Is Automatic Text Simplification Useful?

Advaith Siddharthan

University of Aberdeen
Scotland
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
    - The words in the text reflect the complexity of the topic
    - Replace difficult words with simpler synonyms
    - Replace technical jargon with definitions
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
  - Syntax

  Longer sentences with embedded clauses are more difficult to read than shorter sentences

- Split complicated multi-clause sentences into simpler ones
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
  - Syntax
  - Text Length

  Longer texts require readers to have more stamina and a greater capacity to remember details

- Reduce text length by deleting peripheral information
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
  - Syntax
  - Text Length

  Texts with key information repeated in different ways is more likely to result in retention of information

- Increase text length by adding redundancy (through reformulation, analogy, metaphor, examples...)

---

<table>
<thead>
<tr>
<th>What is Text Simplification</th>
<th>Evaluating Text Simplification</th>
<th>My Personal Experience</th>
<th>Challenges for QATS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is Text Simplification?

• TS aims to make information/meaning more accessible through textual reformulation
  • Lexis
  • Syntax
  • Text Length
  • Discourse

Chronological or cause/effect ordering of sentences is easier to follow than more sophisticated structures - stories within stories, flashbacks, parallel plots, etc.

• Reorder information to simplify argumentation
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
  - Syntax
  - Text Length
  - Discourse

  **Cohesive texts are easier to follow**
  - Make content more transparent by making discourse relations explicit
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
  - Syntax
  - Text Length
  - Discourse
  - Semantics

  The readers’ background knowledge affects their ability to read and comprehend texts
  - Conceptual simplification
  - Numerical simplification
  - Simplify meaning (dumbing down)
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
  - Syntax
  - Text Length
  - Discourse
  - Semantics
  - Quality

  Errorful text is harder to read
  - Check spelling and grammar
What is Text Simplification?

- TS aims to make information/meaning more accessible through textual reformulation
  - Lexis
  - Syntax
  - Text Length
  - Discourse
  - Semantics
  - Quality
  - Style

More engaging text is easier to read
- Paint a picture (use visual words)
- Personal narratives
- Humour
Manual Text Simplification

- Explored extensively with regard to (typically, middle) school performance.
Manual Text Simplification

- Explored extensively with regard to (typically, middle) school performance.
  - L'Allier (1980): lexical and syntactic text revision brings low ability readers above the performance level of middle ability readers on the original text.
Manual Text Simplification

- Explored extensively with regard to (typically, middle) school performance.
  - L'Allier (1980): lexical and syntactic text revision brings low ability readers above the performance level of middle ability readers on the original text.
Manual Text Simplification

- Explored extensively with regard to (typically, middle) school performance.
  - L'Allier (1980): lexical and syntactic text revision brings low ability readers above the performance level of middle ability readers on the original text.
  - Linderholm et al. (2000): Reformulating causal relations for relatively difficult texts facilitates poor readers.
Manual Text Simplification

- Explored extensively with regard to (typically, middle) school performance.
  - L'Allier (1980): lexical and syntactic text revision brings low ability readers above the performance level of middle ability readers on the original text.
  - Linderholm et al. (2000): Reformulating causal relations for relatively difficult texts facilitates poor readers.
Manual Text Simplification

- Explored extensively with regard to (typically, middle) school performance.
  - Similar results have been found for readers with low levels of domain expertise (Noordman & Vonk, 1992; Kamalski et al., 2008; McNamara et al., 1996).
Manual Text Simplification

- Explored extensively with regard to (typically, middle) school performance.
  - Similar results have been found for readers with low levels of domain expertise (Noordman & Vonk, 1992; Kamalski et al., 2008; McNamara et al., 1996).
  - Similar results have been found for L2 (Long & Ross, 1993; Yano et al., 1994; Tweissi, 1998; Gardner & Hansen, 2007).
Evaluating manual text simplification

Typically comprehension of an entire text is tested on target populations:

- Immediate or delayed free recall of passage
- Multiple choice or short-answer comprehension questions
Evaluating manual text simplification

Typically comprehension of an entire text is tested on target populations:

- Immediate or delayed free recall of passage
- Multiple choice or short-answer comprehension questions

*Quality is assumed; Evaluations focus on effect on target populations*
Evaluating manual text simplification

Typically comprehension of an entire text is tested on target populations:

- Immediate or delayed free recall of passage
- Multiple choice or short-answer comprehension questions

*Evaluation of entire texts; Little interest in evaluating sentence-level simplification*
Evaluating manual text simplification

Typically comprehension of an entire text is tested on target populations:

- Immediate or delayed free recall of passage
- Multiple choice or short-answer comprehension questions

This is the evidence that TS can facilitate readers
Some criticisms of (manual) text simplification

Honeyfield (1977); Long & Ross (1993); Yano et al. (1994); Oh (2001):

- Can impede language acquisition (good for short term, but bad in the long run)
- Does not work, even in the short term!
  - Homogenises vocabulary across the text, and makes important information harder to identify
  - Simplified texts are not as interesting
  - Not really simpler (e.g. use of readability formulae as guides to writing)
Some criticisms of (manual) text simplification

Honeyfield (1977); Long & Ross (1993); Yano et al. (1994); Oh (2001):

- Can impede language acquisition (good for short term, but bad in the long run)
- Does not work, even in the short term!
  - Homogenises vocabulary across the text, and makes important information harder to identify
  - Simplified texts are not as interesting
  - Not really simpler (e.g. use of readability formulae as guides to writing)

*But, manually simplified texts are widely used in educational contexts, especially for L2*
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language.
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language.

- If we have a good handle on what these difficult constructs are, text complexity (and hence complexity reduction through ATS) can be evaluated as a counting exercise.
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language.
- There is a long literature on measuring text complexity, starting with readability metrics, but also substantially more sophisticated methods.
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language.

- There is a long literature on measuring text complexity, starting with readability metrics, but also substantially more sophisticated methods.
  - Combine various heuristics for complexity.
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language
- There is a long literature on measuring text complexity, starting with readability metrics, but also substantially more sophisticated methods
  - Combine various heuristics for complexity
  - Generally, these are not very informative for evaluating ATS
    - They assume error-free text
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language.
- There is a long literature on measuring text complexity, starting with readability metrics, but also substantially more sophisticated methods.
  - Combine various heuristics for complexity.
  - Generally, these are not very informative for evaluating ATS.
    - They assume error-free text.
  - Errors add to complexity.
  - ATS struggles to produce error-free text.
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language.
- Instead, evaluations of ATS have tended to focus on Quality.
Evaluating Automatic Text Simplification

- ATS aims to reduce text complexity - typically by reducing the likelihood of encountering the more difficult constructs permitted by a language.

- Instead, evaluations of ATS have tended to focus on Quality.

- Metrics from MT evaluation (Bleu, etc) begin to get at fluency through n-gram overlap.

  - But, a single error can make a sentence very difficult to read for a typical consumer of simplified language. Errors can be generated to bad word choice, or misparsing Siddharthan (2003b) or discourse level problems Siddharthan (2003a).
## Translation vs Simplification

<table>
<thead>
<tr>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
</tr>
<tr>
<td></td>
<td>Native</td>
</tr>
</tbody>
</table>

### What is Text Simplification

### Evaluating Text Simplification

### My Personal Experience

### Challenges for QATS
## Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
<td>??</td>
</tr>
</tbody>
</table>
## Translation vs Simplification

<table>
<thead>
<tr>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
</tr>
</tbody>
</table>

**EW:** Four generations of the Willis family continued the family tradition of organ building until 1997 when Henry Willis 4 retired, and the first non-family Managing Director was appointed.

**SEW:** Four generations of the Willis family continued the family tradition of organ building until 1997, when Henry Willis the fourth retired and the new Managing Director of the firm was not a member of the Willis family.
Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
<td>???</td>
</tr>
</tbody>
</table>

**EW:** Four generations of the Willis family continued the family tradition of organ building until 1997 when Henry Willis 4 retired, and the first non-family Managing Director was appointed.

**SEW:** Four generations of the Willis family continued the family tradition of organ building until 1997, when Henry Willis the fourth retired and the new Managing Director of the firm was not a member of the Willis family.

**Siddharthan & Angrosh (2014):** Four generations of the Willis family continued the family tradition of organ building until 1997. Then, Henry Willis 4 retired.

And the first non-family Managing Director was given.
## Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Translator</strong></td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td><strong>Simplifier</strong></td>
<td>?</td>
<td>???</td>
</tr>
<tr>
<td><strong>Translation User</strong></td>
<td>Unfamiliar</td>
<td>Fluent</td>
</tr>
</tbody>
</table>
# Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
<td>???</td>
</tr>
<tr>
<td>Translation User</td>
<td>Unfamiliar</td>
<td>Fluent</td>
</tr>
</tbody>
</table>

Fluency in Target language helps handle errors in translation
Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
<td>??</td>
</tr>
<tr>
<td>Translation User</td>
<td>Unfamiliar</td>
<td>Fluent</td>
</tr>
<tr>
<td>Simplification User</td>
<td>?</td>
<td>??</td>
</tr>
</tbody>
</table>
## Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
<td>???</td>
</tr>
<tr>
<td>Translation User</td>
<td>Unfamiliar</td>
<td>Fluent</td>
</tr>
<tr>
<td>Simplification User</td>
<td>?</td>
<td>???</td>
</tr>
</tbody>
</table>

Typical user has low tolerance to errors in simplification
## Translation vs Simplification

<table>
<thead>
<tr>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
</tr>
<tr>
<td>Translation User</td>
<td>Unfamiliar</td>
</tr>
<tr>
<td>Simplification User</td>
<td>?</td>
</tr>
</tbody>
</table>

*Text Simplification is harder than Machine Translation*
## Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
<td>???</td>
</tr>
<tr>
<td>Translation User</td>
<td>Unfamiliar</td>
<td>Fluent</td>
</tr>
<tr>
<td>Simplification User</td>
<td>?</td>
<td>???</td>
</tr>
</tbody>
</table>

**Text Simplification is harder than Machine Translation**

- *MT users can’t read the source text, TS users can*
## Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>??</td>
<td>???</td>
</tr>
<tr>
<td>Translation User</td>
<td>Unfamiliar</td>
<td>Fluent</td>
</tr>
<tr>
<td>Simplification User</td>
<td>??</td>
<td>???</td>
</tr>
</tbody>
</table>

**Text Simplification is harder than Machine Translation**

- *MT users can’t read the source text, TS users can*
- *TS users are not fluent in target language, so quality requirements are higher*
# Translation vs Simplification

<table>
<thead>
<tr>
<th></th>
<th>Source Language</th>
<th>Target Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translator</td>
<td>Fluent</td>
<td>Native</td>
</tr>
<tr>
<td>Simplifier</td>
<td>?</td>
<td>??</td>
</tr>
<tr>
<td>Translation User</td>
<td>Unfamiliar</td>
<td>Fluent</td>
</tr>
<tr>
<td>Simplification User</td>
<td>?</td>
<td>??</td>
</tr>
</tbody>
</table>

**Text Simplification is harder than Machine Translation**

- *MT users can’t read the source text, TS users can*
- *TS users are not fluent in target language, so quality requirements are higher*

**Quality assessment of TS is also harder,**

*we can’t reuse MT metrics blindly*
Evaluating Text Simplification

- *How accurate does automatic text simplification need to be?*
  - The typical target reader of a text simplification system has poor reading skills.
  - Errorful system output might be unusable, even when it could be understood by a fluent reader.
Evaluating Text Simplification

- *How accurate does automatic text simplification need to be?*
  - The typical target reader of a text simplification system has poor reading skills.
  - Errorful system output might be unusable, even when it could be understood by a fluent reader.

*Key Evaluation Question:* Does ATS facilitate users relative to original text?
Evaluating Text Simplification

How are text simplification systems evaluated?
Evaluating Text Simplification

How are text simplification systems evaluated?

- Automated evaluations:
  - Compare system output to manually simplified text (BLEU or NIST scores) (Coster & Kauchak, 2011)
  - Intrinsic Readability scores (Louis & Nenkova, 2013)
- None of these account properly for errors that increase comprehension difficulty, or for unintentional meaning change.
Evaluating Text Simplification

- **How are text simplification systems evaluated?**

- **Automated evaluations:**
  - Compare system output to manually simplified text (BLEU or NIST scores) *(Coster & Kauchak, 2011)*
  - Intrinsic Readability scores *(Louis & Nenkova, 2013)*
  - None of these account properly for errors that increase comprehension difficulty, or for unintentional meaning change.

- **Ratings by fluent readers:**
  - Fluency, Simplicity, Meaning *(Siddharthan, 2006; Wubben et al., 2012; Woodsend & Lapata, 2011)*
Evaluating Text Simplification

- *How are text simplification systems evaluated?*

  - **Automated evaluations:**
    - Compare system output to manually simplified text (BLEU or NIST scores) *(Coster & Kauchak, 2011)*
    - Intrinsic Readability scores *(Louis & Nenkova, 2013)*
    - None of these account properly for errors that increase comprehension difficulty, or for unintentional meaning change.

  - **Ratings by fluent readers:**
    - Fluency, Simplicity, Meaning *(Siddharthan, 2006; Wubben et al., 2012; Woodsend & Lapata, 2011)*

  - **Online Methods:**
    - Eye-tracking: Fixation times etc. *(Bott et al., 2012)*
Evaluating Text Simplification

- **How are text simplification systems evaluated?**

  - Automated evaluations:
    - Compare system output to manually simplified text (BLEU or NIST scores) *(Coster & Kauchak, 2011)*
    - Intrinsic Readability scores *(Louis & Nenkova, 2013)*
    - None of these account properly for errors that increase comprehension difficulty, or for unintentional meaning change.

  - Ratings by fluent readers:
    - Fluency, Simplicity, Meaning *(Siddharthan, 2006; Wubben et al., 2012; Woodsend & Lapata, 2011)*

  - Online Methods:
    - Eye-tracking: Fixation times etc. *(Bott et al., 2012)*

  - Offline Methods for recall / comprehension:
    - Free Recall *(Siddharthan & Katsos, 2012)*
    - MCQs *(Canning, 2002; Angrosh et al., 2014)*
Some aspects I’ve tried to evaluate

- Correctness of simplification operations – binary decision (Siddharthan, 2011):

<table>
<thead>
<tr>
<th>System</th>
<th>Av S Len</th>
<th>#Trans/S</th>
<th>%S Trans</th>
<th>%Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>20.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>15.3</td>
<td>0.65</td>
<td>50</td>
<td>84</td>
</tr>
<tr>
<td>B</td>
<td>14.3</td>
<td>0.74</td>
<td>55</td>
<td>88</td>
</tr>
<tr>
<td>C</td>
<td>14.8</td>
<td>0.65</td>
<td>50</td>
<td>71</td>
</tr>
<tr>
<td>D</td>
<td>14.0</td>
<td>0.74</td>
<td>55</td>
<td>78</td>
</tr>
</tbody>
</table>
Some aspects I’ve tried to evaluate

- Acceptability of output text – binary decision *(Siddharthan, 2011)*:

<table>
<thead>
<tr>
<th>System</th>
<th>Av S Len</th>
<th>#Trans/S</th>
<th>%S Trans</th>
<th>%Correct</th>
<th>%Acceptable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>20.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>15.3</td>
<td>0.65</td>
<td>50</td>
<td>84</td>
<td>63</td>
</tr>
<tr>
<td>B</td>
<td>14.3</td>
<td>0.74</td>
<td>55</td>
<td>88</td>
<td>66</td>
</tr>
<tr>
<td>C</td>
<td>14.8</td>
<td>0.65</td>
<td>50</td>
<td>71</td>
<td>30</td>
</tr>
<tr>
<td>D</td>
<td>14.0</td>
<td>0.74</td>
<td>56</td>
<td>78</td>
<td>33</td>
</tr>
</tbody>
</table>
Some aspects I’ve tried to evaluate

- Magnitude Estimation of Acceptability (*Siddharthan & Katsos, 2010*)
  - Participants rate acceptability of sentence relative to a sentence that is borderline acceptable
  - MagEst scores are standardised as z-scores (mean=0, sd=1)
Some aspects I’ve tried to evaluate

- **Delayed Sentence Recall** *(Siddharthan & Katsos, 2012)*
  - Correlates strongly with Magnitude Estimation for longer sentences
  - For shorter sentences, it is easier to spot mistakes and rate them poorly, but recall does not appear to be impaired at all.
  - When sentence is recalled incorrectly, this method provides insights into the nature of the breakdown (is the first clause recalled or the second, etc.)
Some aspects I’ve tried to evaluate

- Quality Metrics using Likert Scales (Siddharthan, 2006; Siddharthan & Angrosh, 2014, etc.):

<table>
<thead>
<tr>
<th>System</th>
<th>Fluency</th>
<th>Simplicity</th>
<th>Meaning Preservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2.48</td>
<td>2.85</td>
<td>2.92</td>
</tr>
<tr>
<td>B</td>
<td>4.04</td>
<td>3.74</td>
<td>4.60</td>
</tr>
<tr>
<td>SEW</td>
<td>4.15</td>
<td>4.05</td>
<td>3.96</td>
</tr>
</tbody>
</table>

What is Text Simplification

Evaluating Text Simplification

My Personal Experience

Challenges for QATS
Some aspects I’ve tried to evaluate

- Comprehension Tests using MCQ (Angrosh et al., 2014)

<table>
<thead>
<tr>
<th></th>
<th>Orig Man ATS</th>
<th>Orig Man ATS</th>
<th>Orig Man ATS</th>
<th>Orig Man ATS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Language Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>Excellent</td>
<td>Good</td>
<td>Mod</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>.69</td>
<td>.84</td>
<td>.74</td>
<td>.64</td>
</tr>
<tr>
<td></td>
<td>.92</td>
<td>.87</td>
<td>.78</td>
<td>.77</td>
</tr>
<tr>
<td></td>
<td>.94</td>
<td>.80</td>
<td>.80</td>
<td>.72</td>
</tr>
<tr>
<td><strong>Recall Skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accuracy</td>
<td>Excellent</td>
<td>Good</td>
<td>Mod</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>.88</td>
<td>.75</td>
<td>.60</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>.98</td>
<td>.87</td>
<td>.70</td>
<td>.53</td>
</tr>
<tr>
<td></td>
<td>.95</td>
<td>.84</td>
<td>.75</td>
<td>.40</td>
</tr>
</tbody>
</table>
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
- Can AQA correlate with comprehension data, and therefore work with longer texts?
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
- Can AQA correlate with comprehension data, and therefore work with longer texts?
- QATS focuses on lexical and syntactic simplification, because this is what people work on today. Will AQA metrics work for other simplification operations?
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
- Can AQA correlate with comprehension data, and therefore work with longer texts?
- QATS focuses on lexical and syntactic simplification, because this is what people work on today. Will AQA metrics work for other simplification operations?
- Be unambiguous and specific
  - But, simpler / more frequent words have more senses
  - But, simpler / more frequent words are less specific
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
- Can AQA correlate with comprehension data, and therefore work with longer texts?
- QATS focuses on lexical and syntactic simplification, because this is what people work on today. Will AQA metrics work for other simplification operations?
- Be unambiguous and specific
  - But, simpler / more frequent words have more senses
  - But, simpler / more frequent words are less specific
  - In a pilot study, readers showed a preference for less polysemous (but also less frequent) words (Walker et al., 2011)
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
- Can AQA correlate with comprehension data, and therefore work with longer texts?
- QATS focuses on lexical and syntactic simplification, because this is what people work on today. Will AQA metrics work for other simplification operations?
- Be unambiguous and specific
  - But, simpler / more frequent words have more senses
  - But, simpler / more frequent words are less specific
  - In a pilot study, readers showed a preference for less polysemous (but also less frequent) words (Walker et al., 2011)
  - Can AQA measures capture ambiguity effects?
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
- Can AQA correlate with comprehension data, and therefore work with longer texts?
- QATS focuses on lexical and syntactic simplification, because this is what people work on today. Will AQA metrics work for other simplification operations?
- Simplify through elaboration or use of analogy/metaphor
  - Ubiquitous in science texts for children
  - Also emphasised by L2 community
Challenges for Automated Quality Assessment

- All simplification systems result in shorter sentences and/or simpler words. Can AQA distinguish these systems?
- Can AQA correlate with comprehension data, and therefore work with longer texts?
- QATS focuses on lexical and syntactic simplification, because this is what people work on today. Will AQA metrics work for other simplification operations?
- Simplify through elaboration or use of analogy/metaphor
  - Ubiquitous in science texts for children
  - Also emphasised by L2 community
  - Can AQA measures handle elaboration?
Conclusions

- ATS needs automated metrics for evaluation, but AQA do not take errors into account.
Conclusions

- ATS needs automated metrics for evaluation, but AQA do not take errors into account. Until they do, it will be hard to adopt AQA.
**Conclusions**

- ATS needs automated metrics for evaluation, but AQA do not take errors into account. **Until they do, it will be hard to adopt AQA**

- QATS is an excellent start
Conclusions

• ATS needs automated metrics for evaluation, but AQA do not take errors into account.  Until they do, it will be hard to adopt AQA

• QATS is an excellent start  Perhaps in future editions AQA metrics should correlate with comprehension data?
Conclusions

- ATS needs automated metrics for evaluation, but AQA do not take errors into account. Until they do, it will be hard to adopt AQA.

- QATS is an excellent start. Perhaps in future editions AQA metrics should correlate with comprehension data?
References


What is Text Simplification
Evaluating Text Simplification
My Personal Experience
Challenges for QATS


