The Adaptation of the ToBI System for Lithuanian Pitch Change

Neringa Krapikaitė

Vytautos Magnus University, Department of the Lithuanian Language, Str. K. Donelaičio 58, LT-44248 Kaunas, neringann@yahoo.com

Abstract. Prosody (intonation) is one of the most complicated phenomena of a spoken language. The most convenient tool for this type of research can be at least minimally annotated speech corpus and properly prepared speech labeling system based on common methodology widely used by foreign researchers. One of the most popular standards for annotation and transcription of prosodic events is ToBI (an acronym for Tones and Break Indices). The transcription system of an utterance consists minimally of a recording of the speech, its fundamental frequency contour, and (in the transcription proper) symbolic labels for prosodic events. The transcription proper is usually arranged in four time-aligned parallel horizontal tiers: the Tone tier, the Orthographic tier, the Break-Index tier and the Miscellaneous tier, for recording additional observations. The main aim of this paper is the following: to outline the use of ToBI transcription system for the transcription of Lithuanian intonation and to reveal the future research in the area of prosody. Therefore the main task of this paper was to make a wide – range investigation of ToBi transcription system in order to present the process of creating transcription system for intonation in Lithuanian. It was investigated, that ToBI transcription proper is applicable for the prosodic annotation of Lithuanian intonation and is the most convenient tool for prosody research of this type.

Keywords: spoken language, ToBI, intonation, transcription, pitch accent.
Introduction

The language can be expressed in both the sounds and graphics. Spoken language is different from the written text in many linguistic respects. There is a tendency in spoken language to avoid those elements of speech, which can be compensated in a general situation. Secondly, speaking involves various non-verbal communication tools (body expression), syllable omission, shortened words, incomplete sentences and ellipsis. Moreover, the abundance in words, showing the excessive usage of certain lexical items is also the characteristic of spoken language. Spoken language is usually spontaneous. Speaking takes place in real time and place and speakers do not have the opportunity to edit their language. For this reason, there always occur the slips of the tongue, errors, corrections, repetitions and filled pauses. The syntax of spontaneous spoken language is in general fragmented. The utterances are short and simple.

Prosody (intonation) is one of the most complicated phenomena of a spoken language. It comprises the segmentation of speech into meaningful units, emphasis of key words, fluctuation in speech tempo, expression of emotions, etc. These features vary in each language depending on its linguistic (language standards) and extralinguistic distinctions (cultural, historical dimensions).

As an important element in language, intonation is correlated with other segmental and suprasegmental linguistic means. The influence between these elements and intonation is mutual. Pitch, loudness and the duration are the most significant features of intonation which organizes speech phonetically, breaking it down according to its meaning into sentences and significant segments or syntagms. Intonation is also related with other important acoustic characteristics, such as words stress, syllable structure, rhythm (the distribution of accented syllables, the duration of a syllable). The most significant features of intonation can also be the characteristics of other phonetic elements (such as word stress). Intonation also exists as an independent system consisting of the changing of the intonation centre position, syntagmatic segmentation and pausing. These particular elements of the system also have their own acoustic characteristics which contribute to the variety of the utterances.

As we can see, intonation is one of the most complicated phenomena of a spoken language. The complexity of prosodic features poses many problems for prosody researchers. The most convenient tool for this type of research can be at least minimally annotated speech corpus and properly prepared speech labeling system based on common methodology widely used by foreign researchers. One of the most popular standards for annotation and transcription of prosodic events is ToBI. The main aim of this paper is the following: to outline the use of ToBI transcription system for the transcription of Lithuanian intonation and to reveal the future research in the area of prosody. Therefore the main task of this paper was to make a wide – range investigation of ToBI transcription system in order to present the process of creating transcription system for intonation in Lithuanian.
1. The Lithuanian Prosodic Corpus

There is not any Lithuanian prosodic corpus available yet, though The Experimental Corpus of the Lithuanian Speech (ECLS), compiled by Vytautas Magnus university (VMU) in 2008, could be thought of as a kind of predecessor. Although this database is very small (10 hrs.), this is the first annotated freely available online Lithuanian speech corpus. Though the ECLS corpus has just phoneme level annotations (the features of intonation are not reflected in textual annotations), the experience obtained while creating it, is used in the process of building The Lithuanian Prosodic Corpus at VMU. The project is sponsored by the Research Council of Lithuania.

Intonation corpus can be thought of as a speech corpus that has its orthographic / phonetic annotations complemented with prosodic labels. Prosodic labels may characterize both prosodic phenomena themselves (e.g. pitch accents and boundary tones) and features that affect prosody (e.g. logical stress).

The process of compiling speech corpus can be divided into a few steps: a) recording of prosody-rich utterances of spoken language b) description of the content of these utterances and utterance markup with tags that describe prosodic features c) automatically assigning timings to prosodic features (as a result of phone level annotation of intonational speech corpus). Every step in this process requires certain procedures to be observed and certain requirements to be met (Kazlauskienė, Raškinis, 2013, 1).

The corpus is being built according to the conventions (for transcribing and annotating the intonation) and common methodology widely used by foreign researchers (known as ToBI (Tones and Break Indices)) transcription system (Pierrehnmbert, Hirschherg, 1990, 271).

2. Tones and Breaks Indices

The term ToBI is used to denote the transcription system to record the intonation in individual languages. It was developed in a series of meetings by representatives of various disciplines (engineers, psychologists, computer experts and phoneticians) who wanted to create tools to work together and common terminology that researchers from different fields could use it for the interpretation of data and contribution for further analysis (Volk, 2012, 169).

The ToBI transcription system is a method for transcribing two distinctive aspects of the prosody of spoken utterances:

a) accents, which contribute to a word’s relative prominence in an utterance and
b) phrasing, which creates a grouping of words.

The transcription system of an utterance consists minimally of a recording of the speech, its fundamental frequency contour, and (in the transcription proper) symbolic
labels for prosodic events. The transcription proper is usually arranged in four time-aligned parallel horizontal tiers:

1. the Tone tier, for transcribing tonal events;
2. the Orthographic tier, for transcribing words;
3. the Break-Index tier, for transcribing boundaries between words;
4. the Miscellaneous tier, for recording additional observations (Hirschberg, Beckmann, 2012).

2.1. The Tone tier

The tone tier is that part of the transcription which corresponds mostly to the phonological analysis of the intonational pattern of each utterance. It consists of labels for distinguishing pitch accent tones which are transcribed as a sequence of high (H) and low (L) tones. A tone may represent a ‘peak accent’ (H*), ‘low accent’ (L*) or a part of it (L*+H and L+H*).

H* means a “peak accent”. It indicates an apparent tone target on the accented syllable which is in the upper part of the speaker’s pitch range for the phrase. This label is also being used for tones in the middle of the pitch range but excludes a very low level of F0 (Pierrehnmbert, Hirschberg, 1990, 275).

![Figure 1. Paskaita baigėsi! (angl. The lesson has finished!,) (sukurta mano – aut. past.)](image)

This example has 2 words and is produced as a single intonational phrase with one pitch accent. The most prominent syllable in this utterance is the syllable (bai-) of the word baigėsi. The f0 contour shows a rise that peaks at the end of bai- and then falls. This is an example of a canonical H* or High pitch accent. In the ToBI transcription, the H stands for a High tonal target and the * indicates that the H tone is associated with an accented syllable (Hirschberg, Beckmann, 2012).
It is important to keep in mind, that the changes in pitch in “peak accented“ (or H*) syllables are strongly related to the syllable structure. For example, syllable initial stops [b], [p] and [t], fricatives [f], [s] and voiced [d], [g] usually disturb the pitch and as a result strongly influence the fundamental frequency pattern. Moreover, it is not possible usually to read the intended pitch during a voiceless consonant by interpolating from the last f0 value before voice offset to the first f0 after voice onset because obstruent consonants (stops, fricatives, affricates) all cause dramatic perturbations in the fundamental frequency contour over and above any interruption of voicelessness per se. Nevertheless, ToBI Annotation Conventions prescribe H* even where it is not possible to read the intended pitch (Beckman, Elam, 1997, 13).

L* accent means low tone target in the accented syllable which is in the lowest part of the speaker’s pitch range (Hirschberg, Beckmann, 2012).

The following example has a falling f0 that reaches a minimum on the bai- syllable in the word baigėsi. This is a Low pitch accent on baigėsi, marked with L*. Again, the L stands for a low tonal target and the * means that the low tone is associated with a prominent syllable.

‘Rising peak accents’ (L+H*) are also common in Lithuanian. This kind of accent means a high tone target on the accented syllable which is immediately preceded by relatively sharp rise from a valley in the lowest part of the speaker’s pitch range. (Pierrehnmbert, Hirschherg, 1990, 275).
In the following example, there is a stretch of low f0 followed by a rise to a high f0 peak, then a fall to another low stretch. It is important to look closely at how the pitch track aligns with the syllables of the word *paskaita*. As it can be seen in figure 3, the H tone aligns with the prominent syllable `-ta-`. The f0 rises through the prominent syllable, with the peak occurring at the end of the syllable. This is a ‘Rising peak accent’ on *paskaita*, marked with L+H*. Again, the H* stands for a “peak accent” or a high tone target on the accented syllable which is immediately preceded by L or a relatively sharp rise from a valley in the lowest part of the speaker’s pitch range.

L*+H represents a low tone target on the accented syllable which is immediately followed by a sharp rise to peak in the upper part of the speaker’s pitch range (Pierrehumbert, Hirschberg, 1990, 275).

However the realization of ‘scooped accents’ (L*+H) is very poor in Lithuanian.
As it can be seen in figure 4, the tone aligning with the prominent syllable -rė- is Low. While there is a peak corresponding to the High tone, this High tone occurs at the non-prominent syllable –jo, which is at the end of the word nenorėjo. The alignment of the L tone with the prominent syllable shows that the pitch accent is the “starred” one in this bitonal pitch accent and the H corresponds to the perception that this low tone is immediately followed by a sharp rise to peak in the upper part of the speaker’s pitch range. This is an example of a canonical L*+H or ‘scooped’ pitch accent.

2.2. More on the ToBI Labelling Tiers

The Orthographic tier contains a transcription of all of the words in the utterance, in ordinary Lithuanian orthography.

The Break index tier represents the strength of a boundary between two orthographic words. Symbols / and // indicate intermediate and full intonation phrase boundary respectively. The symbol _ indicates the absence of a word boundary.

The Miscellaneous tier is essentially a ‘comment’ tier that can be used to mark events such as breaths, coughs, laughter, long silences and other nonspeech events. These are traditionally marked with angle brackets (e.g. <cough>). Despite the fact, that these events are arguably not a part of prosody, they are important for interpreting the analyses on the Tone tier and Break-Index tier, because they disrupt the smooth rhythm of the utterance or interrupt the intonation contour.

The Lithuanian Prosodic Corpus differs from other languages in the completeness of textual annotations. Some common methodologies have been extended in order to create complete textual annotations which indicate significant linguistic and paralinguistic information, such as intermediate or full intonation phrase boundaries, breaks, phrasal tones, theme and rheme, emotion, noise, etc.

Conclusions

The ToBI transcription proper is the most convenient tool for prosody research. The paper presented the use of ToBI transcription in The Lithuanian Prosodic Corpus which is being built in Vytautas Magnus University. It outlined the application of transcription system of prosodic events in automated analysis of large speech corpora with the aim of investigating whether this transcription system is applicable for the prosodic annotation of Lithuanian. The results of the analysis show, that the intonational contour can be transcribed separately in individual tiers representing independent structural types. The use of the adapted ToBI system for Lithuanian therefore gives the possibility of distinguishing intonational contours which share the same phonological structure but different types of pitch accents. The described transcription method show, that the symbolic representation...
of prosodic elements corresponds most to the phonological analyses of the intonational pattern: symbolic labels may represent a pitch accent (H* or L*) or a part of it (H+L* or L+H*). However, further research on the intonation of Lithuanian is needed to clarify the boundaries between different sounds.

References


paplitusių intonacijos anotavimo ir trankribavimo sistemų – ToBI transkripcijos sistema. Pilną ToBI transkripcijos sistemą sudaro tam tikrų rūšių lygmenys (angl. Tiers), išdėstyti skirtinguose laiko ir erdvės požiūriu pasakymo garso taškuose: ortografijos, pauzių indekso, intonacijos ir pastabų. Pagrindinis šio straipsnio tikslas (ir uždavinys) – atskleisti ToBI transkripcijos sistemos pritaikomumą lietuvių kalboje (jos intonacijoje) ir aptarti tolesnių intonacijos tyrimų perspektyvas. Nustatyta, kad ToBI transkripcijos sistema gali būti pritaikyta lietuvių kalbos intonacijos požymiams žymėti. Ši transkripcijos sistema ne tik palengvina bendro pobūdžio lietuvių kalbos prosodijos tyrimus, bet ir sudaro sąlygas lyginti pačios fonologinės struktūros pasakymų požymius.

**Esminiai žodžiai:** sakytinė kalba, ToBI, intonacija, transkripcija, pagrindinio tono kirtis.

Gauta / Received 2015-01-06
Priimta / Accepted 2015-03-02