TEXT SIMPLIFICATION BEYOND SENTENCE-LEVEL

WHAT IS TEXT SIMPLIFICATION?
Text simplification aims to modify the content and structure of a text in order to make it easier to understand while retaining its original meaning.

Original: Owls are the order Strigiformes, comprising 200 bird of prey species. Owls hunt mostly small mammals, insects, and other birds though some species specialize in hunting fish.

Simplified: An owl is a bird. There are about 200 kinds of owls. Owls’ prey may be birds, large insects (such as crickets), small reptiles (such as lizards) or small mammals (such as mice, rats, and rabbits).

- Unusual concepts are explained
- Uncommon words are replaced
- Complex syntactic structures are changed (e.g., splitting)
- “Unimportant” information is removed

OPERATIONS BEYOND SENTENCE-LEVEL

1. Sentence reordering

(a) Facebook Chief Executive Mark Zuckerberg announced Tuesday that they plans to eventually donate 99 percent of the Facebook stock owned by him and his wife, Priscilla Chan, shares that are worth about $45 billion today.

(b) That amount would make it one of the largest philanthropic commitments ever.

2. Deletion of sentences and Addition of explanatory information

Gone, too, is the Hershey Social Club, [...]. “Everything has been destroyed,” said Amparo Delongh, 92, the first person born in the town and one of the few who stayed to see it fall apart. With U.S. businesses pushing harder [...].

(Simplified)
Gone, too, is the Hershey Social Club, [...]. Private business does not exist in communism. Instead, the government controls business. People from the United States who were running businesses in 1959 had to leave. Then Washington put a trade embargo in place, which has prevented U.S. businesses from trading with Cuba for more than 50 years. With U.S. businesses pushing harder [...].

3. Deletion and pronoun resolution

(a) “We can’t keep living like this” said Elis de Cary Rojas, who runs a nail salon in her home, earning 40 cents a manicure. “Something has to change”.

(b) She moved back to the town with her young daughter a few years ago, preferring its peace and quiet and space to a cramped Havana apartment.

(Simplified)
(a) Elis de Cary Rojas moved back to the town with her young daughter a few years ago, preferring its peace and quiet and space to a cramped city apartment.

PROJECT AIMS

1. Identification of Simplification Operations

(a) Facebook Chief Executive Mark Zuckerberg announced Tuesday that he plans to eventually donate 99 percent of the Facebook stock owned by him and his wife, Priscilla Chan, shares that are worth about $45 billion today.

(b) That amount would make it one of the largest philanthropic commitments ever.

(c) Simplified
The couple’s shares are currently worth about $45 billion.

3. Document-level Evaluation

... holdings, the company’s only properties outside the United States... holdings.

(Simplified)
... holdings, where he painted on a canvas, he painted life-size works.

(Simplified)
... before, he had been creating life-size works.

SO FAR: ANNOTATING OPERATIONS IN PARALLEL SENTENCES

Annotation algorithms and heuristics based on word alignments

- Deletions, replacements and additions
- Operations spanning more than single words
- Movements and (some) multiword replacements
- Evaluation on 100 manually annotated sentences: F$_{1}$ = 0.70

1. Learning How to Simplify from Explicit Labeling of Complex-Simplified Text Pairs

(Simplified)
... Operations Predictor

2. Operation-based performance analysis of sentence simplification systems (PWKP)

<table>
<thead>
<tr>
<th>System</th>
<th>D</th>
<th>M</th>
<th>R</th>
<th>RW</th>
<th>RM</th>
<th>RWM</th>
<th>C</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moses</td>
<td>0.38</td>
<td>0.16</td>
<td>0.53</td>
<td>0.04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.75</td>
<td>0.62</td>
</tr>
<tr>
<td>PBSMT-R</td>
<td>0.14</td>
<td>0.00</td>
<td>0.20</td>
<td>0.00</td>
<td>0.11</td>
<td>0.00</td>
<td>0.71</td>
<td>0.55</td>
</tr>
<tr>
<td>Hybrid</td>
<td>0.35</td>
<td>0.08</td>
<td>0.43</td>
<td>0.05</td>
<td>0.22</td>
<td>0.00</td>
<td>0.75</td>
<td>0.61</td>
</tr>
<tr>
<td>TSM</td>
<td>0.32</td>
<td>0.05</td>
<td>0.17</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.66</td>
<td>0.52</td>
</tr>
<tr>
<td>EncDecA</td>
<td>0.18</td>
<td>0.04</td>
<td>0.13</td>
<td>0.09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.71</td>
<td>0.55</td>
</tr>
<tr>
<td>DRESS</td>
<td>0.47</td>
<td>0.03</td>
<td>0.08</td>
<td>0.05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.64</td>
<td>0.53</td>
</tr>
<tr>
<td>DRESS-LS</td>
<td>0.47</td>
<td>0.03</td>
<td>0.07</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.65</td>
<td>0.54</td>
</tr>
</tbody>
</table>