

Text Simplification Beyond Sentence-Level



Fernando Alva-Manchego, Carolina Scarton and Lucia Specia

Department of Computer Science, University of Sheffield, UK {f.alva,c.scarton,l.specia}@sheffield.ac.uk

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

Problem: Only sentencelevel operations

Real scenario: Simplify entire documents

OPERATIONS BEYOND SENTENCE-LEVEL

Manual analysis of professionally produced simplifications: Newsela corpus.

1. Sentence reordering

Original

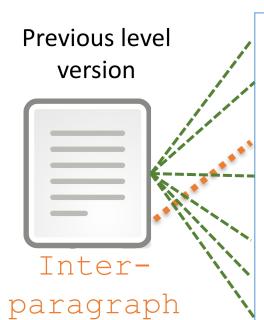
- (a) Facebook Chief Executive Mark Zuckerberg announced Tuesday that he plans to eventually donate 99 percent of the Facebook stock owned by him and his wife, Priscilla Chan, shares that are worth about \$45 billion today.
- (b) That amount would make it one of the largest philanthropic commitments ever.

Intra-paragraph

Simplified

- (a) Facebook Chief Executive Mark Zuckerberg announced that he and his wife, Priscilla Chan, will donate 99 percent of their Facebook stock to charity.
- (b) Their promised gift would be one of the largest charitable donations ever made.
- (c) Together, the couple's shares are currently worth about \$45 billion.

reordering



- (a) Mark Zuckerberg is the head of Facebook.
- (b) He helped start the company when he was a student at Harvard University.
- (c) Today, Facebook is used by 1.5 billion people every month.
- (d) It has made Zuckerberg a very rich man.
- (a) Zuckerberg recently made a big announcement.
- (b) He said he and his wife will give away most of their money.
- (c) His wife's name is Priscilla Chan.

2. Deletion of sentences and Addition of explanatory information

Original

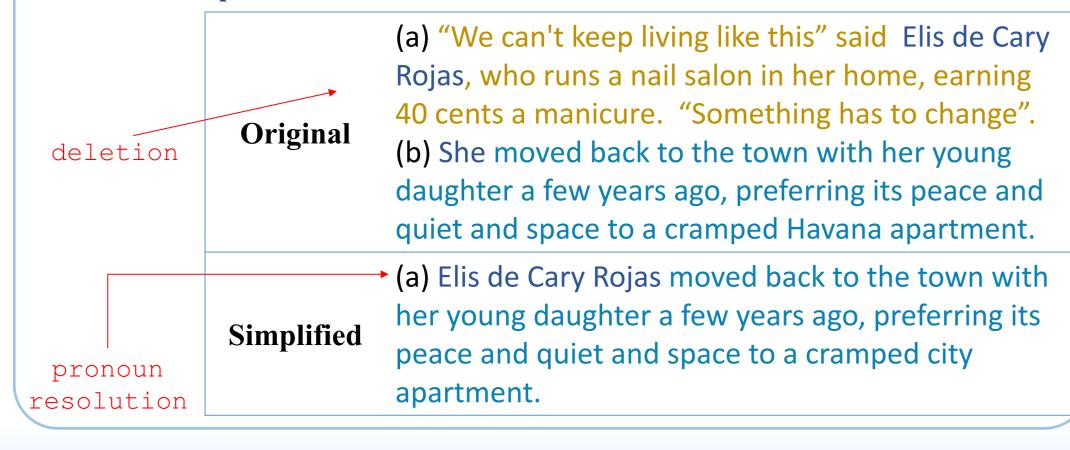
reordering

Gone, too, is the Hershey Social Club, [...] . "Everything has been destroyed," said Amparo DeJongh, 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

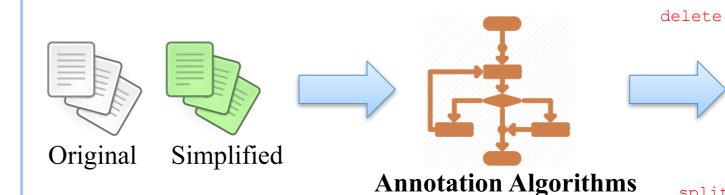
Gone, too, is the Hershey Social Club, [...]. Private business does not exist in communism. Instead, the government controls business. People from the United States who were running businesses in 1959 had to leave. Then Washington put a trade embargo in place, which has prevented U.S. businesses from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

3. Deletion and pronoun resolution.



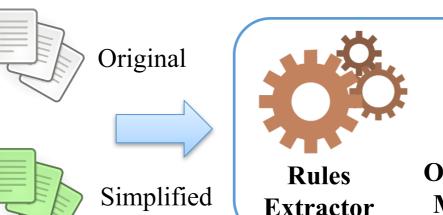
PROJECT AIMS

1. Identification of Simplification Operations



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. That amount would make it one of split the largest philanthropic commitments ever.

2. A Document-level Grammar-based Approach



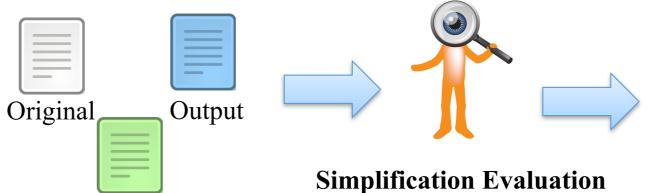




Document Simplifier

3. Document-level Evaluation

Simplified

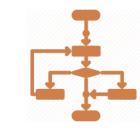


- Which operations were performed at sentence-level?
- Which operations were performed at document-level? - What's the per operation
- performance?
- What's the overall quality of the simplification at document-level?*

SO FAR: ANNOTATING OPERATIONS IN PARALLEL SENTENCES

Annotation algorithms and heuristics based on word alignments

deletions, replacements and additions



Hershey left no heirs when he died in 1945, giving most of his fortune to charity.

Hersey died in 1945 and gave most of his fortune to charity.

B-D I-D B-DB-D I-D ...holdings, the company 's only properties outside the United States.

...holdings.

Operations spanning more than single words

B-DC I-DC I-DC I-DC I-DC I-DC when he painted on a canvas, he painted life-size works.

> before, he had been creating life-size works. B-A B-A B-A

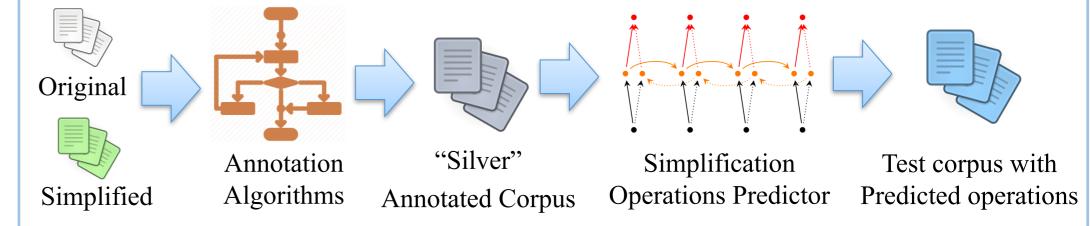
Movements and (some) multiword replacements

[...] DeJongh remembered [...]

Evaluation on 100 manually annotated sentences: $\mathbf{F_1} = \mathbf{0.70}$

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

[...] remembered Amparo DeJongh [...]



- Application of Predicted Operations: delete (not include token in the output) and replace (context-aware neural ranking of substitutions (Paetzold and Specia (2017))
- Evaluation: SARI = 31.29 > Moses (24.58), Nematus (29.89) and NTS (30.44)
- 2. Operation-based performance analysis of sentence simplification systems (PWKP)

	\mathbf{F}_1 measure							
System	D	M	R	RW	RM	RWM	C	Overall
Moses	0.38	0.16	0.33	0.04	0.00	0.00	0.75	0.62
PBSMT-R	0.14	0.00	0.20	0.00	0.11	0.00	0.71	0.55
Hybrid	0.35	0.08	0.43	0.05	0.22	0.00	0.75	0.61
TSM	0.32	0.05	0.17	0.00	0.00	0.00	0.66	0.52
EncDecA	0.18	0.04	0.13	0.09	0.00	0.00	0.71	0.55
DRESS	0.47	0.03	0.08	0.05	0.00	0.00	0.64	0.53
Dress-Ls	0.47	0.03	0.07	0.00	0.00	0.00	0.65	0.54

