THE RULES OF TEXT SIMPLIFICATION FOR FOREIGN LANGUAGE LEARNERS

USE CASE: JANE AUSTEN’S PRIDE AND PREJUDICE

Emily Franzini and Marco Büchler
PRIDE
AND
PREJUDICE:
A NOVEL.
IN THREE VOLUMES.

BY THE
AUTHOR OF "SENSE AND SENSIBILITY."

VOL. I.

London:
PRINTED FOR T. EGERTON,
MILITARY LIBRARY, WHITEHALL.
1813.
Definition:

Graded readers are “simplified books written at varying levels of difficulty for second language learners”, which “cover a huge range of genres ranging from adaptation of classic works of literature to original stories, to factual materials such as biographies, reports and so on” [Waring 2012].
chapter 1 it be a truth universally understand that a single man in possession of a good fortune must seed a wife
so when a wealthy gentleman arrive in a neighbourhood it be clear that he must soon become the property of someone daughter
my dear Mr. Bennet say he wife to he one day have you hear that someone have rent the house at Netherfield Park at last
Mr. Bennet reply that he have not
yes she continue Mrs. Long have just be here and she tell I all about it
Mr. Bennet do not answer
do you not want to know who have take it
cry he wife impatiently
you want to tell I and I have no objection to hear it
well Mrs. Long say that Netherfield have be take by a young man of large fortune from the north of England
what be he name
Bingley
be he marry or single
oh
single my dear
a single man of large fortune four or five thousand pound a year
what a fine thing for we girl

however little known the feeling or view of such a man may be on he first enter a neighbourhood this truth be so well fix in the mind of the surround family that he be consider the rightful property of some one or other of they daughter
my dear Mr. Bennet say he lady to he one day have you hear that Netherfield Park be let at last
Mr. Bennet reply that he have not
but it be return she for Mrs. Long have just be here and she tell I all about it
Mr. Bennet make no answer
do you not want to know who have take it
cry he wife impatiently
you want to tell I and I have no objection to hear it
this be invitation enough
why my dear you must know Mrs. Long say that Netherfield be take by a young man of large fortune from the north of England that he come down on Monday in a chaise and four to see the place and be so much delighted with it that he agree with Mr. Morris immediately that he be to take possession before Michaelmas and some of he servant be to be in the house by the end of next week
what be he name
Bingley
be he marry or single
oh
To computationally analyse the process Y and classifying the changes:

- Do the changes follow strict rules?
- Do they form patterns?
- Can they be computationally reproduced?

Categories of changes:

- Cognitive
- Structural
- Cognitive and structural
Readability tests aim to classify texts by their degree of complexity and understandability. Measured primitives are sentence length and difficulty of the words.

Two tests, the ARI score and the Dale-Chall-Index have been selected:

The ARI score is based on the word length and the sentence length:

\[
R_{ARI} = 4.71 \left( \frac{\text{characters}}{\text{words}} \right) + 0.5 \left( \frac{\text{words}}{\text{sentences}} \right) - 21.43
\]

(1)

The Dale-Chall-Index is based on the word frequency (3000 most frequent words) and the sentence length:

\[
R_{DCI} = 0.1579 \left( \frac{\text{difficult words}}{\text{words}} \times 100 \right) + 0.0496 \left( \frac{\text{words}}{\text{sentences}} \right)
\]

(2)
RESULTS OF THE SIMPLIFICATION WITH READABILITY TESTS

Readability test result matrix:

<table>
<thead>
<tr>
<th></th>
<th>ARI</th>
<th>Dale-Chall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original Novel</td>
<td>14-15 year olds</td>
<td>14-16 year olds</td>
</tr>
<tr>
<td>Graded Reader</td>
<td>11-12 year olds</td>
<td>11-13 year olds</td>
</tr>
</tbody>
</table>
An example of a structural text simplification > many-to-one.
COMPARISON OF WORD LENGTH

Word length distribution

Original text
Graded reader

Length of word

Probability

0  5  10  15  20

0  0.05  0.1  0.15  0.2  0.25

Graded reader

0

Original text
The Dotplot view of original novel against the graded reader on a sentence-wise segmentation uncovers which passages were taken over in the graded reader and which not:
The theme this presentation is based on is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License. Changes to the theme are the work of eTRAP.