The interface between readability and automatic text simplification: identifying difficulties to support simple writing

Thomas François
Assistant Professor at UCLouvain

Invited talk at ATA-18
8 November, 2018
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
5. General conclusion
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
   - Introduction to the CEFRLex project
   - AMesure
5. General conclusion
Current needs for reading

- 18.9% of adults from 33 countries do not have the minimal functioning skills in reading for professional purposes (OCDE report) ([http://www.oecd.org/fr/els/l-importance-des-competences-9789264259492-fr.htm](http://www.oecd.org/fr/els/l-importance-des-competences-9789264259492-fr.htm)).

- In parallel, the place of language in professional contexts has substantially increased [Boutet, 2001]

- In everyday life, we are more and more confronted with specialized texts:
  - medical information in case of chronic pathology (Grabar et al., 2018)
  - administrative texts for various purposes
  - technological knowledge is ubiquitous
Current research efforts

- **Readability**: methods to automatically assess the **global** reading difficulty of a text for a given **reader population**.

- **Automatic text simplification**: methods to automatically adapt the style of a text in order to make it more accessible for reading, while preserving most of the meaning.

- **Simple writing aids**: methods to help writers to express themselves in a more effective and more accessible style.

These fields have reach different points in their development...
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
   - Introduction to the CEFRLex project
   - AMesure
5. General conclusion
A typical classic readability formula

[Flesch, 1948] :

\[
\text{Reading Ease} = 206.835 - 0.846 \times \text{wl} - 1,015 \times \text{sl}
\]

where :

- **Reading Ease (RE)**: a score between 0 and 100 (a text for which a 4th grade schoolchild would get 75% of correct answers to a comprehension test)
- \( \text{wl} \): number of syllables per 100 words
- \( \text{sl} \): mean number of words per sentence.

- Classic formulas made use of linear regression and a few linguistic **surface** aspects.
- Vocabulary difficulty is assessed in a simple way (syllables, list, letters, etc.)
Main periods in readability

5 major periods in readability :

1. **The origins**: first works in the field. A lot of interesting perspectives, often forgotten in the current studies!

2. **Classic period**: formulas are based on linear regression and mostly use two indices (one lexical, one syntactic)

3. **The cloze test era**: concerns arise about motivated features (= cause of difficulty) and difficulty measurement

4. **Structuro-cognitivist period**: takes into account newly discovered textual dimensions (cohesion, structure, inference load, etc.).
   → Period of strong criticisms against the classical formulas

5. **AI readability**: NLP-enabled features are combined with more complex statistical algorithms.
   → Main success: integrating both classic and structuro-cognitivist approaches.
What are the uses for readability formulas?

Readability formula have been used for:

- Selection of materials for textbooks.
- Used in scientific experiments to control the difficulty of textual input data.
- Controlling the difficulty level of publications from various administrations (justice, army, etc..) and newspapers.
- More recently, checking the output of automatic summarization, machine translation, etc. [Antoniadis and Grusson, 1996, Aluisio et al., 2010, Kanungo and Orr, 2009].
- Assessing automatic text simplification systems [Štajner and Saggion, 2013, Woodsend and Lapata, 2011, Zhu et al., 2010]

Contexts where having a single value indicator is fine!
Shortcomings of this single score: lost information

- Producing this estimate of the reading difficulty of a text is costly:
  - [Feng et al., 2010]: compute 273 features, 28 retained
  - [François and Fairon, 2012]: compute 406 features, 46 retained
  - [Vajjala and Meurers, 2012]: compute 46 features
  - ...

- Use of a lemmatiser, a tagger and/or a syntactic parser

- The results of these analyses is forgotten (ex. word frequency, complex syntactic structures, etc.)
Shortcomings of this single score: reliability

- Readability assumes that we know which texts are more difficult than other...
  → what means “difficult”? How can we measured it?
- It is measured through another variable, easier to measure and correlated with difficulty
  → we call it the criterion!
- Several criteria have been used in readability ... none are perfect!
  → Human judgments, textbooks, comprehension tests, cloze test, reading speed, eye-tracking data, etc.
Expert judgments

**Pros and cons**

**Pros**: supposedly reliable, rather convenient (no subjects)

**Cons**: population is not directly tested

→ we model the experts’ view of difficulty for the given population

**Issue of heterogeneity**

- [van Oosten et al., 2011] had 105 texts assessed by experts (as pairs) and clustered them by similarity of judgements (train one model per cluster).
  → this leads to different models, whose intracluster performance > intercluster.

- [François et al., 2014a] had 18 experts annotate 105 administrative texts (with an annotation guide)
  → $0.10 < \alpha < 0.61$ per batch (average = 0.37).

- High agreement seems difficult to reach in readability (SemEval 2012 : $\kappa = 0.398$ on the test set).
Using textbooks

Pros and cons

**Pros**: very convenient (no subjects and no experts!)
→ most popular criterion in AI readability, due to the large training corpus needed

**Cons**: population is not directly tested, heterogeneity of the annotations

- Very few corpora available: Weekly Reader is mostly used
  [Schwarm and Ostendorf, 2005, Feng et al., 2010, Vajjala and Meurers, 2012]
  → **risk**: high dependence towards one training corpus, as McCall and Crabbs lessons in classic period [Stevens, 1980]

- This dependence has consequences:
  - formulas will be specialized towards this corpus (coefficients)
  - always the same population and type of texts considered

- Problem of heterogeneity between textbook series [François, 2014]
### Example of heterogeneity in a corpus

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>/</td>
<td>/</td>
<td>-746</td>
<td>-763</td>
<td>-766</td>
<td>-787</td>
</tr>
<tr>
<td>-705</td>
<td>-723</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>/</td>
<td>/</td>
<td>-749</td>
<td>-757</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>-690</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>-758</td>
<td>-766</td>
<td>-777</td>
</tr>
<tr>
<td>-694</td>
<td>/</td>
<td>-746</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>-725</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>-696</td>
<td>-730</td>
<td>-753</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>-731</td>
<td>-742</td>
<td>-733</td>
<td>-766</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>-787</td>
<td>-778</td>
</tr>
<tr>
<td>-664</td>
<td>-712</td>
<td>-756</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>-711</td>
<td>-740</td>
<td>-752</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>-683</td>
<td>-740</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td>-700.09</td>
<td>-732.9</td>
<td>-750.75</td>
<td>-763.52</td>
<td>-771</td>
<td>-779</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion about criterion

- No optimal criterion!
- Recent investigation of:
  - eye-tracking [Singh et al., 2016, Bingel et al., 2018a]
  - crowdsourcing [De Clercq et al., 2014]
- Criterion (and corpus) is probably the factor that impact the most readability formulas performance (difficult to compare all work)
Other issue: generalization of algorithms

"New" Corpus

Corpus readers A1 to B2 !!!
Corpus textbooks C1 and C2
NB: sampling is different
Sampling (48/lev.)
Taking out outliers

New Model

SVM(38) → R = 0.845; acc = 58.2%; adj. acc. = 87.7%

SVM(41) → R = 0.85; acc. = 56%; adj. acc. = 88%

SVM(41) → R = 0.72; acc. = 38%; adj. acc. = 81.7%

SVM(41) → R = 0.72; acc. = 48%; adj. acc. = 77.9%

Old Corpus

Textbooks (68/level)

Old Model

Not available: meanNGProbG, NCPW, NAColl
Now constant: Infi (1) and med_nbNeighMoreFreq (0)
Effect of the genre

[Nelson et al., 2012] distinguishes between performance of various famous models on narrative and informative texts.
The curse of readability

- Averaging means more noisy data

**Reading test**

**Aggregating data**
Conclusions on readability

- Readability methods produce a single difficulty value whose reliability is subject to caution.
- Probably more efficient to model the reading ability at the individual level (personalised models).
- No much work using neural networks or deep learning so far [Liu et al., 2017] → difficult to combine both!
- Try to better use of the information collected to highlight difficult linguistic phenomena in texts.
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
   - Introduction to the CEFRLex project
   - AMesure
5. General conclusion
Is simplifying useful?

We take for granted that simplifying helps the reading process. What are the evidence?

- Manual simplification seems to help.

- [Rello et al., 2013]: using shorter words causes readers with dyslexia to read significantly faster and significantly increased their comprehension. → no effect on control group, though!

- [Gala and Ziegler, 2016]: simplified texts are read significantly faster and with fewer (speech) errors by dyslexic children. → no significant effect on comprehension

Still true: “reading comprehension can be improved for reader with poor literacy by” manually simplifying texts [Siddharthan, 2014]
Is simplifying useful?

What about automatic simplification?

- “a summary helps, but the information highlighted in texts do not” [Margarido et al., 2008]
  → only percentages; no effect computed

- Use of Lickaert scales
  [Wubben et al., 2012, Woodsend and Lapata, 2011]
  → no effect computed

Still too much evidence for ATS!
The lack of theoretical grounds

- Not clear what should be simplified!
- Good synthesis by [Siddharthan, 2014]:
  - lexical and simplification revision (L'Allier, 1980)
  - making discourse relations explicit (Beck et al., 1991)
  - reformulating causal relations (Linderholm et al., 2000)
- No full-fledged psycholinguistic model about complexity to rank words/syntactic structures/... [Gala et al., 2018]
- Current approaches are based on data... and are as good as their data (Wikipedia, Newsela)
Current challenges

- Goes beyond the lexical and syntactic levels!
- Drawing from readability, investigate adaptative approach to take into account the variety of texts and readers [Bingel et al., 2018b]
- Rely more on psycholinguistics evidences to motivate simplifications.
- Develop more parallel data, especially for languages other than English
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
   - Introduction to the CEFRLex project
   - AMesure
5. General conclusion
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
   - Introduction to the CEFRLex project
   - AMesure
5. General conclusion
Objectives of the CEFRLex project

To offer lexical resources describing word distributions in textbooks across the 6 CEFR levels.

Possible uses:

- Targeted vocabulary learning (which word to learn at which level)
- Comparing the frequency of usage of synonyms
- Using it within a language model for various iCALL tasks (readability, etc.)
- Apply it for automatic text simplification (ATS)
The CEFRLex project: current projects

FLELex (French L2)
- Available at http://cental.uclouvain.be/flelex/
- Publication: [François et al., 2014b]
- Team: Thomas François, Núria Gala, Anaïs Tack, Patrick Watrin, Cédrick Fairon

EFLLex (English L2)
- Available at http://cental.uclouvain.be/cefrlex/
- Publication: [Dürlich and François, 2018]
- Team: Thomas François, Luise Dürlich
The CEFRLex project: current projects

**SVALex (Swedish L2 - reception)**
- Publication: [François et al., 2016]
- Team: Thomas François, Elena Volodina, Ildikó Pilán, Anaïs Tack

**SweLLex (Swedish L2 - production)**
- Publication: [Volodina et al., 2016]
- Team: Elena Volodina, Ildikó Pilán, Lorena Llozhi, Baptiste Degryse and Thomas François
The CEFRLex project: current projects

**NT2Lex - Dutch L2**
- Publication: [Tack et al., 2018]
- Team: Anaïs Tack, Thomas François, Piet Desmet, Cédrick Fairon

**ELELex - Spanish L2**
- Not yet available (publication in preparation)
- Publication: scheduled for 2019
- Team: Thomas François, Barbara De Cock, Irwing Palacios
Common methodology

1. Collect a corpus of texts intended for L2 learners (from textbooks or simplified readers)
   → The texts must be labelled with a CEFR level

2. Find the lemma and the part-of-speech tag of each word in the corpus
   → Issue: what is a word? MWE!

3. Estimate the frequency distribution of each lemma using a robust estimator
   → dispersion index [Carroll et al., 1971] to normalize frequencies

4. Iterative process: manual postprocessing of the resource to correct NLP errors precedes a new frequency estimation step
Example: entries from EFLLex and ELELex

<table>
<thead>
<tr>
<th>lemma</th>
<th>tag</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>NN</td>
<td>2.940</td>
<td>202.796</td>
<td>31.681</td>
<td>33.339</td>
<td>28.9847</td>
<td>65.019</td>
</tr>
<tr>
<td>empty</td>
<td>JJ</td>
<td>86.492</td>
<td>150.888</td>
<td>65.947</td>
<td>194.801</td>
<td>123.405</td>
<td>156.021</td>
</tr>
<tr>
<td>explore</td>
<td>VB</td>
<td>20.578</td>
<td>54.677</td>
<td>73.625</td>
<td>46.070</td>
<td>36.665</td>
<td>56.961</td>
</tr>
<tr>
<td>obviously</td>
<td>RB</td>
<td>0</td>
<td>11.034</td>
<td>2.589</td>
<td>0.315</td>
<td>0.815</td>
<td>0.611</td>
</tr>
<tr>
<td>tiresome</td>
<td>JJ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>video</td>
<td>NN</td>
<td>2.467</td>
<td>0.556</td>
<td>34.825</td>
<td>23.802</td>
<td>13.248</td>
<td>18.431</td>
</tr>
<tr>
<td>write</td>
<td>VB</td>
<td>934.708</td>
<td>378.337</td>
<td>760.734</td>
<td>536.380</td>
<td>713.326</td>
<td>549.909</td>
</tr>
<tr>
<td>shopping centre</td>
<td>NN</td>
<td>0</td>
<td>5.040</td>
<td>2.589</td>
<td>0</td>
<td>0.815</td>
<td>1.946</td>
</tr>
<tr>
<td>sign up</td>
<td>VB</td>
<td>0</td>
<td>0.887</td>
<td>10.789</td>
<td>2.499</td>
<td>6.216</td>
<td>5.302</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>lemma</th>
<th>tag</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>incluir</td>
<td>VM</td>
<td>5.4</td>
<td>60.35</td>
<td>31.3</td>
<td>90.4</td>
<td>258.04</td>
<td>74.1</td>
</tr>
<tr>
<td>llamada</td>
<td>NCF</td>
<td>9.6</td>
<td>45.3</td>
<td>56</td>
<td>40.8</td>
<td>9.1</td>
<td>44.6</td>
</tr>
<tr>
<td>monumental</td>
<td>AQ0</td>
<td>0</td>
<td>0</td>
<td>1.3</td>
<td>2</td>
<td>0</td>
<td>1.14</td>
</tr>
<tr>
<td>malententido</td>
<td>NCM</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
<td>0.04</td>
</tr>
<tr>
<td>acerca de</td>
<td>SP</td>
<td>0</td>
<td>0</td>
<td>4.9</td>
<td>21.4</td>
<td>18.3</td>
<td>8.2</td>
</tr>
<tr>
<td>al fin y al cabo</td>
<td>RG</td>
<td>0</td>
<td>0</td>
<td>0.4</td>
<td>15.5</td>
<td>10.2</td>
<td>4.1</td>
</tr>
</tbody>
</table>
Demonstration

Make a query in EFLLex

Enter a word
book  reserve  Search

Frequencies by CEFR levels for the words book and reserve.

![Bar chart showing frequencies by CEFR levels for book and reserve.](chart.png)
A few figures about the resources

<table>
<thead>
<tr>
<th>Resource</th>
<th># entries</th>
<th># Hapax</th>
<th># MWE</th>
<th>r with list</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLELex-TT</td>
<td>14,236</td>
<td>4,697</td>
<td>2,038</td>
<td>Lexique3 (0.84)</td>
</tr>
<tr>
<td>EFLLex</td>
<td>15,280</td>
<td>/</td>
<td>/</td>
<td>BNC (0.97)</td>
</tr>
<tr>
<td>SVALex</td>
<td>15,681</td>
<td>/</td>
<td>1,450</td>
<td>/</td>
</tr>
<tr>
<td>ELELex</td>
<td>24,804</td>
<td>/</td>
<td>5,456</td>
<td>/</td>
</tr>
<tr>
<td>NT2Lex-CGN</td>
<td>15,227</td>
<td>/</td>
<td>459</td>
<td>Subtlex-NL (0.69)</td>
</tr>
</tbody>
</table>

NB: in NT2Lex-CGN, 4,431 (31%) of the single-word entries are compounds
CEFRLex project: Assets

- Allow to discriminate words within a level:

<table>
<thead>
<tr>
<th>Word</th>
<th>Pos</th>
<th>A1</th>
<th>A2</th>
<th>B1</th>
<th>B2</th>
<th>C1</th>
</tr>
</thead>
<tbody>
<tr>
<td>about</td>
<td>RB</td>
<td>2014,8</td>
<td>1210,2</td>
<td>984,4</td>
<td>483,9</td>
<td>238,2</td>
</tr>
<tr>
<td>(to) camp</td>
<td>VB</td>
<td>38,7</td>
<td>26,4</td>
<td>11,7</td>
<td>12,3</td>
<td>0</td>
</tr>
</tbody>
</table>

- Frequency estimation from authentic pedagogical documents

- Disambiguation as regards POS-tags and/or senses
  [Tack et al., 2018]
FLELex relates words to a given CEFR level

Analyse a text with FLELex

With FLELex, it is possible to analyse the lexical complexity of a French text for a specific CEFR proficiency level. All you need to do is introduce a text of your choice and we'll do the analysis for you. For additional tips and tricks on how to interpret the analysis, please consult the "How-to" tab below.

Lexical complexity for level A2

La présidente, nouvellement élue, demande l'abolition de la taxe sur le capital.
From distribution to level

**Problem**: How to transform a distribution into a single level?

**Example**: the distribution of *capital*

... is transformed into B1

*Frequencies of capital in FLELex-TT.*
Experiment from [Tack et al., 2016a]

- Collect annotations from 4 Flemish learners of FFL (A2 and B1) on 51 short texts → learners report unknown words via a web interface.

- Then, various thresholds (frequency value, quantile) were tested in order to maximize the prediction of unknown words for the 4 subjects.

- Surprisingly, the best discretization function is “first occurrence”!
From distribution to level

Classification accuracy for FLELex prediction with percentiles

- B1-4
- B1-U
- A2-2
- A2-3
Using FLELex as a way to predict lexical knowledge

Rule to decide between known (0) and unknown (1) words:

\[ c = \begin{cases} 
1, & \text{if } l > L \lor \exists! l \\
0, & \text{otherwise}
\end{cases} \]

<table>
<thead>
<tr>
<th>Learner</th>
<th>Lexical words</th>
<th>Grammatical words</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2-2</td>
<td>86.6%</td>
<td>99.2%</td>
<td>89.7%</td>
</tr>
<tr>
<td>A2-3</td>
<td>81.1%</td>
<td>99.2%</td>
<td>87.4%</td>
</tr>
<tr>
<td>B1-4</td>
<td>91.3%</td>
<td>99.7%</td>
<td>92.3%</td>
</tr>
<tr>
<td>B1-U</td>
<td>90.8%</td>
<td>99.8%</td>
<td>92.0%</td>
</tr>
</tbody>
</table>

**Table**: Accuracy for the prediction of the lexical knowledge of the 4 learners using FLELex.
Discussion

- In the interface, predictions appears as overoptimistic (too much A1 words)
- The evaluation seems good, but...
  - The model behave better on known words than on unknown ones (less numerous)
- This is a consequence of the rule “first occurrence”, which appears as too simple!

<table>
<thead>
<tr>
<th></th>
<th>Connu</th>
<th>Inconnu</th>
</tr>
</thead>
<tbody>
<tr>
<td>apprenant A2-2</td>
<td>95.7% (0.92)</td>
<td>4.3% (0.42)</td>
</tr>
<tr>
<td>apprenant A2-3</td>
<td>88.1% (0.94)</td>
<td>11.9% (0.38)</td>
</tr>
<tr>
<td>apprenant B1-4</td>
<td>97.0% (0.94)</td>
<td>3.0% (0.40)</td>
</tr>
<tr>
<td>apprenant B1-U</td>
<td>96.7% (0.94)</td>
<td>3.3% (0.37)</td>
</tr>
</tbody>
</table>

**Table**: Percentage of known and unknown words + recall for predictions based on FLELex [Tack et al., 2016a].
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
   - Introduction to the CEFRLex project
   - AMesure
5. General conclusion
The AMesure project

Administrative texts can be hard to understand:

AMesure is...

- A platform for simple writing of administrative texts, supported by the FWB (Federation Wallonia-Brussels).

- Objectives:
  1. Draw the attention of writers to complex linguistic phenomena;
  2. Provide pieces of advice for manual simplification of administrative texts (based on simple writing guides).
AMesure

AMesure : current situation

- A global readability score (readability formula, in A) [François et al., 2014a]
- Assessment of several linguistics dimensions of the text (B)
- Highlighting complex phenomena in the text (C)
- Suggestions for simple writing for each sentence (D)
Detecting complex phenomena in administrative texts

Currently detected:

- Subordinated clauses:
  - relative clauses
  - object clause (fr. complétive)
  - adverbial clause

- Passive sentence

- Brackets

- Abbreviations (list-based and rules-based)

- Complex words (frequency-based)
Detecting syntactic structures

Implementation based on [Brouwers et al., 2014]

```
Sentences to analyse

Syntactic parsing (Berkeley Parser)

Defining rules (based on a corpus)

Applying regular expressions (via Tregex)
```

En Région wallonne, une taxe annuelle d'un montant de 100 € doit être payée lorsque l'on détient un appareil de télévision, quel que soit l'usage qui en est fait.

"VN < (V\|VINF\|VPP "+etre()\+" $. (VPP "+notVIntransitif()\+"));"
AMeasure

Evaluation

- Results of the previous system (François et al., 2018)
- Test data = 13 administrative texts (319 clauses, 134 passives, 57 abbreviations)

<table>
<thead>
<tr>
<th>Phenomena</th>
<th>Precision</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>all subordinated clauses</td>
<td>0.89</td>
<td>0.88</td>
</tr>
<tr>
<td>relative clauses</td>
<td>0.98</td>
<td>0.85</td>
</tr>
<tr>
<td>object clauses</td>
<td>0.6</td>
<td>0.67</td>
</tr>
<tr>
<td>adverbial clauses</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td>passives</td>
<td>0.9</td>
<td>0.92</td>
</tr>
<tr>
<td>abbreviations</td>
<td>0.57</td>
<td>0.4</td>
</tr>
</tbody>
</table>
Generating the advices

- Theoretical reference = simple writing guides by the administrations...
- 7 cases have been implemented

<table>
<thead>
<tr>
<th>Problem</th>
<th>Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of nested syntactic structures</td>
<td>≥ 3</td>
</tr>
<tr>
<td>total number of clauses</td>
<td>&gt; 3</td>
</tr>
<tr>
<td>length of the sentence</td>
<td>&gt; 15 words</td>
</tr>
<tr>
<td>length of the longest nested clause</td>
<td>&gt; 10 words</td>
</tr>
<tr>
<td>length of text between brackets</td>
<td>&gt; 10 words</td>
</tr>
<tr>
<td>number of subordinated clauses</td>
<td>≥ 3</td>
</tr>
</tbody>
</table>
Le rôle est un document fiscal global qui reprend le nom de l'ensemble des redevables ainsi que le montant de l'impôt ou de la taxe dont ils sont redevables. En matière de redevance télévision, vous êtes tenu par la loi, en tant que redevable, de payer la redevance dans le délai fixé par l'invitation à payer qui vous est adressée par l'administration fiscale wallonne (Direction générale opérationnelle de la Fiscalité). Si vous n'effectuez pas le paiement réclamé par cette invitation, le montant dû est alors enclenché et un avertissement-extrait de rôle (qui est donc un extrait individuel du rôle vous concernant personnellement) vous sera alors envoyé. Le montant de redevance repris sur cet avertissement-extrait de rôle est considéré, au regard de la loi, comme une dette certaine au profit du trésor public, qui doit être payée immédiatement. Attention: L'Administration fiscale n'adresse pas de rappel après envoi de cet avertissement-extrait de rôle. En cas de non paiement, votre dossier sera donc transmis chez un huissier de justice et votre dette initiale sera donc augmentée de frais supplémentaires. C'est pourquoi, en cas de réception d'un tel document, il vous est instamment conseillé de verser sans délai la somme réclamée. Si vous contestez cette taxation, vous pouvez introduire une réclamation par écrit dans les 6 mois selon les formes précisées sur l'avertissement extrait de rôle et explicitées dans le feuillet explicatif qui y est joint. À noter que l'introduction d'une éventuelle demande d'exonération n'est pas suspensive ni de l'obligation du paiement, ni de l'introduction d'une réclamation administrative.

— Extrait pris du site: http://www.wallonie.be/fr/redevance-impayee-et-facilities-de-paiement
Next version in preparation

- Diagnosis organised in three main dimensions (lexical, syntactic, discursive)
- Further investigation of abbreviations:
  - density of abbreviations
  - detecting abbreviations without an extended forms in the text
  - detecting technical terms
- Global ratio of complex structures for the text
- Suggest simpler synonyms using ReSyf [Billami et al., 2018]

Soon out!
Plan

1. Introduction: text accessibility
2. Is assessing text readability possible?
3. The current challenges of text simplification
4. Highlighting difficulties
   - Introduction to the CEFRLex project
   - AMesure
5. General conclusion
Conclusions

- Highlighting reading difficulties is a simpler task than ATS and a more meaningful one than readability
  → It is also easier to test on humans

- Suffer from the same lack of theoretical definition of complexity
  → Allows to test independently each category of phenomena

- Adaptation is also required (Yimam and Biemann, 2018)
  → Personalising the models seems to account for a larger part of the variance
  [Tack et al., 2016b]
Thank you for your attention
References I


References II


References III


References VI


References VII


References IX


A posteriori agreement as a quality measure for readability prediction systems.

Swellex: second language learners’ productive vocabulary.
In *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition at SLTC, Umeå*, number 130, pages 76–84. Linköping University Electronic Press.

Learning to simplify sentences with quasi-synchronous grammar and integer programming.
References XI
