search for terms into the text (e.g. in the quoted speeches of a particular character), explore connections between
characters in their social networks.

The CBook is used by Italian students and their teach-
ers, as an innovative digital tool for reading and studying
literature, in primary and secondary schools. The list of
schools involved in Librare project, for example, can be
found here:

Some Italian and English classics are already present
in the web app, the full work or a portion: I promessi
sposi (complete), Le avventure di Pinocchio (complete),
Decameron (3 novels), Odissea (book VI, Italian transla-
tion), Romeo and Juliet (complete), The Wonderful Wizard
of Oz (chapters 15). Other works requested by teachers
(e.g. Dante’s Purgatorio), will be added to the web app in
the next months.

We can say that CBook is, in some sense, a collabora-
tive anthology (or an Anthology 2.0), i.e. a collection of
literary works selected by the users that are reading, an-
notating and exploring the works, using some DH tools
and methodologies (collaborative annotation, interactive
maps, SNA of characters, text mining, etc).

The web app was designed and tested in collaboration
with high school teachers, students and DH scholars. It is
used on tablets, PCs, interactive whiteboards, and smart-
phones (with less functionalities).

An important feature of CBook is interoperability
between models and tools for TEI XML annotation and
models and tools for semantic annotation. A significant
difference between the CBook and other annotation tools
is its emphasis on user-centered design (for students and
teachers).

The annotation system of CBook is based on Annotator
(http://annotatorjs.org/), “an opensource JavaScript library
to easily add annotation functionality to any webpage”,
following annotation standards for digital documents
developed by the W3C Web Annotation Working Group.
The User Interface is written in HTML5, CSS and Javascript.
The texts were annotated in XML, using a simplified ver-
sion of TEI. On the server side, we use a graph Database
and Java.

The actions of cbook’s users (sign in, sign out, comment)
can be monitored in a public dashboard accessible here:
http://sensori.librare.org/librare/. In this dashboard you
can see the events concerning both ebooks (digital) and
paper books that users were working on in this project.

The first version of “The digital lab of crunched
books” was the Second Runner Up of Digital
Humanities Awards 2014 in the Best DH Tool or Suite
of Tools section (3 march 2015): http://dhawards.org/
dhawards2014/results/

Bibliography

Interactive e-Books for Children IBookC2014 (Denmark), 2nd
Workshop at IDC Interaction Design and Children.

the Betrothed Lovers, in M. Finlayson, B. Fisseni, B. Lowe
and J. C. Meister (eds.), 2013 Workshop on Computational
Models of Narrative, OASIC, vol. 32.


Verbal Identity of a Fictional
Character: a Quantitative Study with
a Machine Learning Experiment

Anastasia Bonch-Osmolovskaya
abonch@gmail.com
National research university ‘Higher school of economics’,
Moscow, Russia

Elena Sidorova
sieleny@gmail.com
National research university ‘Higher school of economics’,
Moscow, Russia

Daniil Skorinkin
skorinkin.daniil@gmail.com
National research university ‘Higher school of economics’,
Moscow, Russia

Introduction

The idea that narrative literature comprises two distinct
types of speech – that of the narrator and that of charac-
ters – dates as far back as Plato’s Republic. The philoso-
pher distinguished diegesis (narrative, narration), when
the author speaks for himself, from mimesis (imitation, enactment), when the author puts words in the mouths of his or her fictional actors.

Modern narratology places high emphasis on the concepts of point of view (POV) and POV structure of a text (Schmid, 2003), which are often expressed through specific combinations of author's and character's speech. Authors may switch between the POV’s by employing character-specific lexica and the use of temporal and spatial references that indicate certain POV (Uspensky, 1983).

Leo Tolstoy was one of the writers known for conscious usage of such means to differentiate between the character POVs. He was a firm proponent of the idea that each character has to speak his/her own language if the book was to be convincing. Critics confirm that Tolstoy’s characters do have their personal styles of conversation.

In this paper we made an attempt to provide quantitative grounds for these claims. For that purpose we extracted all speech activity instances from War and Peace, attributed them to the speaker characters and used the data to train a classifier. Our hypothesis was that if Tolstoy’s characters actually possessed these unique speech features, the classifier would be able to predict the speaker with some tolerable accuracy.

Data

Instances of direct speech were extracted from the text with help of ABBYY Compreno (Starostin et al., 2014). For more details on the extraction procedure see (Bonch-Osmolovskaya A., Skorinkin D., 2015). The total number of extracted speech instances was 6853, of which 4476 had their speakers identified.

Apart from the speaker, a number of additional attributes were extracted for each instance: text of the speech, text of the author’s introduction (‘she cried,’ ‘he said with a laugh’), normalized speech predicate (‘to say,’ ‘to cry,’ ‘to whisper,’ ‘to burst out’), the number of question and exclamatory phrases within one speech, the number of words in the speech and the number of punctuation marks.

Before we carried out the experiment we attempted to analyze the data and detect potentially informative features. It appeared that certain characters (Natasha Rostov, Nikolai Rostov) tend to speak in short intermittent bursts and exclaim a lot, so they were expected to have higher average punctuation marks per word ratio and bigger shares of exclamatory speech. To confirm this intuition we gathered some aggregated statistics (see Fig. 1).

Exclamatory and question phrases together make up the ‘emotional part’ of a character’s speech. Its share (Fig. 2) seems to correlate with age extremities. Prince Nikolay Bolkonsky is probably the oldest of the main characters, and as his age gets the better of him in the course of the novel, he turns more and more emotional and impulsive; Petya Rostov, on the other hand, is an exuberant and emotional boy, the youngest of the Rostov family.

Machine learning experiment

Next step was to try and use some of these features to train a classifier. We used several standard algorithms, of
which Random Forest demonstrated the best results. At the first stage we created a baseline by training the classifier solely on the lemma and word form frequencies of speech. Table 1 shows the results we obtained.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natasha Rostova</td>
<td>0.3</td>
<td>0.394</td>
<td>0.341</td>
</tr>
<tr>
<td>Nikolai Rostov</td>
<td>0.215</td>
<td>0.202</td>
<td>0.209</td>
</tr>
<tr>
<td>Sonya Rostova</td>
<td>0.27</td>
<td>0.181</td>
<td>0.217</td>
</tr>
<tr>
<td>Pierre Bezukhov</td>
<td>0.334</td>
<td>0.44</td>
<td>0.38</td>
</tr>
<tr>
<td>Andrey Bolkonsky</td>
<td>0.218</td>
<td>0.129</td>
<td>0.162</td>
</tr>
<tr>
<td>Mariya Bolkonskaya</td>
<td>0.112</td>
<td>0.195</td>
<td>0.142</td>
</tr>
<tr>
<td>Vasily Denisov</td>
<td>0.667</td>
<td>0.271</td>
<td>0.385</td>
</tr>
<tr>
<td>Fedor Dolokhov</td>
<td>0.238</td>
<td>0.119</td>
<td>0.159</td>
</tr>
<tr>
<td>Mikhail Kutuzov</td>
<td>0.194</td>
<td>0.09</td>
<td>0.123</td>
</tr>
<tr>
<td>Weighted Avg.</td>
<td>0.279</td>
<td>0.269</td>
<td>0.261</td>
</tr>
</tbody>
</table>

Table 1 Baseline results for classifier trained on lemma and word form frequencies

The second stage was to add formal features that we considered informative (number of exclamatory phrases, number of questions and punctuation marks per word ratio) and retrain the classifier. The results we obtained show that the use of these features slightly improved performance.

<table>
<thead>
<tr>
<th>Class</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Natasha Rostova</td>
<td>0.3</td>
<td>0.394</td>
<td>0.341</td>
</tr>
<tr>
<td>Nikolai Rostov</td>
<td>0.215</td>
<td>0.202</td>
<td>0.209</td>
</tr>
<tr>
<td>Sonya Rostova</td>
<td>0.27</td>
<td>0.181</td>
<td>0.217</td>
</tr>
<tr>
<td>Pierre Bezukhov</td>
<td>0.334</td>
<td>0.44</td>
<td>0.38</td>
</tr>
<tr>
<td>Andrey Bolkonsky</td>
<td>0.218</td>
<td>0.129</td>
<td>0.162</td>
</tr>
<tr>
<td>Mariya Bolkonskaya</td>
<td>0.112</td>
<td>0.195</td>
<td>0.142</td>
</tr>
<tr>
<td>Vasily Denisov</td>
<td>0.667</td>
<td>0.271</td>
<td>0.385</td>
</tr>
<tr>
<td>Fedor Dolokhov</td>
<td>0.238</td>
<td>0.119</td>
<td>0.159</td>
</tr>
<tr>
<td>Mikhail Kutuzov</td>
<td>0.194</td>
<td>0.09</td>
<td>0.123</td>
</tr>
<tr>
<td>Weighted Avg.</td>
<td>0.279</td>
<td>0.269</td>
<td>0.261</td>
</tr>
</tbody>
</table>

Table 2 Results for classifier trained with additional features

Results & discussion
Our first attempts to automatically classify speaker in Tolstoy’s text did not prove successful. The best F-measure we were able to obtain so far does not exceed 0.385 for an individual character. However, we were able to show that some formal features, such as punctuation marks per word ratio or the number of exclamatory/question sentences, might improve classification quality. This assumption can be confirmed by the figures in the Data section, where the aggregated values of features correspond with certain character traits that are apparent to the human reader.

Bibliography

Mobile Makerspaces: Te(a)chnology, Design and Digital Humanities
Christina Boyles
christina-boyles@uiowa.edu
University of Iowa, United States of America
Lindsay Kistler Mattock
lindsay-mattock@uiowa.edu
University of Iowa, United States of America

This project explores the development of a mobile makerspace for graduate and undergraduate DH scholars at the University of Iowa and Grinnell College. Blending the approaches of makerspaces like the DHMakerBus <http://dhhmakerbus.com> and University of Victoria’s Maker Lab <http://maker.uvic.ca/>, this DH makerspace will investigate the use of a suite of tools designed to support the development digital literacy and technological proficiency for students across the DH curriculum. By combining the mobility of the DHMakerBus with the experimental computing of the University of Victoria’s Maker Lab, we will produce a new method of digital humanities pedagogy that welcomes the participation of primary and secondary students and educators, local citizens, and digital humanities practitioners.

Designed to support experiential learning - learning