbance-free sound tracks.

Three male and three female newsreaders were taken at random from the radio corpus. In the case of the conversational speech, we took care to select only relaxed interactions. (Occasionally, some of the pairs did not manage to produce easy-going flow of speech. They remained 'stiff' during the recording probably because they were too self-conscious with the microphone.). Additionally, initial two to three minutes of the dialogues were excluded from all the conversations in order to focus on spontaneous flow. Table 1 summarizes the characteristics of our sample.

With regard to our focus, we felt it necessary to extract the values of mean articulation rate (AR), i.e., the speed of articulation with exclusion of all pauses, because potential differences in unit divisions could be related to that. The mean AR in news reading was 15.1 phones per second (ph·s^{-1}), whereas in conversations it was 14.8 ph·s^{-1}. It appears that the mean ARs in the two studied samples are comparable.

Table 1: *Selected descriptors of the sample. Condition Clear refers to news reading, Conv. to conversational speech. Number of words produced is in the 4th column, articulation rate (AR) is in phones per second.*

Speaker	Gender	Condition	n words	AR
NRF1	F	Clear	481	14.3
NRF2	F	Clear	522	15.0
NRF3	F	Clear	581	14.6
NRM1	M	Clear	475	15.2
NRM2	M	Clear	610	15.5
NRM3	M	Clear	428	15.7
CSF1	F	Conv.	380	13.9
CSF2	F	Conv.	550	16.9
CSF3	F	Conv.	499	14.8
CSM1	M	Conv.	399	15.1
CSM2	M	Conv.	430	13.9
CSM3	M	Conv.	523	14.2

2.2. Material pre-processing and analyses

Most of the speech sample handling was carried out in the software Praat [8]. All recordings were carefully manually segmented on levels of phones, syllabic nuclei, words, stress-groups and prosodic (intonation) phrases. As to phones, they were first force-aligned using the Prague Labeller [9] with the orthographic text as input, but during the manual corrections only the actually pronounced segments were left. Syllable nuclei are typically vowels (10 monophthongs and 3 diphthongs), but the Czech language allows for syllabic liquids as well. These were labelled by a Praat script but checked during manual corrections of phone boundaries.

The definition of words is often a matter of debate among linguists, but we opted for conventional orthographic cues – a word in our study is a unit that is separated by spaces from other words in conventional spelling. Such item is also a potential entry in a typical dictionary. (For number of words in our sample see Table 1 above.) Stress-groups and prosodic phrases were established by careful auditory inspection carried out by the two authors and a PhD student who did not know the purpose of the current research. Occasional disagreements between labellers were negotiated with repeated listening. Praat scripts were used to summarize counts of the constituting elements of words, stress-groups and prosodic phrases.

3. Results

3.1. Stress-group composition

Figure 1 displays percentually normalized frequencies of the occurrence of mono- and multi-word stress-groups. It is obvious at first glance that the stress-groups containing just one word are most numerous and that no stress-groups with six or more words in them were found in the material. In fact, there were no five-word stress-groups in the clear speech sample either. The prevalence of single-word stress-groups reached 71.3% in news reading, but only 58.4% in conversations. A similar difference but in the opposite direction was found for three-word stress-groups (SGs): only 3.3% in news reading, but 11.4% in conversations. These differences in the distribution of

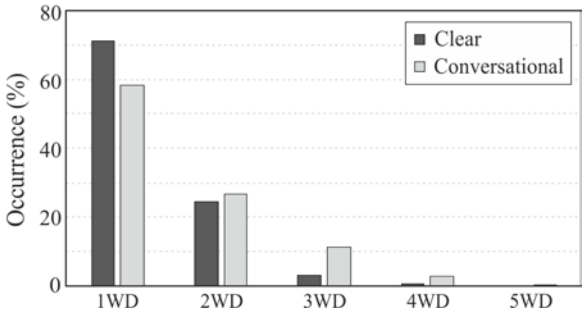

Figure 1: *Percentual representation of stress-groups of varying length in words in the two speech styles. (Clear = news reading)*

SGs of varying lexical complexity were found statistically significant: $\chi^2(4) = 11.15$; $p = 0.025$ (testing goodness of fit).

If the length of stress-groups in the sample is expressed in syllables rather that words, the shape of the distribution changes (Figure 2), but it is still interpretable. It is perhaps useful to realize that the word is a semantic unit by essence, while the syllable is more of a structural or constructional unit. This might provide a cue in hypothesizing why the conversational style possesses fewer short stress-groups if measured in words, but more short stress-groups if measured in syllables. One- and two-syllable SGs in news reading occurred in 34.9% of the cases, while in conversations they represented 43.7% of all SGs (see the first two pairs of columns in Figure 2). For three- and four-syllable SGs the situation is reversed. Figure 2 also suggests that no stress-groups in our sample were longer than 10 syllables, but anything above 6 syllables is relatively rare. Stress-groups of five and more syllables occur in roughly equal proportion in both speech styles.

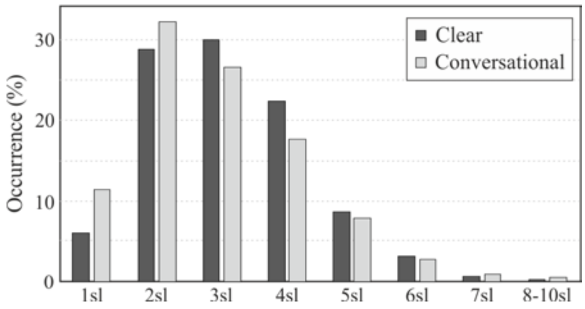

Figure 2: *Percentual representation of stress-groups of varying length in syllables in the two speech styles.*

It might be informative at this point to compare our results with the existing and often cited older statistics based on theoretical assumptions (see above, [5]). Figure 3 shows the comparison of three-, four- and five-syllable stress-groups, where the differences were largest. While the three-syllable theoretical SGs "outperform" our material since they are favoured by the stipulated rules, both our samples produced greater numbers of four- and five-syllable SGs. (The same trend held for 6-syll. and longer SGs without an exception, but the frequencies of occurrence of those were quite low.)

The greatest tension between the existing theoretical (i.e., traditional) models and our data can be expected to concern clustering of polysyllabic into SGs. The traditional models predict that each such word will form a stress-group on its own or will accept a monosyllabic preposition and/or a monosyllabic clitic item. In the following stage of analyses, we focus on stress-groups consisting of two words only to test the prediction.

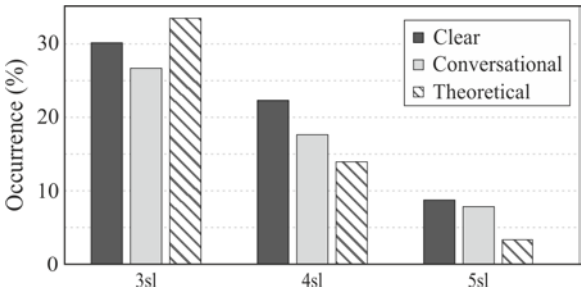

Figure 3: *Percentual representation of three-, four- and five-syllable stress-groups in the two speech styles and a rule-based analysis of a written text.*

The two-word SGs (represented as the second pair of columns in Figure 1) can show whether polysyllables only combine with monosyllables as described in literature, or whether some other options exist. Our sample provided 1028 SGs consisting of two words of which 560 were produced in news reading and 468 in dialogues.

The first column of Table 2 lists twenty-one two-word configurations that were found in the material. In the *x+y* expression, *x* refers to the number of syllables in the first word, *y* to the number of syllables in the following word in the stress-group. The configuration 0+*y* refers to non-syllabic prepositions, i.e., the prepositions that do not form a syllable on their own and, instead, they adhere to the syllabic onset of the following word. There are four such words in Czech: *k* (*to, towards*), *s* (*with*), *v* (*in*) and *z* (*from, out of*). These, however, do not pertain to the main focus of our current study.

Table 2: *Numbers of occurrences of two-word SGs classified by word lengths in syllables.*

Configuration	**Clear**		**Conversational**	
	n	%	*n*	%
0+x	112	20.00	37	7.91
1+1	27	4.82	112	23.93
1+2	103	18.39	73	15.60
1+3	90	16.07	30	6.41
1+4	43	7.68	10	2.14
1+5	11	1.96	5	1.07
1+6	2	0.36	1	0.21
2+1	41	7.32	72	15.38
2+2	40	7.14	55	11.75
2+3	10	1.79	5	1.07
2+4	1	0.18	1	0.21
3+1	29	5.18	28	5.98
3+2	17	3.04	19	4.06
3+3	5	0.89	1	0.21
3+4	0	0.00	1	0.21
4+1	15	2.68	10	2.14
4+2	9	1.61	2	0.43
4+3	0	0.00	1	0.21
5+1	4	0.71	3	0.64
6+2	1	0.18	0	0.00
6+1	0	0.00	2	0.43

Our chief concern is the occurrence of configurations with numbers (i.e., syllable counts) greater than 1. That is, we know that 0+*y*, 1+*y* and *x*+1 are legal and well described two-word SGs. Other cases, however, would be considered illegal by [5] and [6], and improbable by [7]. Our results reveal that exactly these cases accounted for 14.3% in clear speech and 17.7% in dialogues. Clearly, such large proportions of 'illegal' structures are not insignificant. (Moreover, for the sake of brevity and clarity we are considering only stress-groups formed by two words, which represent slightly over one quarter of our material.)

3.2. Prosodic-phrase composition

As repeatedly suggested in [6] and [10], it is potentially misleading to study division of speech into stress-groups without considering a larger context. Therefore, we explored the composition of prosodic phrases in terms of the number of SGs. Figure 4 displays the outcome.

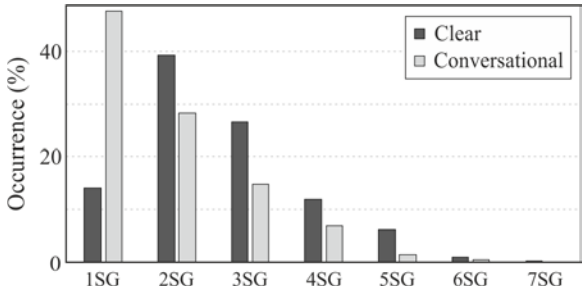

Figure 4: *Percentual representation of prosodic phrases of various lengths counted in number of stress-groups in them, in two speech styles.*

There is a striking, but not inconceivable difference in the utilization of prosodic phrases consisting of just one SG. In conversational speech, such phrases made 47.9% of all phrases produced, whereas in clear speech sample it was only 14.2%. Subsequently, there are much lower relative occurrences of phrases consisting of two or more SGs in dialogues. This disparity is about 11 to 12% for two-SG and three-SG phrases. Very short prosodic phrases in conversational speech are often related to frequent hesitations and pausing – elements that are easy to avoid in read-out speech.

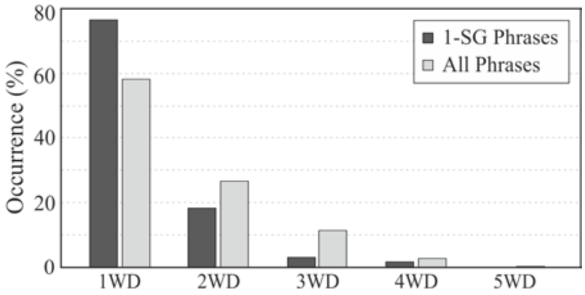

Figure 5: *Percentual representation of stress-groups of varying length in words in conversational speech. Dark columns refer to phrase that comprise one SG only, lighter one all phrases in conversations.*

Interestingly, the very short prosodic phrases in conversations do not behave analogically to the rest of the phrases. Proportions of their lengths in words are different. Figure 5 indicates that if a prosodic phrase is made of just one stress-group, it will likely comprise just one word. It follows that if such phrases were excluded from analyses, the numbers of more complex structure would increase quite substantially. In other words, conversational speech favours longer stress-groups on the one hand, yet on the other hand, short prosodic phrases pull the numbers of complex stress-groups down.

4. Discussion

The chief objective of our study was to map the current speech production in two Czech speaking styles in terms of stress-grouping. The study provides data that are not entirely congruent with traditional descriptions of Czech stress-groups. Older literature predicts a large prevalence of two- and three-syllable stress-groups, but in our sample neither of the speech styles reached 60% of these 'canonical' feet. More importantly though, words of two syllables and longer should not occur in other than stress-group initial positions, but our sample is quite rich in these 'illegal' instances.

One point that could be raised is that of tempo. It is generally argued that faster speech rates lead to fewer prosodic breaks. The differences found in our material cannot be attributed to this: both our speech styles displayed a comparable mean articulation rate. Actually, the news reading was slightly faster (by 0.3 ph·s^{-1}, see Table 1), so it should be less prosodically partitioned. Not only this is not the case, but it is also partitioned differently.

It might seem that the description of Czech stress-groups can be easily amended by mere addition of the more complex clusterings of words which we found in both clear and conversational speech. However, we suggest that a change in terminology is worth considering. Since lexical stress might be conceptualized as an abstract potential of a syllable prominence, stress-group should also be considered a hypothetical unit. It would be perhaps more suitable to talk about an *accent-group* (or possibly *accentual phrase*) in Czech, to refer to actual groups of words joined into one prosodic unit by specific surface manifestations of prominence. This argument is also supported by the fact that the main cue of both internal coherence and external breaks for what we have still called stress-groups, is *F*0 acoustically or melody perceptually (see, e.g., [11] and [12]).

Future research should bring some additional substantiation of our claim. Clearly, more speakers and more speech styles are needed. Apart from expanding our sample, our we would also like to address the rules for joining words into prosodic units in the synthesis system. To the best of our knowledge, Czech speech synthesis relies on the traditional descriptions and we wonder if addition of more complex structures would lead to higher naturalness ratings in perception tests.

To conclude, the current Czech language (in both conversational and news reading styles) forms prosodic units that do not entirely adhere to descriptions in older literature. The explicatory factors could be a) language change, and b) different methodology of research.

5. Acknowledgements

This study was supported by the European Regional Development Fund-Project "Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World" (No. CZ.02.1.01/0.0/0.0/16_019/0000734) and by CUNI project Progres 4, "Language in the shiftings of time, space, and culture".

6. References

[1] J. Zeman, “K přízvukování prvotních jednoslabičných předložek,” *Naše řeč*, 63, pp. 145–149, 1980.

[2] R. Skarnitzl, “O slovním přízvuku na jednoslabičných předložkách v češtině,” *Naše řeč*, 97, pp. 78–91, 2014.

[3] H. J. Giegerich, *English Phonology: An Introduction*. Cambridge: Cambridge University Press, 1992.

[4] J. Volín and R. Skarnitzl, *Segmentální plán češtiny*. Praha: Faculty of Arts, Charles University, 2018.

[5] J. Ondráčková, “O mluvním rytmu v češtině,” *Slovo a slovesnost*, 25/1, 2. pp. 24–29 and 145–157, 1954.

[6] B. Hála, *Uvedení do fonetiky češtiny na obecně fonetickém základě*. Praha: ČSAV, 1962.

[7] Z. Palková, *Fonetika a fonologie češtiny*. Praha: Karolinum, 1994.

[8] P. Boersma and D. Weenink, “Praat: doing phonetics by computer [Computer program],” Version 6.0.47, retrieved 8 February 2019, http://www.praat.org/, 2019.

[9] P. Pollák, J. Volín, and R. Skarnitzl, “HMM-based phonetic segmentation in Praat environment,” *Proceedings of XIIth “Speech and Computer – SPECOM 2007”*, 537–541, 2007.

[10] Z. Palková, *Rytmická výstavba prozaického textu*. Praha: Academia, 1974.

[11] Z. Palková, “Přízvukový takt ve struktuře češtiny,” In: Hladká, Z.–Karlík, P. (Eds.): *Čeština – univerzália a specifika 5,* Praha: Lidové noviny, 399–408, 2004.

[12] T. Duběda, “Prosodic boundaries in Czech: an experiment based on delexicalized speech”, *Proceedings of Interspeech 2006,* 309–312, Pittsburgh, Pennsylvania, 2006.