

# 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).

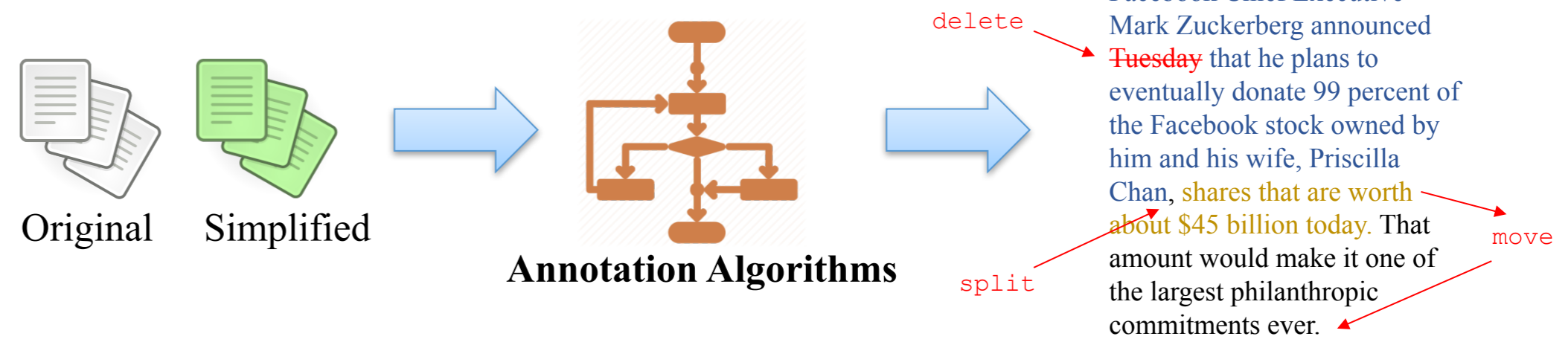
- Unusual concepts are explained
- Uncommon words are replaced
- Complex syntactic structures are changed (e.g., **splitting**)
- "Unimportant" information is removed

**Problem:**  
Only sentence-level operations

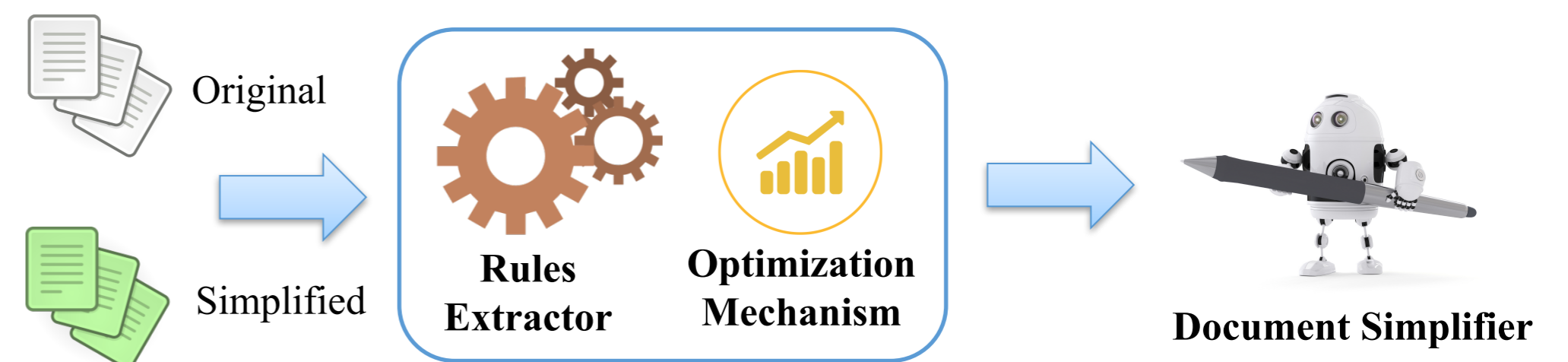
**Real scenario:**  
Simplify entire documents

## PROJECT AIMS

### 1. Identification of Simplification Operations



### 2. A Document-level Grammar-based Approach



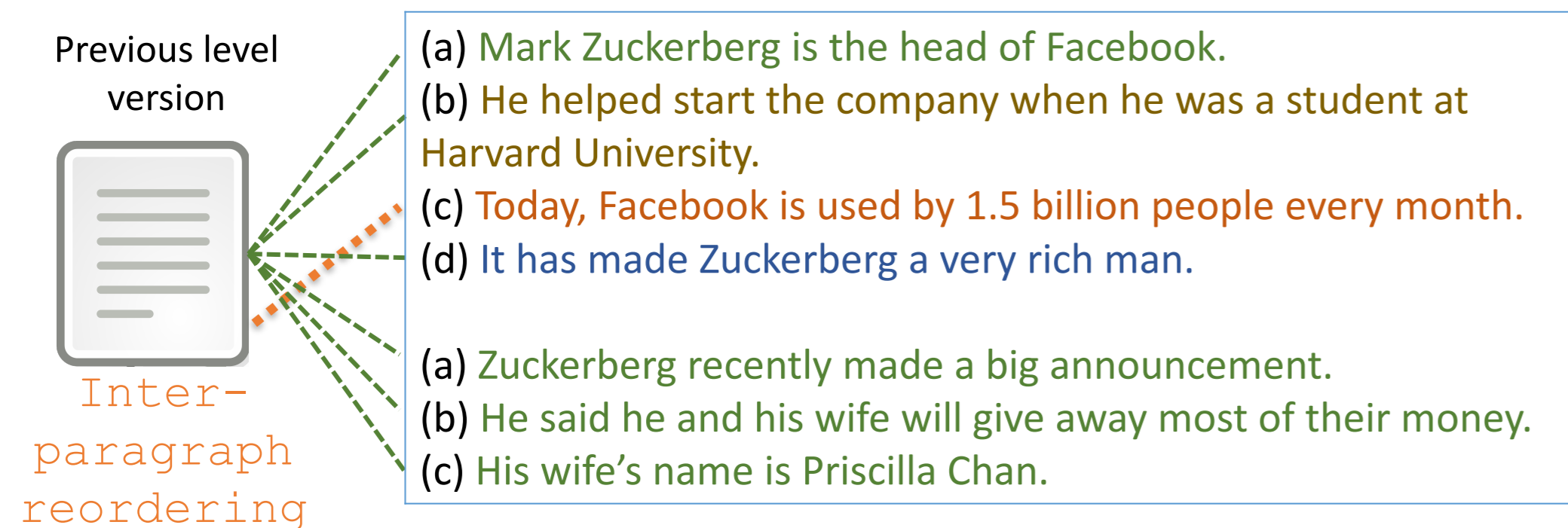
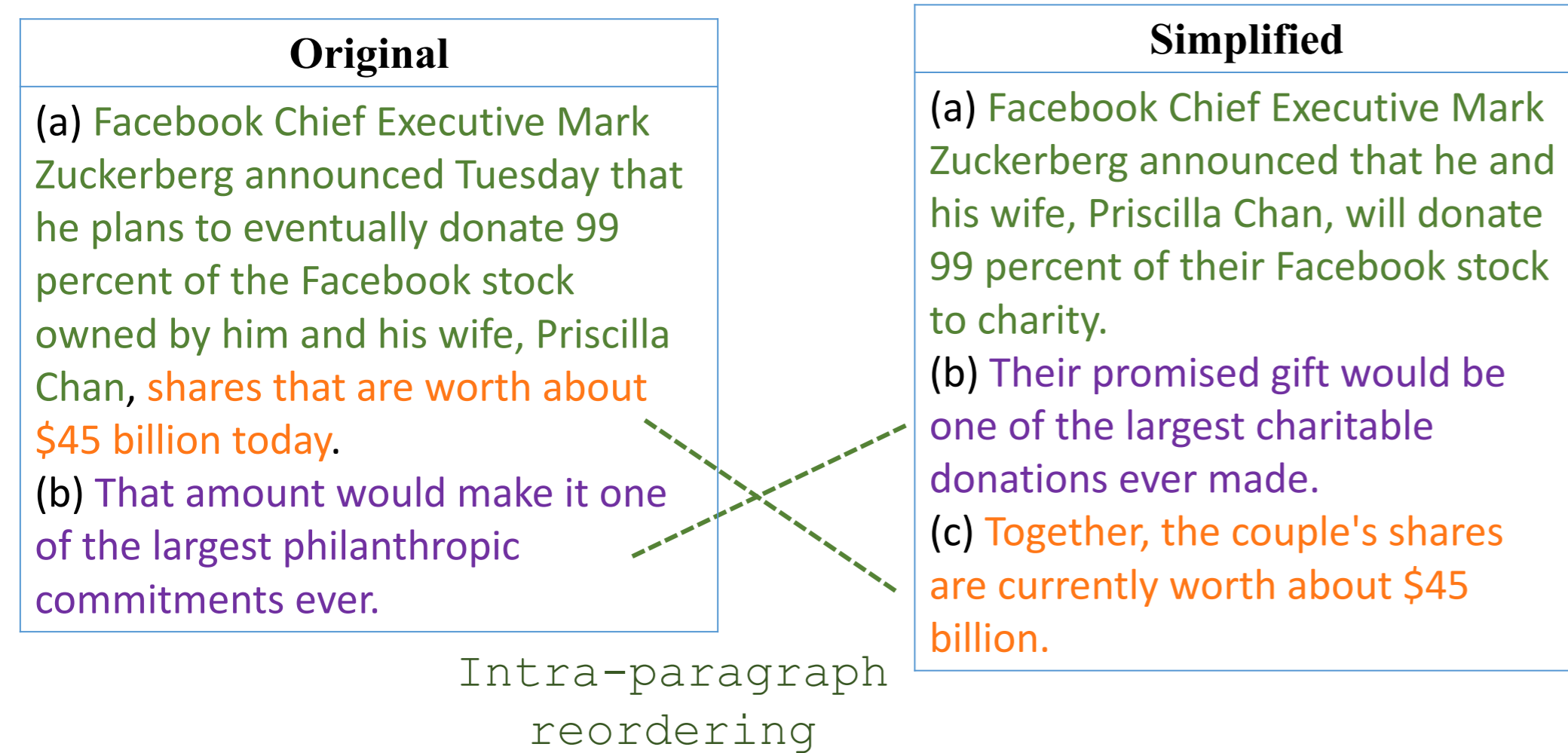
### 3. Document-level Evaluation



## OPERATIONS BEYOND SENTENCE-LEVEL

Manual analysis of professionally produced simplifications: Newsela corpus.

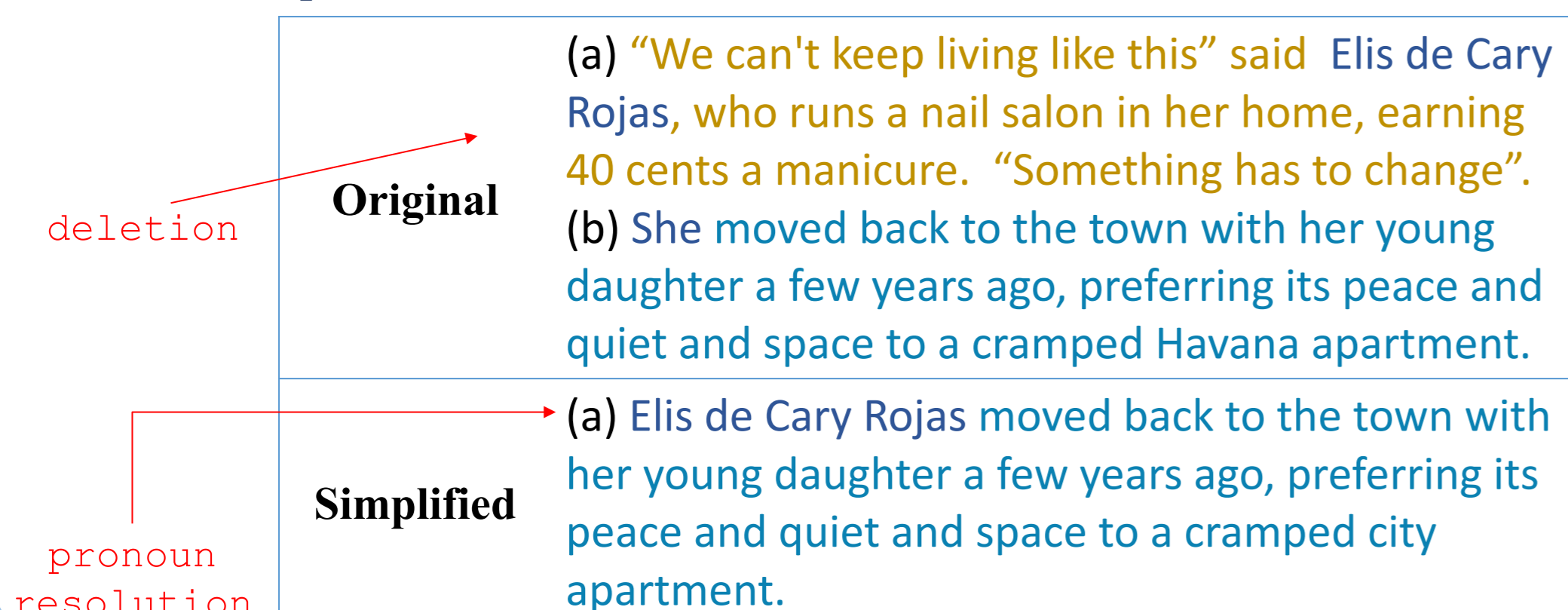
### 1. Sentence reordering



### 2. Deletion of sentences and Addition of explanatory information

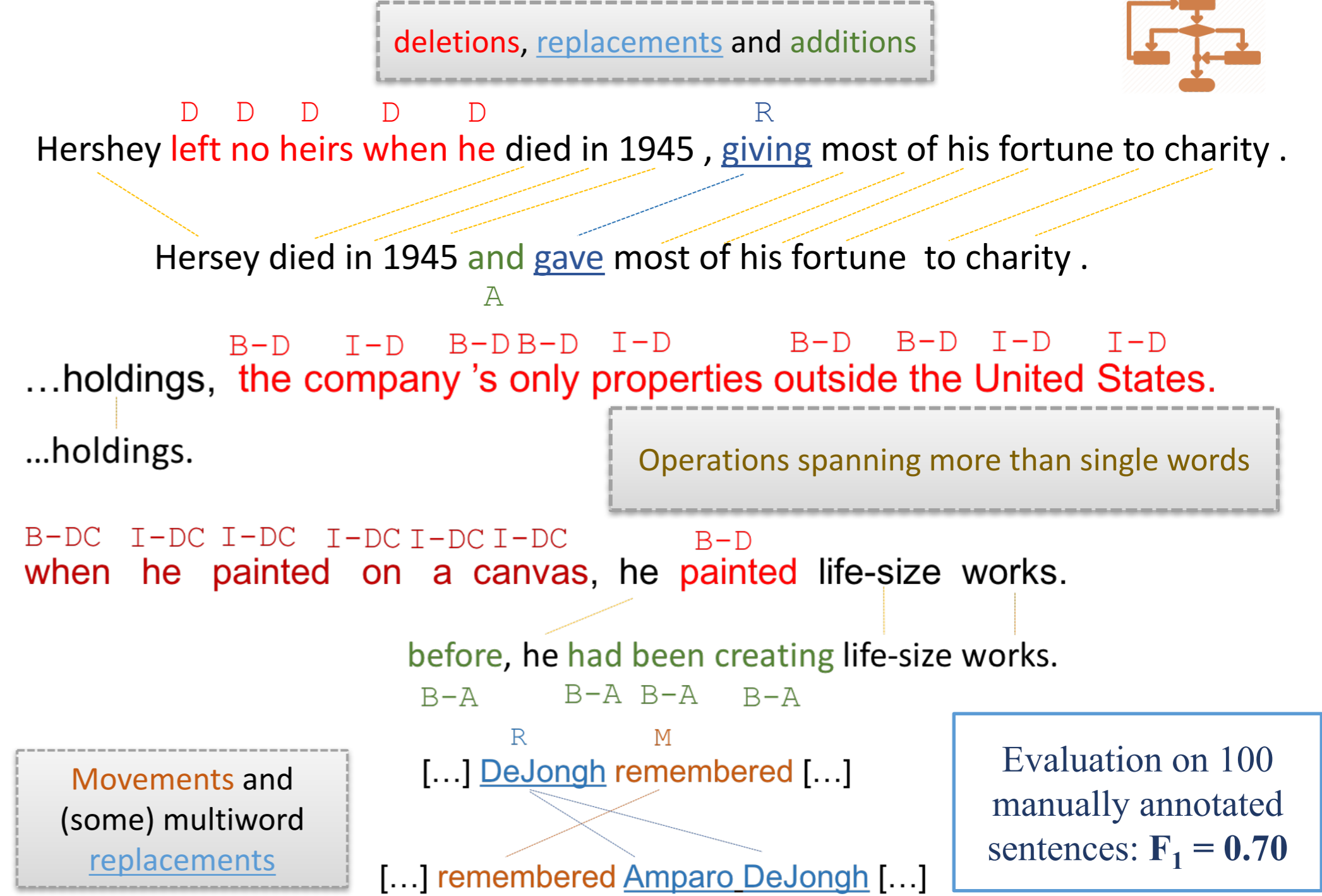
<b>Original</b>	Gone, too, is the Hershey Social Club, [...]. "Everything has been destroyed," said Amparo DeJongh, 92, the first person born in the town and one of the few who stayed to see it fall apart. With U.S. businesses pushing harder [...].
<b>Simplified</b>	Gone, too, is the Hershey Social Club, [...]. 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.

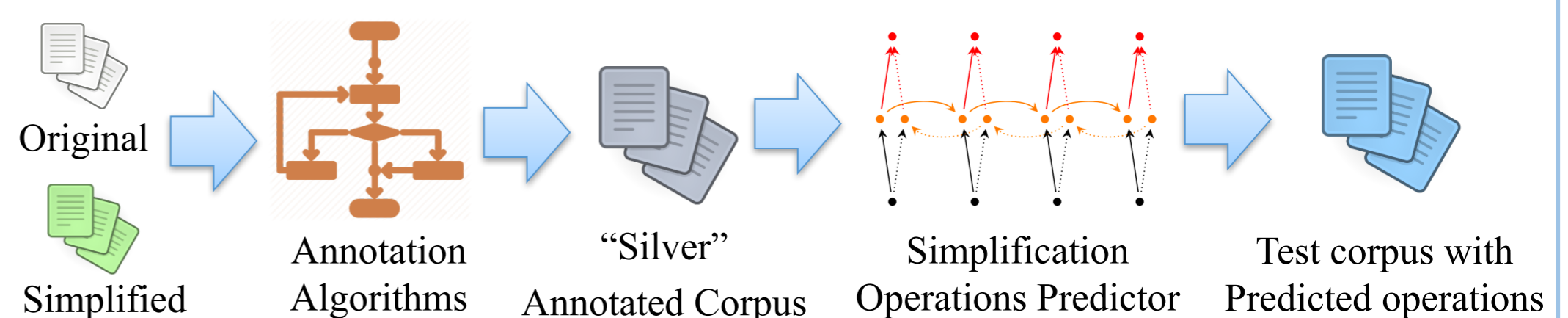


## SO FAR: ANNOTATING OPERATIONS IN PARALLEL SENTENCES

Annotation algorithms and heuristics based on word alignments



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



- **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)

System	F <sub>1</sub> measure							Overall
	D	M	R	RW	RM	RWM	C	
Moses	0.38	<b>0.16</b>	0.33	0.04	0.00	0.00	<b>0.75</b>	<b>0.62</b>
PBSMT-R	0.14	0.00	0.20	0.00	0.11	0.00	0.71	0.55
Hybrid	0.35	0.08	<b>0.43</b>	0.05	<b>0.22</b>	0.00	<b>0.75</b>	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	<b>0.09</b>	0.00	0.00	0.71	0.55
DRESS	<b>0.47</b>	0.03	0.08	0.05	0.00	0.00	0.64	0.53
DRESS-LS	<b>0.47</b>	0.03	0.07	0.00	0.00	0.00	0.65	0.54

