Abstract

It is the main motivation of our present study to identify some further criterion for the oppositions of topic, contrastive topic and, as the case may be, of focus in the prosodic characteristics of utterances. For this purpose, we have examined three sets of naturally occurring speech and analyzed them in order to find whether the character of intonation contours distinguishes between the three notions.

0. Introduction

The main motivation for our study is the effort to find out, how the investigation of the intonation contour of Czech sentence and of other characteristics of spoken language may help to look for criteria relevant for distinguishing some of the TFA values (Topic-Focus Articulation, or, in other words, information structure) in the annotation framework of the Prague Dependency Treebank (PDT [1]). In addition to this primary aim, a comparison of this type may lead to some other interesting observations, which may uncover some important features of signalization of meaning by means of suprasegmental means of expression in spoken language in general.

1. The Annotation Scenario of the Prague Dependency Treebank

1.1. The Prague Dependency Treebank

The Prague Dependency Treebank (PDT in the sequel) is an annotated corpus of Czech texts (the texts are taken from the Czech National Corpus, the first release of which contains a hundred million word occurrences in journalistic fiction and other texts). At present, its scenario contains three layers of annotation: the morpho-syntactic layer (about 1100 tags are in actual use), the layer of dependency structures representing the surface shape of sentences (the so-called analytic layer) and the underlying syntactic layer (so-called tectogrammatical tree structures, TGTSs).

1.2. The tectogrammatical layer of annotation

The TGTSs are based on the theoretical framework of the functional generative description (FGD) proposed by Petr Sgall in the sixties and developed further by him and his followers until the present time (see esp. Sgall et al. 1986 [2]). The tectogrammatical level of FGD is distinguished from other theoretical descriptive frameworks esp. by the following two features: (i) it is dependency based, with coordination as a 'third' dimension of the trees, and (ii) the information structure of the sentence, its topic-focus articulation (TFA), is claimed to be an integral part of the tectogrammatical level, i.e. a distinction belonging to the underlying level of linguistic description.

1.3. Topic-focus articulation in a theoretical description

The articulation of the sentence into T(opic) and F(ocus) is based on what is from a cognitive point of view understood as the "given-new" strategy; the semantic basis of this articulation is the relation of aboutness (for a formal treatment, see Peregrin 1994 [3]): a prototypical declarative sentence asserts that its Focus holds (or does not hold) about its Topic: F(T) or non-F(T). A detailed empirical study of more complex sentences has led us to an introduction of the notion of contrastive topic, i.e. a contextually bound item that stands in contrast. Hajičová, Partee and Sgall (1998 [4], p. 151) introduce the notion of contrastive (part of) topic in connection with the occurrences of the so-called focusing particles (focalizers) in topic.

2. Spoken data

2.1. Description

In our research we used three sets of annotated spoken data from a spoken corpus compiled in the Institute of Formal and Applied Linguistics, Charles University, Prague. These three sets are recordings of three talkshows of Czech TV called Na plovárné (On a Bathing Place), moderated by a well-known and highly intelligent moderator Marek Eben; in these shows, he interviews three personalities of Czech cultural and sport life, namely Vladimír Komárek, a painter, Magdalena Kožená, a concert and opera singer, and Petr Jírmus, a pilot of an acrobatic plane.

The dialogues were recorded with 22kHz sample frequency and with 16-bit resolution. We generated for each dialogue the smoothed F0-contours (fundamental frequency
contours) with the help of the Edinburgh Speech Tools (EST 2000 [5]) software. The F0-contour extraction was done with the time step 0.01 sec and for the 40Hz–400Hz frequency range.

2.2. Segmentation
The dialogues were transcribed manually with the help of the Transcriber program (Transcriber 2002 [6]). The following segmentation of the text has been performed:

Phonemes - segmentation into phonemes and words was carried out automatically by the HTK software (HTK 1999 [7]).

Segments - segmentation into speech segments has been done manually. The transcribers inserted synchronization marks at those places where the end of the sentence was followed by a pause. These synchronization marks define the speech segments.

Turns - segmentation into turns is carried out by hand.

3. Processing and analyzing the data
3.1. Modifications of the segmentation
As already mentioned, we had to make some additions and modifications of the scheme described in Section 2.2.

The word boundaries remain as characterized above, because their definition is motivated phonologically. This has influenced also our characterization of utterance events, see below.

For the reasons mentioned above, we had to modify the specification of the boundaries of speech segments for our purposes. In addition to the phonetic criterion the labelers added marks for the boundaries of utterances: the delimitation of an utterance is crucial for our analysis, because it serves as the basic unit for the determination of TFA.

3.2. Sectors of an utterance
3.2.1. General description
The segmentation of utterances into sectors is guided by structural and semantically relevant criteria rather than by the phonetic shape of these sectors. A sector is a node or a subtree of the tectogrammatical tree structure (which represents a complementation of the main verb or of some dependent node); this node/subtree may be rendered in the surface shape of the sentence by a single word, a prepositional or prepositionless nominal group, or by a whole clause in case of a complex (‘subordinate’) sentence.

3.2.2. Types of sectors
Three main types of sectors are differentiated in our material (see above in Section 1 for the theoretical prerequisite of such distinctions):

Non-contrastive topic: The category of topic can be paraphrased as ‘what is the sentence about’. From the TFA point of view, in unmarked cases the initial position of a Czech sentence may be occupied in two ways: by a contextually bound non-contrastive item OR by a contrastive topic. Our starting hypothesis was that these two cases should be distinguished in the intonation contours, so that if we succeed to distinguish in a sufficiently clear way the realizations of topic and contrastive topic, we can speak about a typical contrastive contour.

Contrastive topic: Since contrast is a rather unclear and oscillating category, contrastive topics have been determined by means of a number of criteria, the main being the (unival) contrastive character of their ‘linking’ in the text. In the first phase of our analysis, we have therefore tried to exclude vague or somewhat doubtful cases. The quantitative limit was to obtain at least 100 cases of contrastive topics, which was the lowest number of cases we wanted to have for the purpose of further analysis.

Focus: As a rule, focus is understood as a part of the sentence that is signaled by including the bearer of the intonation center and by its unmarked position at the end of the utterance, i.e. also by a characteristic final cadence. Most theories of TFA assume that every utterance must have a focus. Since one of the main tasks of our analysis was to attempt to demonstrate a difference in the contours of focus sectors and contrastive topics, we have singled out for our analysis of foci those utterances in which a contrastive topic was identified. This was a simple and at the same time a transparent criterion from the point of view of comparison of differences in the intonation contours of the two categories (i.e. both occurred in the same sentence).

Focus segments, however, may also occur in a marked position, i.e. either at the beginning or in the middle of an utterance. Such a marked order of elements was called by Mathesius (Mathesius 1947 [8], p. 241) a subjective order. An analysis of such utterances may help us to find out, whether an eventual difference between the intonation con-
tours of contrast and focus is not conditioned just by position in the utterance (initial vs. final) and whether the characteristics obtained are not typical just for a specific sentential position rather than for a TFA category.

3.2.3. A more detailed classification of the basic types

The category of focus has been further divided into two types: (i) the respective sector is the final sector of the given utterance (focus1) (ii) the sector is the focus of an utterance that stands in a coordination relation within a larger whole (focus2). The reason for such a subdivision was to make it possible to find out the influence of a final cadence or half-cadence on the examined intonation contour.

According to the strength of the contrast, the category of contrast in topic has been subdivided into three types; the decision of the assignment of one of these categories was based on the size of the set of alternatives from which the contrastive topic has been chosen, its means of expression and the way of the choice of the contrastive topic from this set (see Hajičová, Sgall, Veselá 2003 [9]).

Contrast 1: the strongest type of contrast, when an enumeration of elements of a set of alternatives continues, these elements are in the same semantic class and in an analogous position in the sentence.

Contrast 2: a weaker type of contrast, the contrastive topic is the first choice of an element of the set of alternatives and it carries information about the articulation of this set.

Contrast 3: the weakest type of contrast, the contrastive topic is selected from the 'hypertheme' of the discourse segment, and the main signal of contrastivity is the non-derivability of the contrastive element from the (immediately) preceding context.

4. Outputs of the analysis

4.1. Tables

We present first the results in a tabular form, followed (in Section 4.2) by the description of the factors measured.

Table 3:

<table>
<thead>
<tr>
<th>type of sector</th>
<th># of occur.</th>
<th>length (msec)</th>
<th>mean value</th>
<th>begin. (F0)</th>
<th>end (F0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>97</td>
<td>615.6</td>
<td>137.0</td>
<td>134.31</td>
<td>135.91</td>
</tr>
<tr>
<td>Contrast1</td>
<td>31</td>
<td>617.7</td>
<td>136.7</td>
<td>123.86</td>
<td>142.39</td>
</tr>
<tr>
<td>Contrast2</td>
<td>30</td>
<td>614.3</td>
<td>149.5</td>
<td>137.06</td>
<td>156.43</td>
</tr>
<tr>
<td>Contrast3</td>
<td>50</td>
<td>809.4</td>
<td>138.3</td>
<td>122.65</td>
<td>142.48</td>
</tr>
<tr>
<td>Focus1</td>
<td>32</td>
<td>837.8</td>
<td>126.1</td>
<td>129.92</td>
<td>115.50</td>
</tr>
<tr>
<td>Focus2</td>
<td>83</td>
<td>669.2</td>
<td>131.2</td>
<td>136.21</td>
<td>122.03</td>
</tr>
<tr>
<td>Sfocus</td>
<td>25</td>
<td>646.4</td>
<td>136.6</td>
<td>139.25</td>
<td>126.12</td>
</tr>
</tbody>
</table>

Table 4:

<table>
<thead>
<tr>
<th>type of sector</th>
<th>range (F0)</th>
<th>rise—fall (F0)</th>
<th>difference (F0)</th>
<th>dif. std. deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic</td>
<td>30.78</td>
<td>14.79</td>
<td>1.61</td>
<td>34.29</td>
</tr>
<tr>
<td>Contrast1</td>
<td>45.48</td>
<td>32.59</td>
<td>18.53</td>
<td>48.33</td>
</tr>
<tr>
<td>Contrast2</td>
<td>40.63</td>
<td>32.67</td>
<td>19.37</td>
<td>39.37</td>
</tr>
<tr>
<td>Contrast3</td>
<td>48.37</td>
<td>37.28</td>
<td>19.83</td>
<td>34.85</td>
</tr>
<tr>
<td>Focus1</td>
<td>49.49</td>
<td>30.82</td>
<td>-14.42</td>
<td>52.46</td>
</tr>
<tr>
<td>Focus2</td>
<td>44.57</td>
<td>30.20</td>
<td>-14.18</td>
<td>46.93</td>
</tr>
<tr>
<td>Sfocus</td>
<td>40.18</td>
<td>29.73</td>
<td>-13.13</td>
<td>37.57</td>
</tr>
</tbody>
</table>

4.2. Quantified parameters

Table 3 and Table 4 present the results of the phonetic analysis of our data, selected from the collections of authentic speech records described in Section 2 above. We summarize there the maximum of data considered potentially relevant from some point of view. All these values are taken from the F0 curve, because we take the intonation contour to be the most important factor of the intonation of Czech sentences. All F0 values presented here are calculated as mean values of all sectors of the given type. Both tables reflect aggregated data (across all the speakers).

The following values have been calculated:

- the number of occurrences of the given sector type,
- maximum - the maximal value of F0 measured on the given sector,
- minimum - the minimal value of F0 measured on the given sector,
- mean - the mean height of F0 on the given sector,
- beginnings of the sectors - the value of F0 of the initial point of the sector,
- ends of the sectors - the value of F0 of the final point of the sector,
- length - length of the sector in milliseconds,
- difference - the overall tendency of the sector - the difference between the initial and the final values of F0 of the given sector,
- range - the difference between the maximal and minimal value of F0 of every sector,
- rise — fall - the difference between the maximal and lesser of the initial and final F0 values of the given sector,
- standard deviation of difference - square root of difference variability.
4.3. Results

The column “difference” indicates quite clearly that the material can be subdivided into three groups: the sectors of the type topic occur at the point near to zero, the sectors of the focus type in the negative values and the sectors of the type contrastive topic in the positive values.

There are also no evident differences between the individual groups of contrastive topic. The resulting values rather indicate that the strength of contrast does not have an influence on the phonetic shape of the sentence.

5. Deviations

Few errors could have been introduced into our statistic by some wrong word boundary detection (usually due to a wrong boundary detection of plosive phones) or by some wrong F0-values extraction at voiced consonants and at speech segments boundaries where a small F0-contour flattening occurs by smoothing.

Quite understandably, the mean values of the measured values are influenced by the manner of speech used by individual speakers. In this domain, there are only relatively few phenomena that can be expressed in a quantitative way, out of which the following ones can be mentioned here:

- The mean height of the speaker’s voice.
- An overall ‘tendency’ of the utterance.
- The range of the utterance.

In addition to the height of voice and other quantified features, also purely individual tendencies can be observed, that is a specific way of realizations of the individual sectors.

Table 5:

<table>
<thead>
<tr>
<th>speaker</th>
<th>mean (F0)</th>
<th>range (msec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>speaker1</td>
<td>118.4</td>
<td>164.4</td>
</tr>
<tr>
<td>speaker2</td>
<td>99.5</td>
<td>136.6</td>
</tr>
<tr>
<td>speaker3</td>
<td>132.7</td>
<td>205.3</td>
</tr>
<tr>
<td>speaker4</td>
<td>159.4</td>
<td>214.6</td>
</tr>
</tbody>
</table>

6. Conclusion

On the basis of the results presented above we can attempt at a hierarchization of the measured parameters and their combinations according to their ability to help to distinguish between individual types of sectors. It is quite evident that the most important of the values is the difference between the initial and final value, which perspicuously differentiates the types of focus, contrast and topic.

We can thus say that the typical contour of contrast is a rising one, the typical contour of focus the falling one and that topic has a more or less constant contour.

7. Acknowledgement

The work reported on in this paper has been carried out under the project of the Czech Ministry of Education LN00A063.

8. References