



# The Impact of Prosodic Position on Post-Stress Rise in Three Genres of Czech

Jan Volín, Radek Skarnitzl

Institute of Phonetics, Faculty of Arts, Charles University  
Czech Republic

jan.volín@ff.cuni.cz, radek.skarnitzl@ff.cuni.cz

## Abstract

In general phonetics, stressed syllables are described as more prominent due to greater duration, higher intensity, less steep spectral slope and/or higher fundamental frequency. However, there are languages in which lexical stress commonly manifests with a post-stress rise (L\*+H). Previous studies dedicated to post-stress rises in Czech were limited in material and/or methodology. Our current study extends the material to sizeable samples of three genres of speech: professional storytelling, poetry reciting and news reading. Over 30,000 syllables were manually labelled in terms of their accent-group status. The main focus of the study was the step between the stressed and the following syllable, but apart from the frequency of occurrence and size of the step, we also examined the influence of the position within a prosodic phrase. The results suggest that the post-stress rise should be considered a typical pitch accent in Czech, but that it does not occur uniformly across the examined genres and prosodic positions.

**Index Terms:** accent-group, Czech, post-stress rise, prosodic phrase, speech styles, speech genres, stress-group.

## 1. Introduction

### 1.1. Research background

Bitonal pitch accents have been reported in a number of languages, particularly in the prenuclear field. In the ToBI framework [1], these accents may be described as, for example, L+H\* where the low (L) tone is leading and the high (H) tone is associated with the stressed syllable, or L\*+H where the low target is associated and the high one is trailing. It is the latter situation, the post-stress rise, which is examined in this study, namely in Czech.

The post-stress rising tone, L\*+H, has been described as “the prenuclear accent par excellence in Greek” [2: 2]. It was found that, in prosodic words of at least three syllables, the L target was aligned near the beginning of the stressed syllable and the H target in the post-stressed vowel [3]. It was later found that lexical stress had no effect on this alignment in the phrase-medial (i.e., pre-nuclear) position [4]. While the prenuclear L\*+H tones appear most frequently in Greek statements, questions and negatives, [2] documented systematic differences in their alignment and scaling between statements and questions – such that listeners were able to discriminate between the two melodies even with the nuclear melody removed from the stimuli.

A similar melodic pattern has been reported in some varieties of English. In a comparison of Orkney and Shetland dialects of English, [5] found that differences in the tonal peak alignment were sufficient for an LDA model to discriminate

between the two dialects, with Orkney English having a low target on the stressed and a high target on the post-stressed syllable. [6] examined the Valleys accent of Welsh English and found that over 70% of pre-nuclear contours had a rising tone, with L\*+H being the most frequent one. In addition, he noted that the post-stress syllable may also manifest as great duration and intensity as the stressed syllable.

The occurrence of the melodic peak on the post-stressed rather than stressed syllable has also been noted for various dialects of Spanish; according to [7], such observation was made already by Navarro Tomás in 1944 and has been supported in a number of studies since. [8] found a slightly greater variability in Standard Peninsular Spanish than was previously reported, but still confirmed L\*+H to be the most frequent contour in the prenuclear field, particularly at phrasal onsets. Post-stress rise was documented in a comparative study of Standard Peninsular Spanish and Catalan [9], as well as in Brazilian Portuguese [10] or Danish [11].

In Czech, the language of interest in the current study, it has been known since the 1970s that listeners often identify as stressed syllables those which have lower fundamental frequency ( $f_0$ ) than neighbouring syllables [12]. The authors of this study suggested that it may not be prominence of the stressed syllable, but rather the acoustic configuration of the entire prosodic word that marks lexical stress in Czech. In a later acoustic study, [13] examined specially designed minimal pairs such as /'svjetlo 'vni:maji:/ ‘they perceive light’ vs. /'svjetlovni: 'maji:/ ‘they have light in her’ and documented a typically rising-falling melodic movement in prosodic words of three or more syllables, with the stressed syllable manifesting lower  $f_0$  than the post-stressed one. In a follow-up study, [14] investigated the speech of radio broadcasters and confirmed the delayed rise (L\*+H) to be the prototypical melodic configuration, in all positions within the prosodic phrase and particularly in stress groups comprising three or more syllables. Apart from these smaller samples of laboratory speech and news reading, an experiment with an extensive sample of unsupervised machine-labelled speech was also carried out, documenting post-stress rise in the source corpus [15]. Finally, in a comprehensive study of Czech lexical stress [16], [17], which included the same speech material uttered in spontaneous speech and, subsequently, in read-out phrases, the results were slightly less conclusive. No melodic movement (i.e., within +/- 0.5 ST) between the stressed and post-stressed turned out to be most frequent, but a post-stress rise was still approximately twice as frequent as a post-stress fall.

The question is, therefore, to what extent the L\*+H tone is genre-specific in Czech. It should not be surprising that speech and particularly intonation recorded in controlled laboratory conditions differs from spontaneous speech. For example, in his study of Standard Peninsular Spanish, [18] showed

significant differences between lab and spontaneous speech in  $f_0$  movements (i.e., in the phonetic make-up of individual tones), in peak alignment, downstepping, and final lowering. In line with our efforts to investigate communicatively “meaningful” genres (i.e., those with a clear communicative intent and with an envisaged audience on the side of the speaker), this study sets out to investigate the post-stress rising tone, L\*+H, in three different genres of Czech. We focus on its presence, melodic extent, as well as positional dependence within prosodic phrases.

## 1.2. Research questions

- Do the three investigated genres differ in parameters of stress-groups?
- Is there a difference among genres in the size of the mean post-stress step?
- How frequent the post-stress rise (i.e., L\*+H pitch accent) actually is in Czech texts of the three genres?
- Is the size of the post-stress step influenced by the position of the accent-group in the prosodic phrase?

## 2. Method

### 2.1. Material

The three selected genres were news reading, poetry reciting and professional story-telling (i.e., narratives). They will be referred to as NSR, POR and PNV, respectively. These genres probably represent the same speech style: clearly articulated monologuing based on a written text. An important aspect of the material is its indisputable communicative intent: the speakers wish to be understood and appreciated.

The recordings representing news reading (NSR) were authentic news-bulletins from a national broadcaster (channels Czech Radio 1 and Czech Radio 2). The current Czech Radio news readers are expected to represent civilized normative speech without any colloquialisms, salient idiosyncrasies or fashionable mannerisms: they are considered guarantors of model Czech speech production. Our NSR sample consisted of 12 such experienced professionals (6 female + 6 male). The news bulletins are typically 3 to 4 minutes long (with voices of correspondents between individual news items excluded). A typical extent of one news bulletin in the sample was 40 sentences, which corresponds to approximately 1,100 syllables and 350 stress-groups (SGs).

The samples of poetry reciting (POR) were recorded in the sound treated studio of the Institute of Phonetics in Prague. The speakers (8 female + 4 male) were volunteering students of philology who expressed an inclination to poetry. They were given several poems and were asked to get familiar with the contents and form of each of them, practice the lines for as long as they needed, and then read the poems out loud as if reciting in front of an audience. Four poems by each speaker were selected into the current dataset. Each speaker produced 750 syllables in about 285 SGs.

The story-telling (PNV) was extracted from audiobooks produced by renowned publishers in professional recording studios. The narrators were experienced actors of national level of prominence. The stories were all written by famous authors (so that the publisher could expect justifiable sales). There were 6 female and 6 male speakers, who produced about 970 syllables or 320 SGs each.

## 2.2. Measurements

First of all, the individually produced prosodic phrases and stress-groups had to be established. Both major and minor (intermediate) phrases were identified by expert auditory analysis (the authors of the study) guided by [19]. The actually materialized prominences, i.e., in the surface stress-groups, rather than some sort of canonical were captured. Our chief concern was a stressed syllable with real, not potential stress. The remaining syllables were also labelled according to their position in the SG, and all SGs were given a label according to their position in a prosodic phrase. *Initial*, *medial*, and *final* SGs were recognized, although in phrases containing only one SG the category initial-final had to be used. These were less frequent and will be referred to as *ini-final*. All labelling was done in Praat [20], and 11,540 stress-groups were found.

It has to be noted that it was not possible to analyze all the produced SGs. First, monosyllabic SGs with or without anacrusis had to be excluded, as there is no post-stress syllable. (We use the term “post-stress syllable” only for syllables after the stressed ones within the same SG.) We also excluded disyllabic final stress-groups because the melodic movement in these is purely functional. After the exclusion of these two categories, 8,934 SGs remained for analysis.

The  $f_0$  tracks were extracted using the autocorrelation method in Praat and manually corrected for octave jumps, missed voiced signal in regions of disturbed regularity, or spurious values. Since the  $f_0$  values relevant for melody perception are found in the second third of the syllabic nucleus [21], we used a Praat script to determine five equidistant points spanning the second third of the syllabic nucleus, and subsequently calculated the arithmetic mean of these  $f_0$  values. In order to approximate human hearing, the values in Hertz (Hz) were converted to semitones (ST).

To tell whether the step between stressed and the following syllable should be considered falling, rising or level, a threshold based on [21] was decided on: all steps of less than -0.5 ST are considered falling, those that exceed 0.5 ST are rising, while steps of equal or smaller than the absolute value of 0.5 ST are considered level.

## 3. Results

### 3.1. Structural properties of spoken texts

The most trivial yet interesting question is that of the length of the stress-groups, expressed in syllables. Table 1 reveals that the maximum length was 10 syllables, and it occurred only in news reading. In poetry, the longest stress group had 7 syllables. However, these long SGs were quite exceptional. Importantly, in all three styles it was three-syllabic SGs that were most frequent, even though in poetry and narratives only marginally. It is worth noting that monosyllabic SGs are also relatively rare in Czech (due to its inflectional character). Statistical significance of the differences was calculated for 2-, 3-, 4-, and 5-syllable SGs using chi-square test (naturally with counts, not percentages):  $\chi^2(6) = 32.05; p < 0.001$ .

The mean lengths of SGs in syllables is displayed in Figure 1. It can be noticed that medial and final SGs (shown in white and dark grey) are always shorter than the initial (black) and ini-final ones (light grey) SGs. In terms of genres, poetry seems to have the shortest SGs, whereas news reading the longest. In news and narratives, only final SGs do not reach the mean length of 3 syllables.

Table 1: Occurrences of stress-groups of various lengths in syllables (1<sup>st</sup> column) for the three genres in percentages (each column adds up to 100%).

n-sylls	poetry	news	narrative
1	7.46	5.44	6.11
2	39.06	27.25	32.20
3	39.66	31.30	32.60
4	11.06	22.43	18.60
5	2.56	8.90	7.34
6	0.18	3.29	2.41
7	0.04	0.94	0.43
8	0.00	0.28	0.25
9	0.00	0.14	0.06
10	0.00	0.03	0.00

This result naturally leads to a question of relative counts of SGs in the initial, medial, final positions and the number of ini-final items, i.e., the stress-groups that were initial and final at the same time as they formed a prosodic phrase on their own. The answer is provided in Figure 2, which clearly indicates that the differences are caused by medial and ini-final SGs. News reading has the lowest number of ini-final SGs and at the same time a relatively high number of medial stress-groups. This is related to the fact that the news bulletins are produced in longer prosodic phrases. The opposite was found out about poetry reciting. Medial SGs are quite rare, while one-SG prosodic phrases (= ini-final SGs) occur in higher counts. Individual verses apparently dictate division into shorter prosodic phrases.

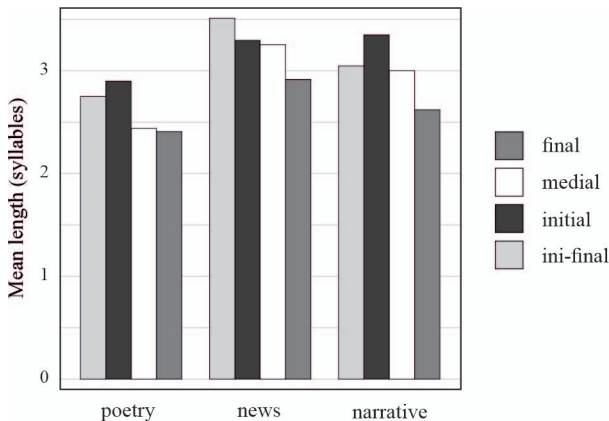


Figure 1: Mean length of stress-groups in syllables by genre and position within a prosodic phrase.

### 3.2. Size of the post-stress step

The mean melodic step from stressed to post-stress syllable in the whole data set was 0.6 ST. This positive value means that the step goes, on average, upwards. The mean step in poetry was 0.66 ST, in news reading 0.78 ST, and in narratives 0.34 ST. Although a one-way ANOVA confirms the difference as significant ( $F [2, 8931] = 33.7, p < 0.001$ ), the benefit of the result for linguistic interpretation is not particularly great. There are too many diverse melodic configurations collapsed into one pool, and averaging rises with falls is conceptually problematic since they form two functional categories.

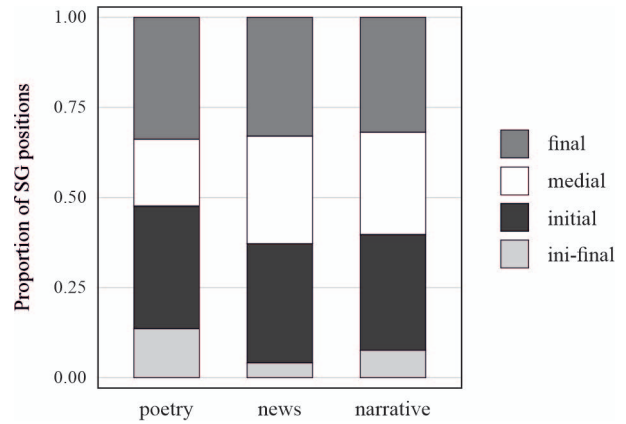


Figure 2: Relative occurrence of stress-groups in various positions within a prosodic phrase. (For 'ini-final' position see text.)

The analysis was, therefore, repeated with rising steps only, i.e., with 4,484 cases (level and falling steps were left out; see Section 2.2 for determining the direction of melodic steps). The results suggest that if a perceptually relevant rise is present, then its overall mean is 2.16 ST. In poetry it is 2.03 ST, in news reading 2.07 ST, and in narratives 2.39 ST. The effect of GENRE is statistically significant:  $F (2, 4472) = 39.9, p < 0.001$ , with narratives differing from the other two genres. Given that after exclusion of the falls and levels the narratives displayed the highest post-stress rise, while before the exclusion the lowest, we can assume that the range of the steps in both directions is generally wider in story-telling. This might be related to the liveliness or dynamics of the contents.

The position of the stress-group in a phrase influenced the mean step between the stressed and post-stress syllable visibly. This can be observed in Figure 3, which shows boxplots of initial, medial and final cases in all three genres. Ini-final SGs were not included this time since they were infrequent and their behaviour was extremely varied.

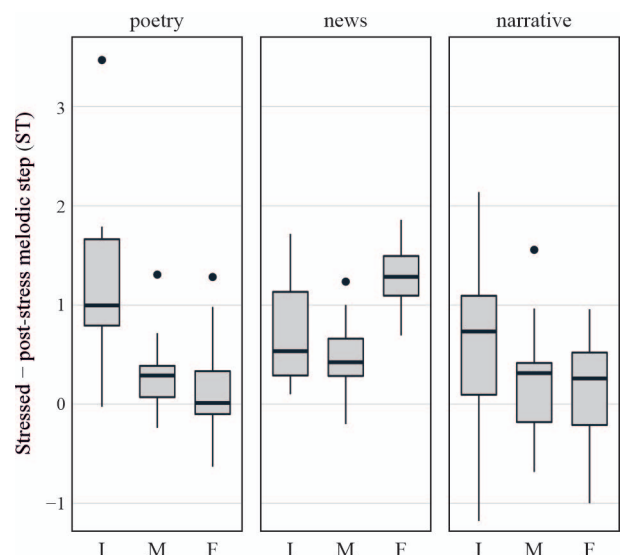


Figure 3: Mean heights of the post-stress steps by genre and position within a prosodic phrase.

An interesting pattern can be observed within the three panels of Figure 3, as well as in their mutual comparison. First of all, the medial step (i.e., the melodic interval in phrase-medial SGs) was always lower and less varied than the initial one, even if the difference is not equally salient in all three genres. Second, the final step is slightly smaller than the medial in poetry and narratives, but actually very prominent in news reading. That resonates with an informally observed mannerism of Czech news readers: some sort of emphatic expansion of the pitch range on the phrase-final stress-group.

Similarly to the previous procedure, we also processed the rising steps on their own. A two-way ANOVA was used with GENRE and POSITION as factors. The effects of both factors were highly significant, but the interaction between them only reached marginal significance:  $F(4, 4224) = 2.01, p = 0.091$ . This means that after the exclusion of level and falling steps, the difference between genres remained only in the height of the step up (the narrative style produces the highest steps), but not in the relative treatment of the initial, medial and final positions. That is, the initial post-stress rise is always the highest, while the medial is always the lowest. The exact values are given in Table 2.

Table 2: Mean heights of post-stress rises in ST by SG position in a phrase and by genre.

Position	poetry	news	narrative
initial	2.31	2.50	2.64
medial	1.53	1.65	2.31
final	1.72	2.05	2.47

### 3.3. Occurrence of post-stress rise

The material offered 8,934 opportunities to post-stress rise, i.e., there were almost nine thousand SGs with a post-stress syllable. The previous analyses already showed that the arithmetic mean only calculates with the size of the steps. The calculation disregards the fact that, generally speaking, a rise is still interpreted as a rise in prosodic structure, even if it is smaller than a fall elsewhere. Therefore, we used the method explained in Section 2.2 to classify all the stressed – post-stress; Figure 4 summarizes the outcome.

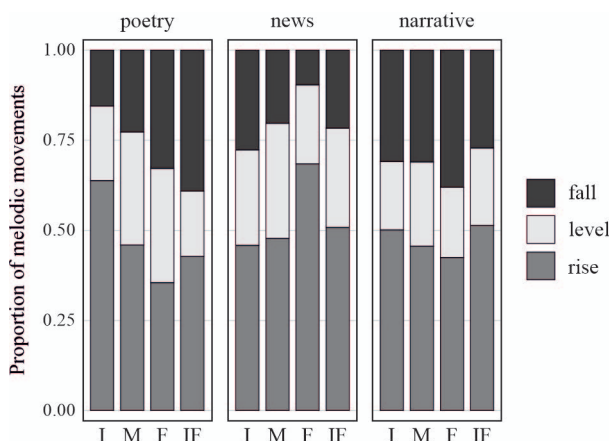


Figure 4: Relative occurrences of rises, falls and level movements between stressed and post-stress syllables by genre and position (initial, medial, final, ini-final) within a prosodic phrase.

First, it is obvious that a post-stress rise is more frequent than a falling or level melodic movement across all three genres and SG positions. However, the mutual proportions of these three possibilities are not uniform. Initial SGs in poetry and final SGs in news reading seem to be the post-stress rising champions. On the other hand, medial SGs in all three genres and final SGs in poetry reciting exhibit the lowest occurrence of post-stress rises, even if still higher than the occurrence of the other two options.

## 4. Discussion

Our study succeeded in providing answers to the research questions in Section 1.2. As to the structural properties, we found differences among the three genres in the mutual ratios of two-, three-, four-, and five syllable SGs. Also, the mean length of a SG according to the position in a phrase was not uniform across the genres. The initial and ini-final SGs are on average longer than medial and final ones. A tentative explanation could be that this is caused by anacrusis. They can occur only after a prosodic break, hence in an initial or ini-final SGs. (We count an anacrusis into the SGs to which it inclines. Leaving it out as some sort of residual material would not be suitable for Czech with its lexically fixed stress on the first syllable.)

The mean steps between the stressed and post-stress syllables, as well as mean post-stress rises (i.e., after exclusion of falls and levels) were found significantly higher in initial positions than in medial positions. That is not particularly surprising since the need to mark beginnings of prosodic phrases is known from other languages. Yet, the quantification of rises was not previously sought in three different genres (Table 2). The mean height of 2 ST is consistent with the occasionally reported narrower pitch range of Czech intonation. A noteworthy detail is also the unusual behaviour of the final post-stress rise in news reading (middle panel of Figure 3), which we tentatively explain as a feature peculiar just to this specific genre.

We also argue that the arithmetic mean in intonology could be deceptive to some extent. As it is generally accepted that a rise is more demanding than fall to produce, it is possible that despite a negative mean, there are more rises in the spoken texts, even if the melodic steps upwards are smaller. The number of rises relative to falls would then be more ‘phonological’ than the mean step across all movements.

Be that as it may, the numbers of perceptually relevant rises in our material were always higher than the numbers of falls or levels. This is a fact that was indicated in the past but never entered prominent textbooks or manuals, perhaps because the suggestions were based on less extensive material or material that was only machine-processed, without human corrections. We believe that our study will inspire future research in the prosodic role of various forms of post-stress rises and their communicative functions.

## 5. Acknowledgements

This work was funded by Czech Science Foundation (GA-CR), project GA21-14758S. The second author was also 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).

## 6. References

- [1] J. B. Pierrehumbert, *The Phonology and Phonetics of English Intonation*. Cambridge: Massachusetts Institute of Technology, MA, 1980. (Ph.D. thesis)
- [2] M. Baltazani, E. Kainada, K. Nicolaidis, and A. Lengeris, “The prenuclear field matters: Questions and statements in standard modern Greek,” in *Proceedings of the 18th ICPhS*, Glasgow, 2015, paper 941.
- [3] A. Arvaniti, and M. Baltazani, “Intonational analysis and prosodic annotation of Greek spoken corpora,” in *Prosodic Typology: The Phonology of Intonation and Phrasing*, S.-A. Jun, Ed., Oxford, Oxford University Press, 2005, pp. 84–117.
- [4] M. Baltazani, “Effects of stress on intonational structure in Greek,” in *Proceedings of Speech Prosody 2006*, Dresden, Germany, 2006, paper 156.
- [5] K. van Leyden, and V. J. van Heuven, “On the prosody of Orkney and Shetland dialects,” *Phonetica*, 63, pp. 149–174, 2006.
- [6] J. R. Walter, “On the intonation of a South Wales ‘Valleys accent’ of English,” *Journal of the International Phonetic Association*, 33/2, pp. 211–238, 2003.
- [7] J. Michnowicz, and H. Barnes, “A sociolinguistic analysis of pre-nuclear peak alignment in Yucatan Spanish,” in *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, C. Howe et al., Eds., Somerville, MA: Cascadilla Proceedings Project, 2013, pp. 221–235.
- [8] E. Martínez-Celdrán, A. M. Fernández-Planas, and N. Fullana-Rivera, “Pre-nuclear tonal inventories of Spanish intonation,” in *Proceedings of the 15th ICPhS*, Barcelona, Spain, 2003, pp. 595–598.
- [9] M. Ortega-Llebaria, and P. Prieto “Acoustic correlates of stress in Central Catalan and Castilian Spanish,” *Language and Speech* 54/1, pp. 73–97, 2011.
- [10] J. A. de Moraes, “F0 declination in Brazilian Portuguese in read and spontaneous speech,” in *Proceedings of 14th ICPhS*, San Francisco, USA, 1999, pp.2323-2326.
- [11] N. Thorsen, “Two issues in the prosody of Standard Danish,” in *Prosody: Models and Measurements*, A. Cutler, and D. R. Ladd, Eds., Berlin: Springer Verlag, 1983, pp. 27–38.
- [12] P. Janota, and Z. Palková, “The auditory evaluation of stress under the influence of context,” in *Acta Universitatis Carolinae – Philologica: Phonetica Pragensia*, 4, M. Romportl, and P. Janota, Eds., Praha: Universita Karlova, 1974, pp. 29–59.
- [13] Z. Palková, and J. Volín, “The role of F0 contours in determining foot boundaries in Czech”, in *Proceedings of the 15th ICPhS*, Barcelona, Spain, 2003, pp. 1783–1786.
- [14] J. Volín, “Z intonace čtených zpravodajství: výška první slabiky v taktu,” *Čeština doma a ve světě*, 2008/1–2, pp. 89–96. 2008.
- [15] M. Jůzová, and J. Volín, “F0 post-stress rise trends consideration in unit selection TTS,” in *Proceedings of the 21st Int. Conf. on Text, Speech and Dialogue*, P. Sojka, A. Horák, I. Kopeček, and K. Pala, Eds., Cham: Springer Int. Publishing, 2018, pp. 360–368.
- [16] R. Skarnitzl, and A. Eriksson, “The acoustics of word stress in Czech as a function of speaking style,” in *Proceedings of Interspeech 2017*, Stockholm, Sweden, 2017, pp. 3221–3225.
- [17] R. Skarnitzl, “Fonetická realizace slovního přízvuku u delších slov v češtině,” *Slovo a slovesnost*, 79, pp. 199–216, 2018.
- [18] T. L. Face, “Intonation in Spanish declaratives: Differences between lab speech and spontaneous speech,” *Catalan Journal of Linguistics*, 2, pp. 115–131, 2003.
- [19] M. E. Beckman, and G. Ayers Elam, *Guidelines for ToBI Labelling, version 3*. Ohio State University: The Ohio State University Research Foundation, 1997.
- [20] P. Boersma, and D. Weenink, *Praat: doing phonetics by computer* [Computer program] version 6.0.47. Retrieved Feb. 8, 2019, <http://www.praat.org/>, 2019.
- [21] D. J. Hermes, “Stylization of pitch contours”, in *Methods in empirical prosody research*, S. Sudhoff et al., Eds., Berlin: Walter de Gruyter, 2006, pp. 29–61.