ОБРОБЛЕННЯ СИГНАЛІВ І ЗОБРАЖЕНЬ ТА РОЗПІЗНАВАННЯ ОБРАЗІВ

Signal/Image Processing and Pattern Recognition

ПРАЦІ ✵ PROCEEDINGS

ДВАНАДЦЯТА ВСЕУКРАЇНСЬКА МІЖНАРОДНА КОНФЕРЕНЦІЯ

3 – 7 листопада 2014 року ✵ Київ, Україна

THE TWELFTH ALL-UKRAINIAN INTERNATIONAL CONFERENCE

November 3 – 7, 2014 ✵ Kyjiv, Ukraine

ISBN 978-966-479-069-4

Статті публікуються без змін, зі збереженням авторського оригіналу
The papers are published without any change in the text that have been presented by authors

Веб-сторінка конференції:
The conference homepage: http://www.uasoiro.org.ua/ukrobraz2014/

Видання УАсОІРО
(40, проспект Академіка Глушкова, Київ 03680, Україна)

Наклад 60 примірників

III. Усномовні інформаційні технології Speech Information Technologies

IV. Комп'ютерні ресурси та технології письмової мови Natural Language Resources and Technologies

Processing of Direct Speech in Belarusian Texts with NooJ

Yury Hetsevich, Tatsiana Okrut,
Boris Lobanov

United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus
yury.hetsevich@gmail.com, tatberrie@gmail.com, lobanov@newman.bas-net.by

Abstract

This paper focuses on processing of direct speech in Belarusian electronic texts for the purpose of audiobook creation. Usually, for creation of an audiobook, synthesis with only one voice is used. It gives us perspective on the likelihood of making text-to-speech synthesis many-voiced, thus making audiobooks more approximate to the representation of characters' unique speech features.

1. Introduction

The work on direct speech processing was started in 2013 by the group of researchers from the United Institute of Informatics Problems of the National Academy of Sciences of Belarus. The main goals of the study include the identification of all cues in electronic texts and the identification and processing of cues with authors' text insertions for the aims of audiobook creation.

Actually, European scientists have already developed algorithms for character identification and automatic determination of the character's role with the help of NooJ syntactic grammars [1]. As for the Slavic languages, the work of Croatian scientists on the direct speech identification should be noted [2], though they do not consider the problem of gender identification. Such programs for audiobook creating as MP3book2005 and AUDIOBOOK are also developed in this direction. They have special inbuilt units for logical analysis of dialogues, which can provide the marking of the characters' and author's words in a dialogical text. In AUDIOBOOK steps were taken to read dialogues in character, but the program does not cover all the cases. It ignores the cue structures with more than one insertion of the author's words. In addition, it is not able to identify the gender of a character on such indicator as a "verb + masculine noun" combination in the author's words:

– Трэба напісаць "яць", – адказвае вучань.
(– We should write "яць", – the pupil (he) answers.)

Thus, the tasks confronting the authors include the algorithm development for direct speech processing to formalize as much syntactic structures of dialogical text as possible, and to identify automatically the gender of a character by the insertions of the author's words in the direct speech. We also discuss the use of the developed algorithms in a TTS system.

2. The development of automated algorithm for direct speech and author's text identification

At the first stage we have selected texts in Belarusian and identified all the paragraphs with direct speech. The found paragraphs were separated according to the characters' gender and all the cues with author's text insertions were also marked. Then the cues were analyzed to define the syntactical direct speech structures and to detect gender indicators (such as past tense verbs and nouns with gender attributes) in author's text insertions.

The following syntactic structures were revealed in direct speech:

Direct speech apart from the author's text:
– C (! | !! | !!! | ? | ?! | … | .).
Direct speech followed by the author's text:
– C (, | ! | !! | !!! | ? | ?! | … | .) – A (… | .).
Direct speech with one or more insertions of the author's text:
– C (, | ! | !! | !!! | ? | ?! | … | .) – A (, | … | . | : | .) – C (, | ! | !! | !!! | ? | ?! | … | .) (– A (, | … | . | : | .) – C (, | ! | !! | !!! | ? | ?! | … | .)).

The structures contain the following annotations: C – the words of a character (speaker), A - the author's text, brackets (,) – the beginning and the end of a choice set of punctuation marks, | - symbol or (separation of punctuation marks in a choice set).

On the basis of these findings the algorithm for direct speech identification was developed. The main idea is that only those paragraphs are taken into consideration that begin with a dash. After a dash being found, the following elements of the paragraph are alternatively defined as the character's words and the author's words. The algorithm's complexity consists in indicating of a set of characters that separate the character's part from the author's part.

Let us describe the developed algorithm:

1. Process the next paragraph TT of a text T. If TT = Ø, then go to Step 14, otherwise go to Step 2.
2. If TT begins with a dash, then go to Step 3, otherwise – Step 1.
3. If a sequence of any number of SSw1 set's elements is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 4.
4. If a sequence of any number of SSw1 set's elements with SSp2 set's elements placed between them (not several elements in succession) is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 5.
5. If a sequence, starting with any element of SSp1 followed by any number of SSw1 set's elements, is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 6.
6. If a sequence, starting with any element of SSp1 followed by any number of SSw1 set's elements with SSp2 set's elements placed between them (not several elements in succession), is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 7.

7. If a sequence of any number of SSw1 set's elements with SSp1 set's elements placed between or after SSw1 set's elements (not several elements in succession) is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 8.

8. If a sequence, starting with any element of SSp1 followed by any number of SSw1 set's elements with SSp1 set's elements placed between or after SSw1 set's elements (not several elements in succession), is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 9.

9. If a sequence of any number of SSw1 set's elements with SSp1 set's elements placed between or after SSw1 set's elements (not several elements in succession) and SSp2 set's elements placed between SSw1 set's elements (not several elements in succession) is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 10.

10. If a sequence, starting with any element of SSp1 followed by any number of SSw1 set's elements with SSp1 set's elements placed between or after SSw1 set's elements (not several elements in succession) and SSp2 set's elements placed between SSw1 set's elements (not several elements in succession), is found next, and at the end of which there is any element of SSp3 set, then go to Step 11, otherwise – Step 1.

11. If a dash is found next, then go to Step 12, otherwise conclude that TT belongs to direct speech and go to Step 1.

12. If a sequence of any number of SAw1 set's elements is found next, and at the end of which there is any element of SAp2 set, then go to Step 14, otherwise – Step 13.

13. If a sequence of any number of SAw1 set's elements with SAp1 set's elements placed before, between or/and after them is found next, and at the end of which there is any element of SAp2 set, then go to Step 14, otherwise – Step 1.

14. If a dash is found next, then go to Step 4, otherwise conclude that TT belongs to direct speech and go to Step 1.

15. The end of the algorithm.

Within the given algorithm, Steps 3-10 correspond to the speaker's words identification and Steps 12-13 correspond to the author's words identification. For the description of direct speech processing, the following sets of elements were used: SSw1, SSp1, SSp2, SSp3. Respectively, to describe the author's word's identification the SAw1, SAp1 i SAp2 sets were used.

Let us describe the elements of the sets being used. SSw1 and SAw1 sets include the rules of possible speaker's and author's words formation and have the following filling: SSw1 = SAw1 = {<WF>, <NB>, <WF>#-#<WF>}, where WF is any word form, and NB is any number. SSp1, SSp2, SSp3, SAp1 and SAp2 sets represent a range of different punctuation marks and are filled as follows:

SSp1 = {/\"/, /, \"/, /. \"/, /\".\/, /\",/ , /...\"/, /… \"/, /\"...\/, /\: \"/, /\: \"...\/, /!.. \"/, /\"... \"/, /\"…/, /\", \"/, /\". \"/, /- \"/, /– \"/, /\" -/, /\" –/, /\"?/, /?!\"/, /!\", \"/, /!\" \"/, /\")/, /\" (/, /...\/, /…/, /“/, /”/};

SSp2 = {<P>, /?!/, /...\/, /…/, /…/, /…/, /-/, /–/, /;/, /!../,/?../, /!!!/, /).\/, /),/, /\:/, /(/, /)/, /) -/, /) –/};

SSp3 = {/./, /,/, /!/, /?/, /.../, /?!/, /…/, /!../, /?../, /)./, /\"?/, /…\"/, /...\"/, /\"./, /”/};

SAp1 = {/./, /,/, /.../, /\:/, /(/, /)/};

SAp2 = {/./, /,/, /…/, /)./, /\:/}.

The developed algorithm was implemented with the help of a computational linguistic tool NooJ, notably with its Belarusian module. NooJ allows to develop syntactical and morphological grammars and to test them on a large number of texts. Thus, the authors have developed a NooJ syntactic grammar DS_All for automated identification of all paragraphs containing direct speech (Fig. 1). Its main parts - graphs Speaker (Figure 2) and Author (Figure 3) - serve to identify respectively the character's words and the author's words.

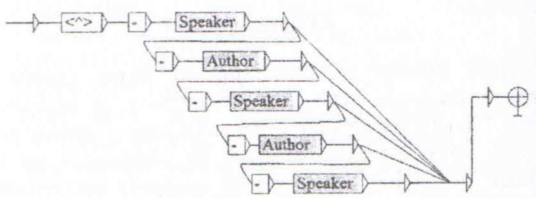

Figure 1: A general view of the DS_All grammar

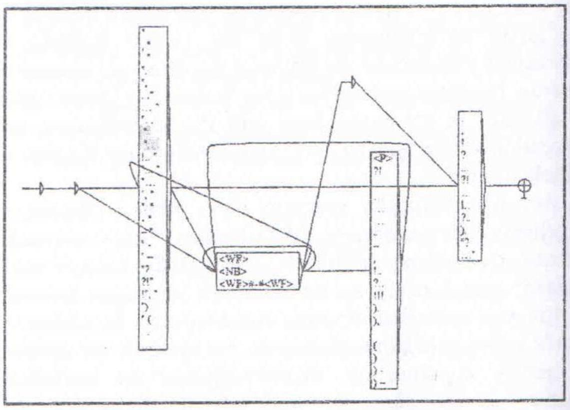

Figure 2: Syntactic subgraph Speaker of the DS_All grammar

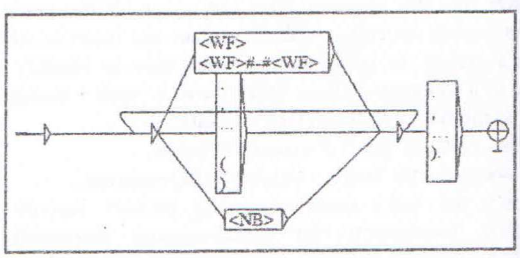

Figure 3: Syntactic subgraph Author of the DS_All grammar

The resulting grammar can be applied sequentially to any Belarusian electronic text through NooJ, and the program function Locate Pattern allows to view the results of the grammar applied in the form of concordance (Fig. 4).

Before	Seq.	After
ўчына.	- Вось бачыце, шкада толькі, што вы ад нас далёка, а то б...	- А хі(
ськаю?	- А хіба тут няма каму гэтай справай заняцца? Вось мой к...	Сахан
іяліся.	- Не, я ўжо зусім страціў там ласку, дзякаваць Богу.	Айцеі
заў ён.	- Апрача таго, я чуў, што ў яе жаніх ёсць ужо.	- Ці м
ць ужо.	- Ці мала на свеце дурняў, - зноў дадаў а. Кірыл.	Матуш
а сваё:	- Ну, то што? Хіба жаніхам свіней не падкладаюць?	- Гэта
даюць?	- Гэта было б не па-хрысціянску.	- Зато

Figure 4: The results of applying the DS_All grammar to a Belarusian text

3. The development of automated algorithms for character's gender identification from the author's text

In the Belarusian language singular past tense verbs may have gender attributes, for example, *направіў* 'he corrected' and *сказала* 'she said'. As such verbs may often occur in author's commentaries to direct speech, as well as some nouns having gender attributes, they may serve as gender indicators and be considered suitable for character's gender identification.

For the purpose of gender identification we have modified the algorithm mentioned above, namely in Steps 12-13 we used one more set with gender indicators. Then on the basis of the grammar DS_All two separate grammars were developed – one for masculine gender identification (DS_M), and one for feminine gender identification (DS_F) (see [3] for more details). For this purpose, we worked further the graph Author and added resources for gender identification (Figure 5). In the Figure 5, one can see the subgraph VERBSmasculine. It includes the list of masculine verbs, which were selected at the stage of manual marking of texts in the Belarusian and Russian languages. The similar list of verbs was created within the subgraph VERBSfemenine for feminine gender identification.

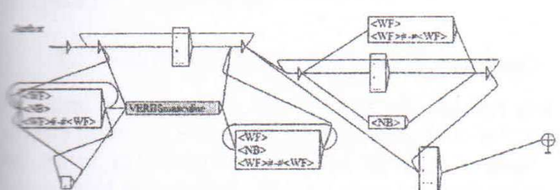

Figure 5: The subgraph Author, DS_M

In order to use the outputs of the grammars in TTS system SAPI 5.1, it is necessary to adapt a text to a SAPI TTS XML format. Therefore, to select an appropriate speech synthesizer, a syntactic grammar should provide annotations of the following kind:
<VOICE Required="name=[a synthesizer's name in TTS system]">
...A text for synthesis...
<VOICE>.

Thus, in the Figure 6 one can see, that the speech synthesizers BorisBel and AlesiaBel will be respectively applied to the character's words (Speaker) and to the author's words (Author).

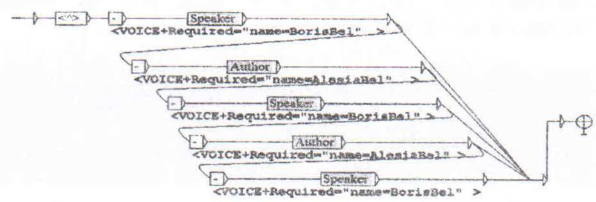

Figure 6: The DS_All grammar after being adapted for SAPI 5.1

For example, after being processed by the DS_M and DS_F grammars, a dialogical text will be annotated as in the Figure 7. A female voice AlesiaBel is applied to the author's words, and voices ElenaBel and BorisBel are used for the female and male characters' words. Such annotation allows to input texts into the TTS system SAPI 5.1, where the indicated voices switch over are automatically (Figure 8).

```
<VOICE Required="name=ElenaBel">- Бацька вады,</VOICE> <VOICE Required="name=AlesiaBel">
- шэптам сказала Майка.</VOICE>
<VOICE Required="name=BorisBel">- Бацька вод,</VOICE> <VOICE Required="name=AlesiaBel">
- направіў Алесь.</VOICE>
<VOICE Required="name=BorisBel">- Вось так і Дняпро пачынаецца недзе.</VOICE>
<VOICE Required="name=ElenaBel">- Жывая вада,</VOICE> <VOICE Required="name=AlesiaBel">
- сказала Яня.</VOICE>
І яна апусцілася на калені і зламала пальчыкамі крыштальную паверхню.
- Піце. Будзеце жыць сто год.
```

Figure 7: The sentences from the Table 1 after being annotated with VoiceXML tags

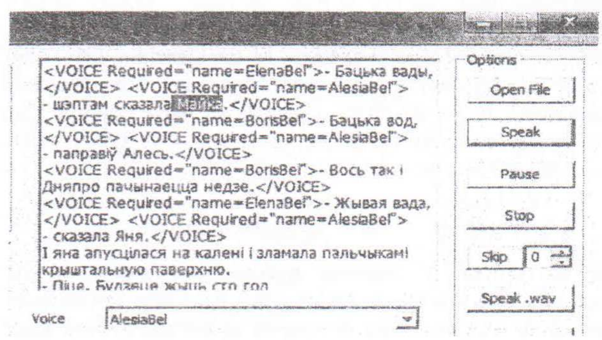

Figure 8: The speech synthesis of the annotated sentences

4. Multi-coloured marking of a text

One more application for the described annotation is the multi-coloured marking of a text for visual presentation of the author's words and of the female and male characters' words. Such marking may be used by an editor to quickly analyze direct speech in a text and to select an optimal number of speech synthesizers or speakers.

To provide the multi-coloured marking, the authors have developed the VoiceXmlToColorReplacer software. The program process VoiceXML-files and allows converting VoiceXML-tags of speech synthesizers into HTML-tags with different styles of direct speech visual presentation.

After a text passes through the VoiceXmlToColorReplacer software, the character's cues and the author's insertions are marked with different colours: namely, author's words (AliesiaBel) are in black, the male character's cues (BorisBel) – in blue, and the female character's cues (ElenaBel) – in red (Figure 9).

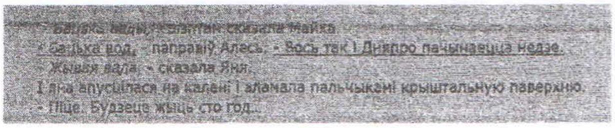

Figure 9: A fragment of a text with multi-coloured marking of direct speech

5. Evaluation

Initially, a training text corpus with 106 000 word usages was used in developing the grammar. Then, in the process of testing, the experts have collected a test text corpus with 23 867 word usages. According to the performance evaluations, the total number of cues in the corpus was equal to 481 (N=481). Among them 233 cues include the author's text insertions, where 165 cues belong to male characters, 68 cues belong to female characters.

The quantity of all cues found by the algorithm DS_All (be) is L=462; the number of those which have been correctly processed is M=461. The calculations have showed the following results for DS_All (be): precision ≈ 99,5 %, recall ≈ 95,8 %, and F-score ≈ 97,6 %.

The quantity of all cues found by the algorithm DS_M (be) is L=145; the number of those which have been correctly processed is M=143. The calculations have showed the following results for DS_M (be): precision ≈ 98,6 %, recall ≈ 86,6 %, and F-score ≈ 92,2 %.

The quantity of all cues found by the algorithm DS_F (be) is L=58; the number of those which have been correctly processed is M=67. The calculations have showed the following results for DS_F (be): precision ≈ 98,2 %, recall ≈ 83,8 %, and F-score ≈ 90,4 %.

6. Conclusion

In the process of character gender identification on the author's text insertions, rather good operating results were obtained. Moreover, the developed algorithm showed itself suitable for the use in combination with a TTS system and later may be applied in audiobook creation with reflecting the unique speech characteristics of characters.

However, text processing at the paragraph level is not sufficient for character gender identification in all cues. There are a lot of cues without author's text insertions, that is why now we face thè task of gender identification directly from the character's words, and the most significant challenge is to provide text-level gender identification through the analysis of the text going before and after the cues. Moreover, further work needs to be done to create dictionary resources with verbs-indicators identifying, to expand the punctuation base (dash and quotation types, etc) and the test corporas.

7. References

[1] Lendvai, P, "Assignment of character and action types in folk tales". In: Gavriilidou Z., Chatzipapa E., Papadopoulou L., Silberzstein M. (eds.), *Selected Papers from the NooJ 2010 Intern. Conf. Formalising Natural Languages with NooJ.* Democritus University of Thrace, Greece. 102–111.

[2] Jurić, T., Stupar, M., Boras, D. "Direct Speech Recognition in Text". In: Vučković K., Bekavac B., Silberztein M. (eds.), *Selected Papers from the NooJ 2011 Intern. Conf. Automatic Processing of Various Levels of Linguistic Phenomena.* Cambridge Scholars Publishing, Newcastle, pp. 122–127, 2012.

[3] Hetsevich, Yu. S., Okrut, T. O. and Lobanov, B. M. "Алгарытмы ідэнтыфікацыі рэплік са словамі аўтара ў электронных тэкстах на беларускай мове", *Informatics*, 1 (41): 68-76, 2014.