

Mosey, Bryan; Chamonikolasová, Jana

Nucleus position and tone unit length in English and Czech

Brno studies in English. 1996, vol. 22, iss. 1, pp. [15]-21

ISBN 80-210-1421-0

ISSN 1211-1791

Stable URL (handle): <https://hdl.handle.net/11222.digilib/104304>

Access Date: 30. 11. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

BRYAN MOSEY & JANA CHAMONIKOLASOVÁ

NUCLEUS POSITION AND TONE UNIT LENGTH IN ENGLISH AND CZECH

This paper presents the results of a corpus-based, contrastive study of English and Czech spoken texts. It looks at the possibilities of carrying out a contrastive study of the prosodic features of two languages with different traditions in prosodic transcription. It examines two features of prosody – the position of the nucleus in a tone unit and the length of a tone unit.

Prosodic features analyzed

Tone unit

The *tone unit* is the basic prosodic unit in our analysis. The expression is used in Svartvik 1990 and Svartvik and Quirk 1979. Other authors writing in English refer to the the basic prosodic unit as the *tone group*, *intonation group*, *sense-group*, *breath-group*, *phonological phrase*, or *phonological clause*. Czech authors use the expression *clause segment* or *speech segment* (Palková 1990, Daneš et al. 1987).

Nucleus

The *nucleus*, as defined by Crystal 1969, is the most prominent accented stress in a tone unit. The corresponding expressions in Czech terminology are *clause accent* and *cadence*. (Palková 1990, Daneš et al. 1987).

Choice of material

We have selected the material for our analysis from the London–Lund Corpus of Spoken English (LLC), compiled at the English Department of Lund University (i.e. the computerised version of the material from the Survey of English Usage at University College London) and the Corpus of Spoken Czech (CSC; Korpus mluvené češtiny) compiled at the Institute of the Czech National Corpus of Charles University in Prague. The CSC is a subcorpus of the Czech National Corpus and has been set up mainly as a database for a frequency dictionary of

spoken Czech. The corpus is grammatically and semantically tagged but it does not contain any prosodic transcription. We had to provide our own prosodic transcription of the Czech material, in which another Czech linguist experienced in prosodic analysis helped us.

We have used dialogue S.1.6 of the LLC and dialogue JP 122 of the CSC. The two dialogues are comparable as to the topic of conversation and the social background of the speakers. We left out 197 tone units (out of 1235) of dialogue S.1.6 in order to make the two texts comparable also in terms of length. The topics of both dialogues are related to university education and the speakers in each dialogue are one female and one male academic. The English dialogue was recorded in 1964, the Czech dialogue in the early 1990s. Both dialogues are non-scripted but while the English dialogue was recorded surreptitiously, the Czech dialogue is, like all the material in the CSC, non-surreptitious.

Problems of comparison

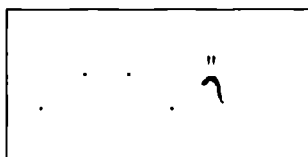
The main problem of our analysis was the lack of prosodic transcription in the Czech text and also the difference between the English and the Czech systems of prosodic notation. Below is an example of the prosodic transcription used in the LLC.

- 17,B, sta[^]tistics is what I :sh\ould know#
 18,B, ((and)) I [^]don't know 'anything a:b\out it#
 19,B, "[^]r\eally#
 22,B, *(([^]th\at`s what /I do#))
 23,A, *[^]y\es#
 24,A, do* [^]you know 'Malcolm B\owen#
 25,A, [^]over at the comp\uter /unit#
 26,B, [^][m]#
 27,A, [^]nice b/oy#
 28,A, [^]sure !he`d h/elp you#
 29,A, if you [^]got st\uck#
 31,A, [^]I !I "[^]I've been a :{fr\iend of} :{M\alcolm`s} :m\other#
 32,A, for "[^]d\onkey`s *'years#*
 33,B, *[^]h\ave you#*
 34,A, [^]oh I 'knew :M\alcolm#
 35,A, [^]when he was in kn\icker'bockers#
 37,B, *[^]y\es#
 38,A, [^]oh *y\es#
 39,B, *that`s [^]\interesting#
 40,B, [^]how \old is 'he#
 41,B, cos [^]I _found this _very _difficult to !g\uess#
 42,B, on [^]\looking *at him#

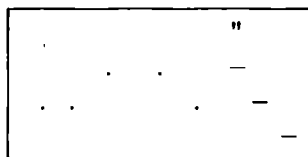
The text is divided into tone units, tone unit boundaries being marked by a slash. Different prosodic marks are used to denote different types of accented stress, unaccented stress, and boosters. We were looking for the most prominent accented stress in a tone unit, i.e. the nucleus. In tone units containing a compound nucleus, the most prominent accented stress – according to Crystal 1969 – is the last fall. In the above text, words containing the nucleus are underlined. We left out all tone units containing a subordinate tone unit in final position because we did not work with the concept of a subordinate tone unit in the Czech text and including data derived from subordinate tone units in the English text would have distorted the statistics.

Czech literature on prosody and intonation speaks about different melodies of speech and relates the pitch pattern of different melodies to different types of clauses (indicative clauses, questions, emotively coloured clauses, etc.). The pitch pattern of a clause is indicated graphically in the following way.

Na červenou stát!



V této místnosti nekuřte!



The dots and dashes indicate syllables and the syllable carrying what is in Czech terminology referred to as clause accent or cadence is marked with quotation marks. Although this prosodic notation does not seem to be very similar to the system used in the LLC, the prosodic theory behind it is compatible with the theory behind the English system. It is based on the recognition of clause sections whose definition corresponds to the definition of tone units. The clause accent or cadence is the most prominent accent in a clause section and that corresponds to the definition of the nucleus (Crystal 1969). We were therefore able to apply the same method of analysis to both texts and compare the results. We made a simple prosodic transcription of the Czech text. We removed punctuation from the text version and then listened to the recording of the Czech dialogue and determined tone unit boundaries and nuclei. We adapted the format of the Czech text so it was comparable to the English text.

43,B, asi máš pravdu#

44,B, něco se s tím asi dělat bude mušet#

45,B, a @m ale co se týče mě#

46,B, tak já chodím tady na práva#

47,B, na vobědy#

48,B, vaří tam celkem slušně#

49,A, a co příbor#

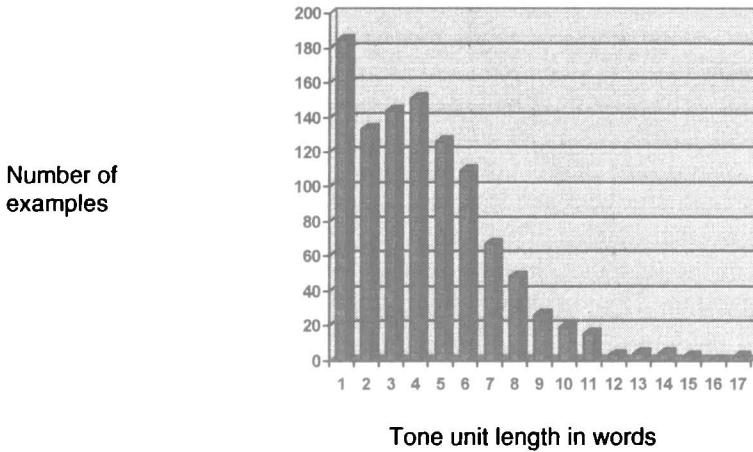
- 50,A, máte ten hliníkovej#
 51,A, takovej ten vošklivej#
 52,A, ten když to cvrkně vo plombu#
 53,A, tak to zajiskří#
 54,B, nó tak to myslim že ani nic takovýho#
 55,B, nic takovýho sem se tam#
 56,B, s ničim takovým sem se tam nesetka#
 57,A, nemáte hliníkovej#
 58,A, a nosíš si svůj příbor#
 59,A, nenosíš#
 60,B, to né no#
 61,B, myslim že se tomu dá celkem důvěřovat#
 62,B, tam je to teda v takový velký míse všechno#
 63,B, ale– já to tak prohlížim namátkou#
 64,B, a zdá se mi to že to je poměrně čistý#
 65,A, hm#
 66,B, takže tomu dost důvěřuju#
 67,B, ale hlavně jako se mi tam líbí docela ta kvalita jídel#
 68,B, i když–#
 69,A, jo#
 70,A, a maso#

Findings

Tone unit length

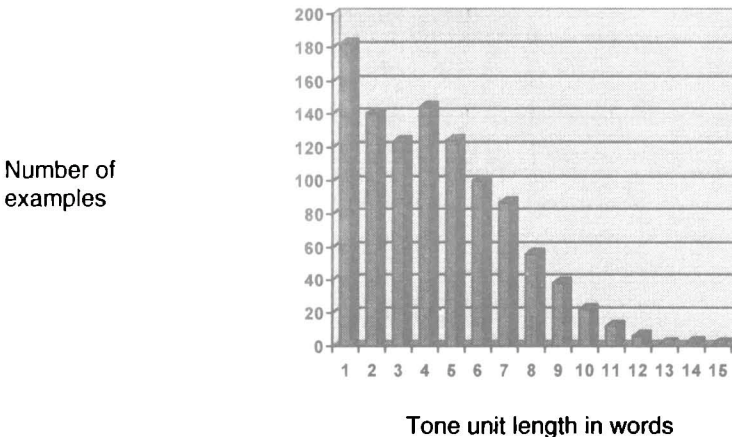
The English text had an average tone unit length of 4.26 words which corresponds closely to the figure of 4.34 words Bryan Mosey found when studying the so-called minor tone units of the Lancaster IBM Spoken English Corpus (Mosey 1994) and to that found by Altenberg in his study of LLC dialogue S.12.6 (Altenberg 1987). This further confirms the general homogeneity of tone unit length across genres, these earlier studies being based on monologue texts while the current study uses a dialogue.

A noticeable difference, however, is the distribution about the average. In the monologue texts of the Spoken English Corpus, tone units of only one word in length were rare and fell clearly outside the standard deviation range. In the current text, the standard deviation about the average is greater by almost half a word, i.e. 2.73. Here, one-word tone units form the largest single group as can be seen in the chart below.

Figure 1: Variation in tone unit length – English text

Although these fall outside the normal range, this is only by a very small margin. This suggests somewhat greater variation in tone length in dialogues than in monologues and a substantially higher percentage of one word tone units, which is perhaps not surprising considering the occurrence of back-channel items and the like in dialogue.

In the case of the Czech text, the average tone unit length was very slightly higher, 4.34 words, and the standard deviation was slightly lower, 2.69. Again, one word tone units fell just outside the standard deviation range but again, they did form a fairly large group (there were 185 one word tone units in the English text and in the Czech text there were 182).

Figure 2: Variation in tone unit length – Czech text

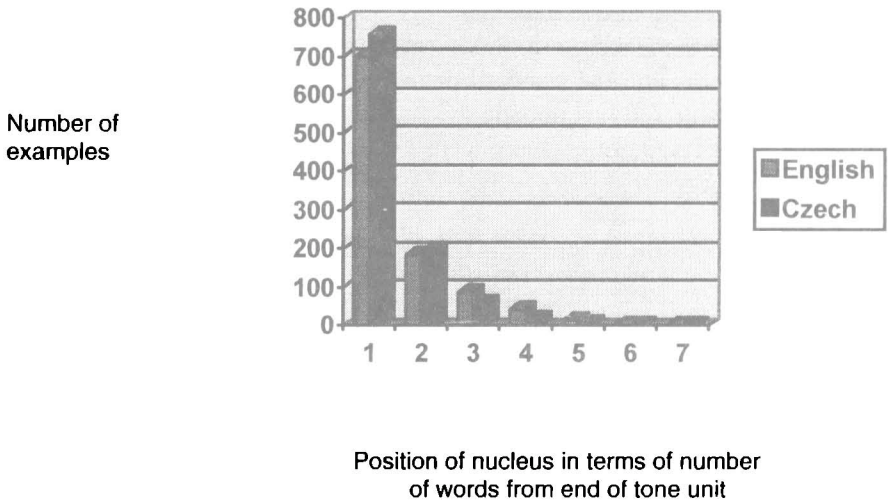
Unfortunately, corresponding data on Czech monologues are not yet available.

Nucleus position

While the patterns regarding tone unit length seemed to differ little between the English and Czech texts (lending credence to the notion of tone units as phonological realisations of units of information chunks of a convenient size for processing by both speakers and listeners), nucleus position did vary more noticeably. The average nucleus position in the English text was 1.6 words back from the end of the tone unit. In the Czech text, it was 1.4 words back from the end of the tone unit. This difference may sound small, but bearing in mind the propensity for nuclei to occur in the very last word of a tone unit, it does have some significance. The standard deviation was 0.96 in the English text and 0.98 in the Czech text.

Tone units in Czech dialogue are, thus, even more likely than their English equivalents to have the nucleus on the last word. This is borne out if we look at the actual figures for the two texts studied here.

Figure 3: Nucleus position in English and Czech texts



SUMMARY

The present article is a study of two prosodic features, tone unit length and nucleus position. It is based on an analysis of spoken texts selected from the London–Lund Corpus and the Corpus of Spoken Czech. The study supports the definition of tone units as the phonological realisation of information chunks of a convenient size for processing by both speakers and listeners; the average length of tone units was found to be around 4.3 words with a standard deviation of about 2.7 in

both the English and Czech texts. The examined texts, both English and Czech, contained a high percentage (almost 18 %) of one-word tone units. The frequent occurrence of one-word tone units seems to be typical of unprepared conversation. Monologues, as other studies have shown, contain a much lower percentage of one-word tone units. The study suggests that both English and Czech speakers have a strong tendency to place the nucleus towards the end of a tone unit. The tendency seems to be stronger in Czech, where the average nucleus position was 1.4 words from the end of a tone unit, than in English, where the average was 1.6, standard deviation being about 1.0 in both English and Czech.

Abbreviations

LLC London–Lund Corpus
CSC Corpus of Spoken Czech

References

- Altenberg, B. (1987) *Prosodic Patterns in Spoken English* (Lund: Lund University Press).
 Couper–Kuhlen, E. (1986) *An Introduction to English Prosody* (Tübingen: Niemeyer).
 Cruttenden, A. (1986) *Intonation* (Cambridge: CUP).
 Crystal, D. (1969). *Prosodic Systems and Intonation in English* (Cambridge: CUP).
 Daneš, F. (1957) *Intonace a věta ve spisovné češtině* (Prague: Academia).
 Daneš et al. (1987) *Mluvnice češtiny 3* (Prague: Academia).
 Firbas, J. (1992) *FSP in written and spoken communication* (Cambridge: CUP).
 Gimson, A.C. (1970) *An Introduction to the pronunciation of English* (London).
 Gimson & Cruttenden (1994) *Pronunciation of English* (Cambridge: Edward Arnold).
 Mosey, B. (1994) A study of nuclei in the SEC – report from the Public Speaking Project in Lund. *Hermes journal of linguistics 13*: 123–134.
 Palková, Z. (1994) *Fonetika a fonologie* (Prague: Academia).
 Svartvik, J. (1990). *London–Lund Corpus of spoken English – description and research* (Lund: Lund University Press).
 Svartvik, J., Quirk R. (1979). *A Corpus of English Conversation* (Lund: Lund University Press).
 Quirk, R., Greenbaum, S., Leech, G., Svartvik, J. (1985). *A comprehensive grammar of the English language* (New York).

Note

Some of the materials used in this analysis were procured using funding from the Grant Agency of the Czech Republic.

