Grammars for Sentence into Phrase Segmentation: Punctuation Level

Yuras Hetsevich(✉), Tatsiana Okrut, and Boris Lobanov

United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus
yury.hetsevich@gmail.com

Abstract. This paper deals with so-called punctuational phrases that make up sentences and with their marking according to intonation type in Belarusian electronic texts using NooJ. Such markings may be used for the implementation of an algorithm for intonationally-coloured text-to-speech synthesis in order to obtain expressive synthetic speech.

Keywords: Punctuational phrases · Intonation-type marking · Belarusian language · Expressive text-to-speech · NooJ

1 Introduction

To date, text to speech (TtS) systems have reached a certain level of development and are now used in a number of practical applications. But under real operational conditions, the way synthesized speech is perceived is not always fully satisfactory. As our verbal experiences show, depending on the degree of speech expressiveness, the same text read by two different speakers may have different effects on an audience. The global trends of speech technology development have created a real demand for expressive TtS systems.

Therefore, this paper aims to give an overview of the initial step in formal representation of the expressive speech phenomena using NooJ.

The concept of expressive speech was formed as an interdisciplinary concept characterizing one of the human speech functions [1–3]. The main components of oral speech expressiveness include speech prosody and, in particular in speech synthesis, correctness and quality of prosodic marking [4]. Within this work, prosodic marking involves the segmentation of sentences into punctuational phrases and the determination of their intonation type.

In speech synthesis, such markings make it possible to avoid frequent repetition of similar intonation constructions, which in turn promotes better intelligibility and naturalness of synthesized speech.

Apart from the application in TtS, prosodic marking may be used for educational purposes. The authors have also developed an algorithm aimed to visual prosodic marking for practical training of intonation.

T. Okrut et al. (Eds.): NooJ 2015, CCIS 607, pp. 74–82, 2016.
DOI: 10.1007/978-3-319-42471-2_7

2 Sentence and Phrase Segmentation

Speech synthesis is carried out sentence by sentence, with each sentence prosodically independent from another sentence in a text.

A sentence is regarded as a text passage separated by punctuation marks such as a full stop {.}, a question mark {?}, an exclamation mark {!}, an exclamation/question mark combination {!?}, several exclamation marks {!!!}. Ellipses {…} may also serve as an end-of-sentence punctuation mark, provided the word after the ellipsis starts with a capital letter.

Then each sentence is separated into punctuational phrases by using the following punctuation marks:

- semicolon {;},
- colon {:},
- comma {,},
- dash {–},
- left bracket {(},
- right bracket {)},
- combination of comma and dash {,–}.

Thus, if a sentence contains n punctuation marks (including the end-of-sentence mark), it is divided into n punctuated phrases (n = 1, 2, 3,…).

It is obvious that phrases may be different in length (where length is regarded as a number of words). If the phrase length is too large (for instance, more than four words), you must determine whether the phrase contains any simple lexical signs (certain words or word combinations) allowing it to be divided into smaller phrases.

Experimental studies show that, in many cases, such lexical signs may include:

- coordinate conjunction “і” (*and*) – phrase segmentation before “і”;
- disjunctive conjunction “ці” (or) – phrase segmentation before “ці”;
- proper names – phrase segmentation after the last in a list of proper names;
- abbreviations – phrase segmentation after abbreviations;
- digit position names – phrase segmentation after each digit position name;
- names of months, the words “гадзіна, хвіліна” (*hour, minute*) when writing dates and times in full unabbreviated form – phrase segmentation after dates and times.

This list is not complete and may be expanded through the analysis of increasingly more large-scale text and speech corpora.

3 Intonation Types of Phrases

Based on the punctuational phrase segmentation described above, four main categories of punctuational phrase intonation types may be identified: finality (P), non-finality (C), interrogation (Q) and exclamation (E).

The category of a declarative sentence is characterized by final intonation. There are following types of final intonation:

– “full stop” intonation - {.},
– “ellipsis” intonation - {…},
– “title” intonation - {*},
– “paragraph” intonation - {$}.

Apart from these types of final intonation, characteristic of end-of-sentence phrases, there are two additional punctuational types of intonation with different degrees of finality:

– “semicolon” intonation – {;},
– parenthetical intonation – {)}, {,–}, {–}.

Parenthetical intonation presupposes that the marks specified should be preceded by the following marks respectively: {(}, {,–}, {–}.

There may also be four punctuational subtypes of intonation with different degrees of non-finality in a sentence:

– “comma” intonation - {,},
– “dash” intonation - {–},
– “colon” intonation - {:},
– pre-parenthetical intonation - {(}, {,–}, {–}.

Pre-parenthetical intonation presupposes that the marks specified should be followed by the {(}, {,–}, {–} marks.

Punctuational phrases may in turn contain lexical phrases with three types of non-final intonation:

– intonation of conjunction “i” (*and*)
– intonation of conjunction “ці” (*or*),
– intonation of lexical phrases with proper names, abbreviations, digit position names and date and time names.

Interrogation and exclamation intonations are identified respectively by question and exclamation marks.

4 Realization of Phrase Segmentation and Intonation Type Marking with NooJ

Based on the segmentation techniques proposed above, the authors have developed a NooJ syntactical grammar (Fig. 1) representing the initial stage of prosodic processing in a speech synthesizer.

The grammar carries out marking of the following intonation types:

– Non-finality (C) – 11 types;
– Finality (P) – 11 types;
– Exclamation (E) – 2 types;
– Interrogation (Q) – 2 types.

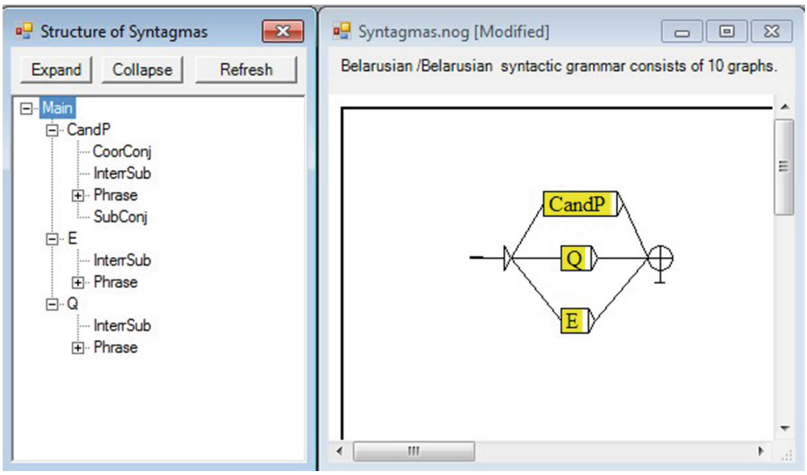

Fig. 1. General view of the grammar for marking phrases at punctuation level

As punctuational phrases may be different in length, in the grammar, by punctuational phrase, we mean any sequence of word forms or phonetic words separated by punctuation marks (Fig. 2).

When marking the intonation type of a phrase, not only punctuation marks are taken into account but also the nearest context in a text. For example, the intonation type "C31" corresponds to each second phrase in a line of consecutive "C3"-phrases – phrases with non-final "comma" intonation (Fig. 3).

In general, depending on the nearest context, there are six types of non-final "comma" intonation:

– C3 - simple "comma" intonation (C3),
– C7 - "comma" intonation with a comma followed by a coordinate conjunction (which indicates the beginning of P7 phrase with a "full-stop" final intonation),

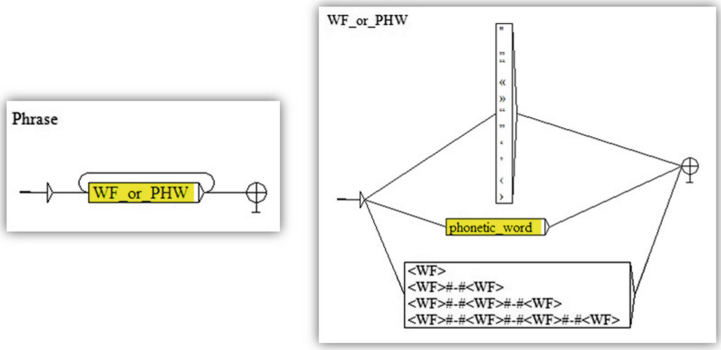

Fig. 2. Phrase

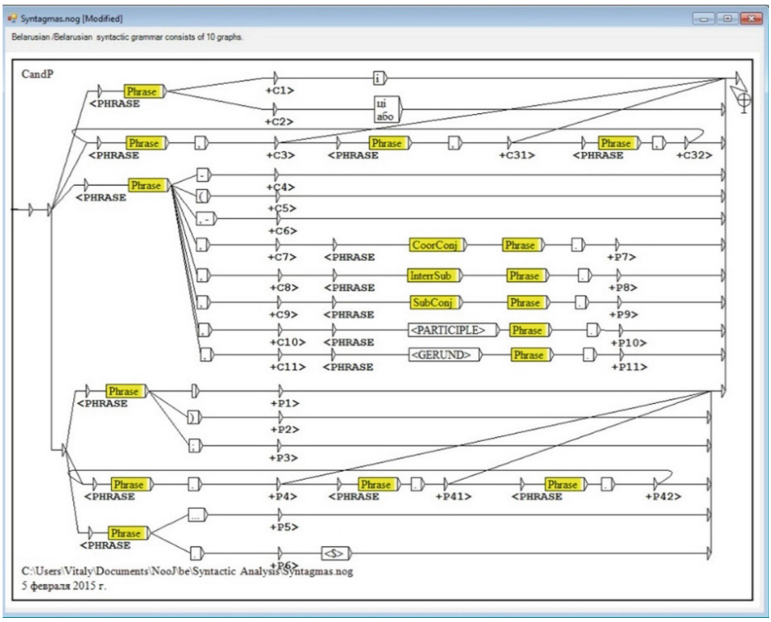

Fig. 3. Non-finality and finality

- C8 - “comma” intonation with a comma followed by an interrogative word (which indicates the beginning of P8 phrase with a “full-stop” final intonation),
- C9 - “comma” intonation with a comma followed by a subordinating conjunction (which indicates the beginning of P9 phrase with a “full-stop” final intonation),
- C10 - “comma” intonation with a comma followed by a participle (which indicates the beginning of P10 phrase with a “full-stop” final intonation),
- C11 - “comma” intonation with a comma followed by a gerund (which indicates the beginning of P11 phrase with a “full-stop” final intonation).

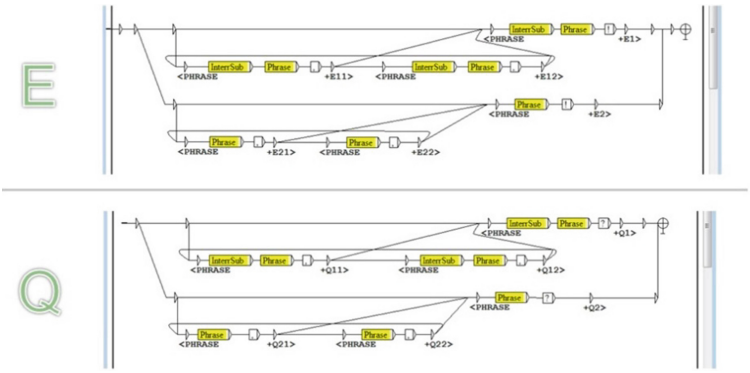

Fig. 4. Exclamation and interrogation

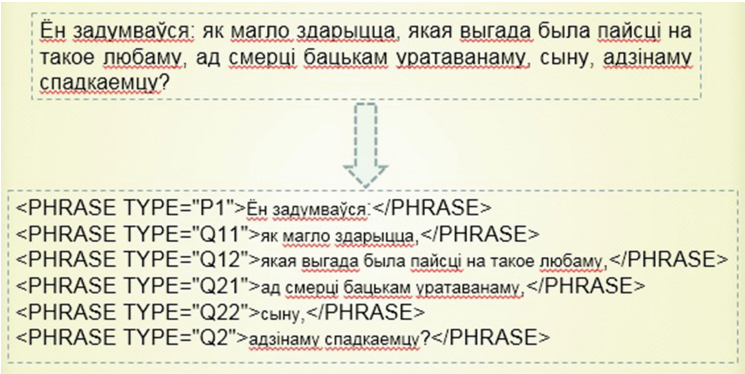

Fig. 5. Applying the grammar to a text

Figure 4 shows sub-graphs for marking exclamation and interrogation. Depending on whether there is an interrogative word starting an exclamation or interrogation phrase, there are two types of exclamation intonation and two types of interrogation intonation.

After applying the grammar to a text, we obtain a text annotated with corresponding intonation indexes (Fig. 5).

In Fig. 5, phrases in the sentence are annotated with the following indexes:

- "P1" – final "colon"-intonation,
- "Q11" and "Q12" – interrogative intonations corresponding respectively to each uneven and each even phrase in a line of consecutive "Q1"-phrases - interrogative non-final phrases containing an interrogative word,
- "Q21" and "Q22" – interrogative intonations corresponding respectively to each uneven and each even phrase in a line of consecutive "Q2"-phrases – interrogative non-final phrases without an interrogative word,
- "Q2" – intonation type corresponding to an end-of-sentence interrogative phrase without an interrogative word.

5 Graphical Representation of Phrase Intonation

One more option for using the results of this grammar is for the marking of a text by graphical intonation. The idea is that each phrase in a text is represented by a particular colour and will be accompanied with a prosodic portrait (contour) corresponding to the phrase's intonation type.

By prosodic portrait, we mean a graphical representation of phrase intonation. This method was developed by one of the authors, Boris Lobanov, and is called Portraits of Accentual Units, where an accentual unit is a minimal prosodic unit in speech (discussed in greater detail in [5]).

```
while (<IN>)
    {
        s/PHRASE/FONT/g;
        s/<FONT TYPE="(.*?)"[^>]*>(.*?)<\/FONT>/ <FONT TYPE="$1">$2<\/FONT> <FONT STYLE=\"text-decoration:underline;
color:black;\">($1)<\/FONT><img src="..\/bd\/$1a1.jpg" width="70px">/g;
        if (1)
        {
            s/TYPE="C1"/STYLE="text-decoration:underline; color:blue;"/g;
            s/TYPE="C2"/STYLE="text-decoration:underline; color:blue;"/g;
            s/TYPE="C3"/STYLE="text-decoration:underline; color:blue;"/g;
                s/TYPE="C31"/STYLE="text-decoration:underline; color:blue;"/g;
                s/TYPE="C32"/STYLE="text-decoration:underline; color:blue;"/g;
            s/TYPE="C4"/STYLE="text-decoration:underline; color:blue;"/g;
            s/TYPE="C5"/STYLE="text-decoration:underline; color:blue;"/g;
            s/TYPE="C6"/STYLE="text-decoration:underline; color:blue;"/g;
        }
        #s///g;
        #s///g;

      $lines++;
        print OUT $_;
    }
```

Fig. 6. Intonation marker program code

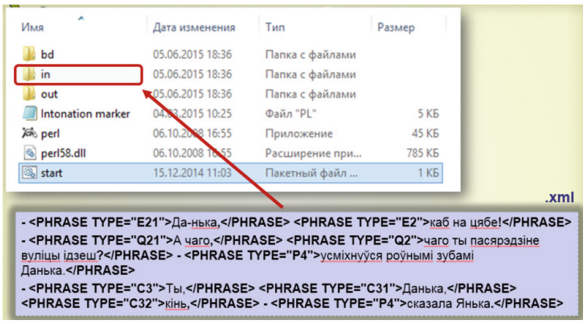

Fig. 7. Marking phrases

For the purpose described, the authors have developed a Perl program which processes texts after having been annotated with the NooJ grammar for phrase segmentation and intonation marking (Figs. 6 and 7).

When processing texts, the program highlights and marked segments of text and applies the base of intonation contours created previously within the framework for developing models of prosodic portraits. All this results in the creation of documents in HTML format, with texts marked as shown in Fig. 8. Such markings serve as instructions on how to read a text.

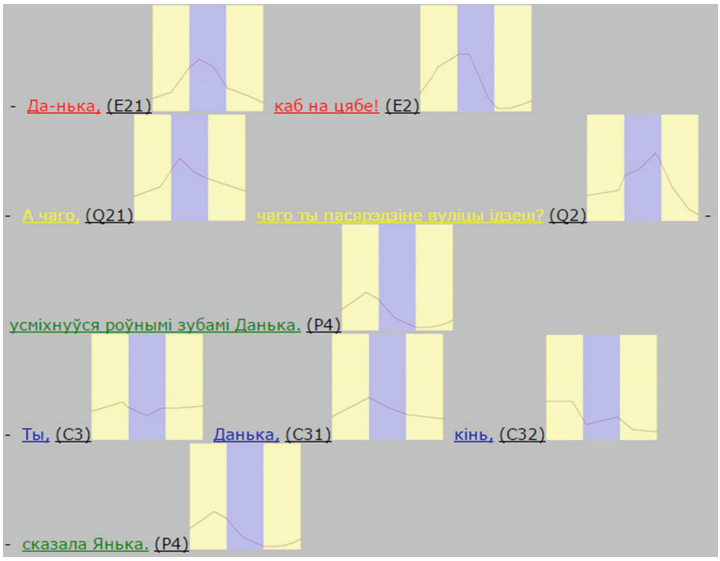

Fig. 8. Graphical representation of phrase intonation

6 Conclusion

In this paper we have presented a technique for automated phrase segmentation at the punctuational level and a system of marking types of phrase intonation in electronic Belarusian texts using NooJ.

This technique seeks to boost the prosodic performance of the Belarusian text-to-speech system and may also serve to improve the Belarusian NooJ module with so-called prosodic transcription.

Alternatively, this technique could be used for educational purposes, for example in learning Belarusian phonetics.

References

1. Pitrelli, J.F., et al.: The IBM expressive text-to-speech synthesis system for American English. IEEE Trans. Audio Speech Lang. Process. **14**(4), 1099–1108 (2006)
2. Froehlich, P., Hammer, F.: Expressive text-to-speech: a user-centred approach to sound design in voice-enabled mobile applications. Telecommunications Research Centre Vienna (FTW). (http://userver.ftw.at/~froehlich/papers/JDS2004_ExpressiveTTS.pdf)
3. Leskovec, L.: Lexical Stress Assignment and Pronunciation Formalization in Expressive TTS/ Trinity College Dublin. (http://www.cs.tcd.ie/research_groups/clg/COST2102.IS2009/content/program/node61.html)

4. Kawanami, H., et al.: Designing Speech Database with Prosodic Variety for Expressive TTS system, Nara Institute of Science and Technology, Takayama-cho. (http://gandalf.aksis.uib.no/lrec2002/pdf/337.pdf)
5. Lobanov, B.M., Tsirulnik, L.I.: Computer speech synthesis and cloning. Belarussian science, p. 342 (2008). (in Russian)