The University Of Sheffield.

Multi-Operation Automatic Text Simplification

Fernando Alva Manchego

@feralvam

BeyondMinds

05 July 2021

Outline

- What is (Automatic) Text Simplification?
- Latest Contributions:
  - A **new dataset** for evaluation of multi-operation simplifications
  - A **hybrid approach** for controllable simplification

What is Text Simplification?

To modify the content and structure of a text so that it is easier to understand while preserving its main idea and as much as possible of its 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~~.

Simplification

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).

- **Elaboration:** Unusual concepts are explained
- **Lexical Paraphrasing:** Uncommon words are replaced by simpler synonyms
- **Sentence Splitting:** A long sentence is divided into several smaller ones
- **Compression:** "Unimportant" information is removed

From: http://videolectures.net/esslli2011_lapata_simplification/

What is Text Simplification useful for?

- Information Accessibility
  - Comprehension in low-ability readers (Mason and Kendall, 1978)
  - Hard-of-hearing children (Quinley et al., 1977; Robbins and Hatcher, 1981)
  - Adults suffering from aphasia (Shewan, 1985)
  - People with dyslexia (Rello et al., 2013)
  - Non-native speakers and ESL learners (Crossley et al., 2007)
- NLP Tasks
  - Parsing (Chandrasekar et al., 1996)
  - Summarisation (Siddharthan et al., 2004; Silveira and Branco, 2012)
  - Machine Translation (Štajner and Popovic, 2016)
  - …

Simplification Scope

- **Word-Level** (a.k.a Lexical Simplification)

The cat perched on the mat. → The cat sat on the mat.

- **Sentence-Level**

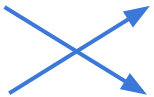
The second largest city of Russia and one of the world's major cities, St. Petersburg has played a vital role in Russian history.

→ St. Petersburg is the second biggest city in Russia.

→ St. Petersburg has played an important role in Russian history.

- **Document-Level**

(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.



(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.

Automatic Sentence Simplification

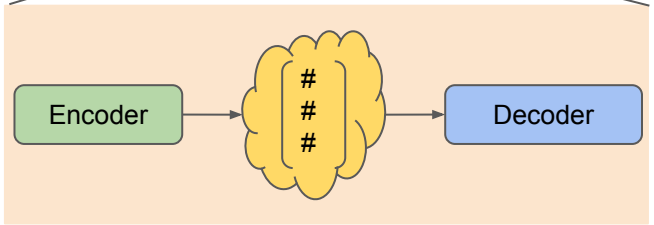
Slightly more fourth-graders **nationwide** are **reading proficiently compared with** a decade ago**, but** only a third of them **are now reading well**, according to a new report.

(Neural) Simplification Model

Fourth-graders are **better readers than 10 years ago. But few** of them **read well**.

Sequence-to-Sequence Model
- Machine Translation
- Summarization
- Caption Generation

...



How do you determine the quality of an automatic simplification?

Metrics used in Machine Translation

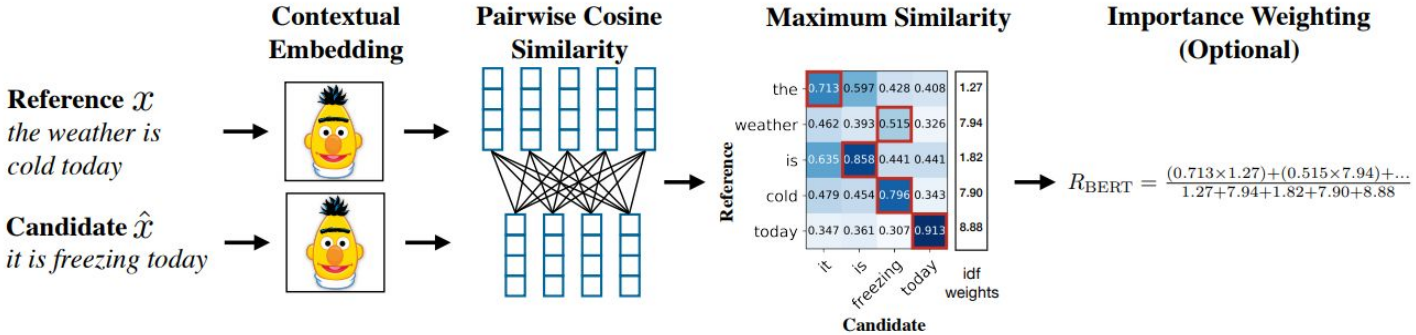
- **BLEU** (Papineni et al., 2002)

$$p_n = \frac{\sum_{S \in C} \sum_{ngram \in S} Count_{matched}(ngram)}{\sum_{S \in C} \sum_{ngram \in S} Count(ngram)}$$

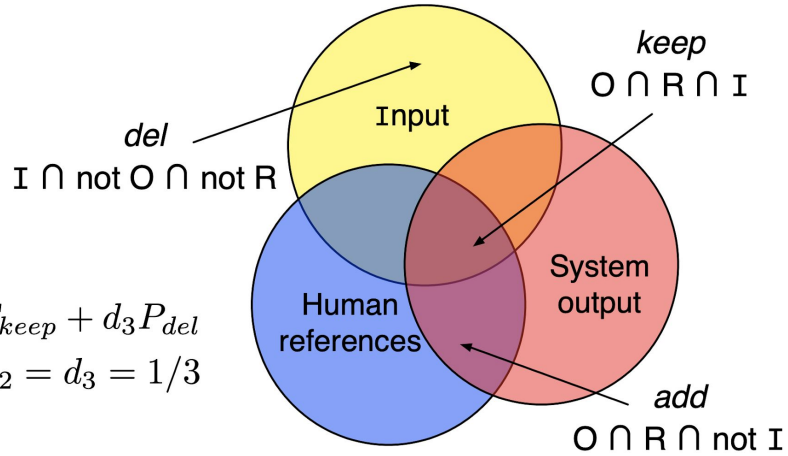
$$BP = \begin{cases} 1 & if \ \ c > r \\ e^{1-\frac{r}{c}} & if \ \ c \le r \end{cases}$$

$$BLEU = BP \times exp\left(\sum_{n=1}^{N} w_n \log p_n\right)$$

- **BERTScore** (Zhang et al., 2020)



SARI (Xu et al., 2016)



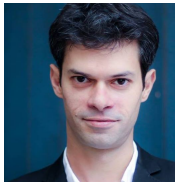
$$\text{SARI} = d_1 F_{add} + d_2 F_{keep} + d_3 P_{del}$$
$$d_1 = d_2 = d_3 = 1/3$$

Input: About 95 species are currently accepted.

REF-1: About 95 species are currently known .
REF-2: About 95 species are now accepted .
REF-3: 95 species are now accepted .

Output-1: About 95 you now get in . → 0.2683
Output-2: About 95 species are now agreed . → 0.7594
Output-3: About 95 species are currently agreed. → 0.5890



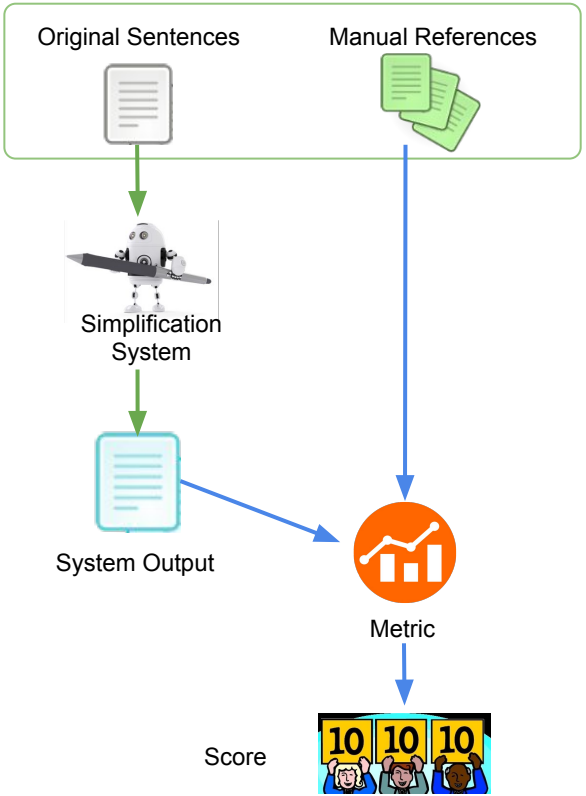

ASSET: A Dataset for Tuning and Evaluation of Sentence Simplification Models with Multiple Rewriting Transformations

Fernando Alva-Manchego[1*] and **Louis Martin**[2,3*] and **Antoine Bordes**[3]
Carolina Scarton[1] and **Benoît Sagot**[2] and **Lucia Specia**[1,4]
[1]University of Sheffield, [2]Inria, [3]Facebook AI Research, [4]Imperial College London

ACL 2020

https://github.com/facebookresearch/asset

Automatic Evaluation in Sentence Simplification



- Human editors perform **multiple operations** simultaneously
  - *Shouldn't automatic systems be expected to do the same?*
  - *Are we evaluating them for that?*
- **<u>Multi-reference</u> evaluation datasets** focus on only one operation:

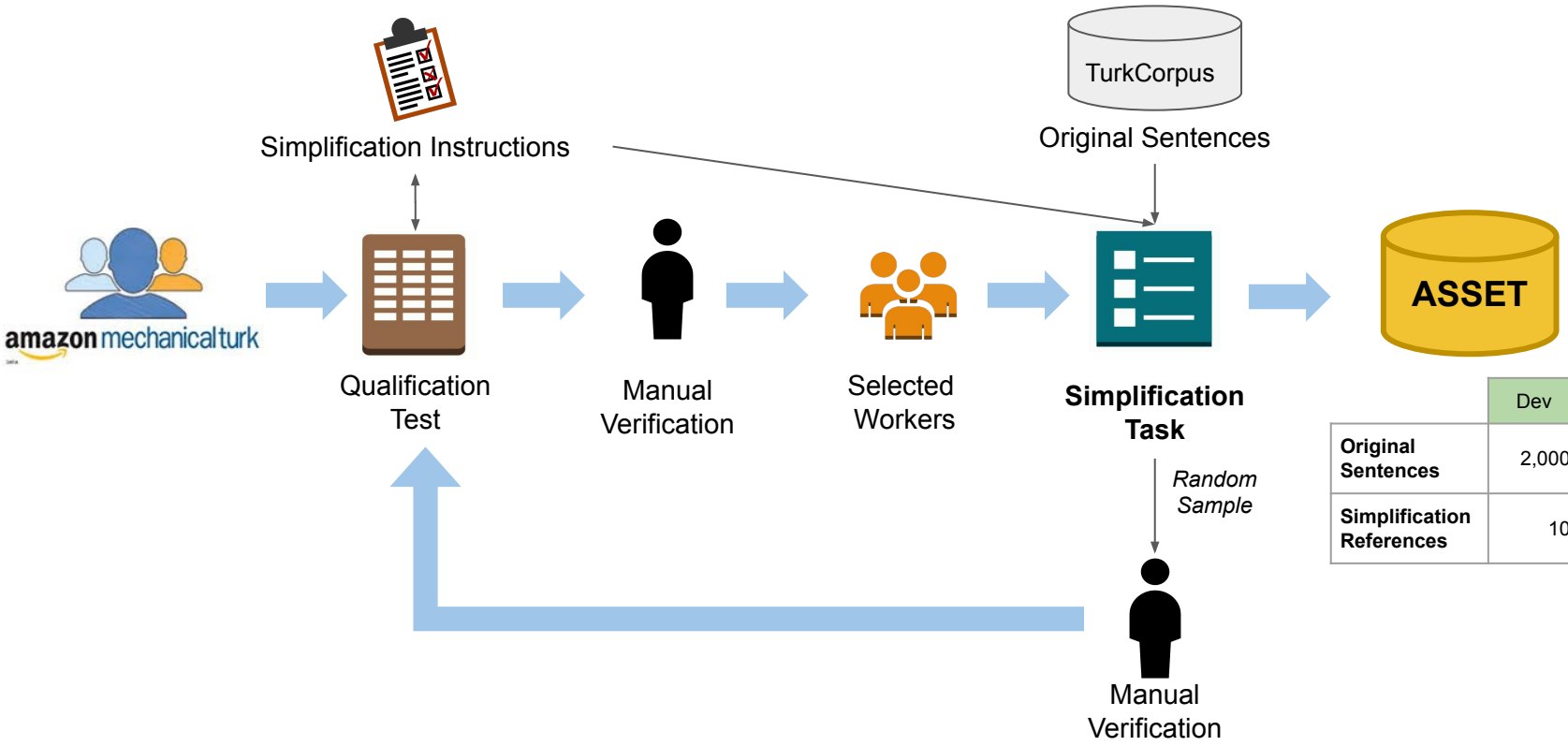
Dataset	Dev	Test	Operation
TurkCorpus (Xu et al., 2016)	✓	✓	Lexical Paraphrasing
HSplit (Sulem et al., 2018)		✓	Sentence Splitting

Introducing ASSET

- A **multi-reference** dataset for sentence simplification in **English**
- Human editors were instructed to perform **3 types of operations**:
  - Lexical Paraphrasing
  - Compression
  - Sentence Splitting

Original	He settled in London, devoting himself chiefly to practical teaching.
ASSET	He lived in London. He was a teacher.
TurkCorpus	He rooted in London, devoting himself mainly to practical teaching.
HSplit	He settled in London. He devoted himself chiefly to practical teaching.

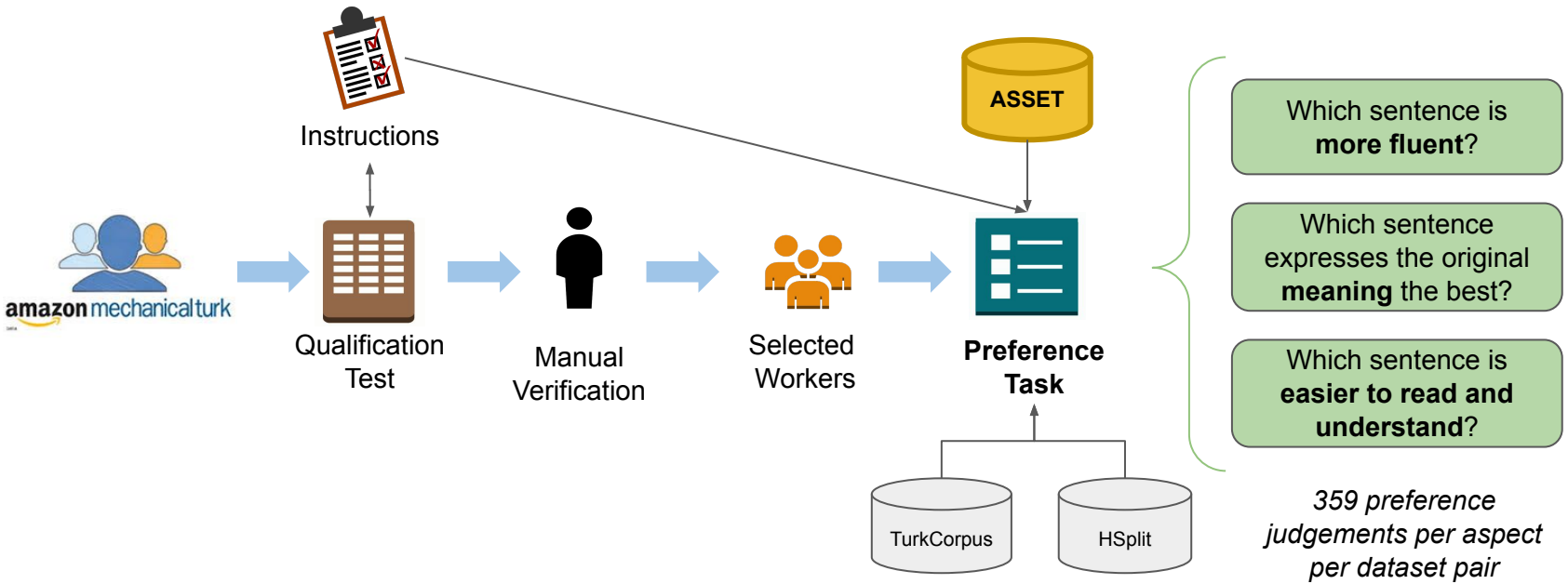
How was ASSET created?



	Dev	Test
Original Sentences	2,000	359
Simplification References	10	10

Quality of Simplifications in ASSET

Do humans prefer multi-operation over single-operation simplifications?



Quality of Simplifications in ASSET

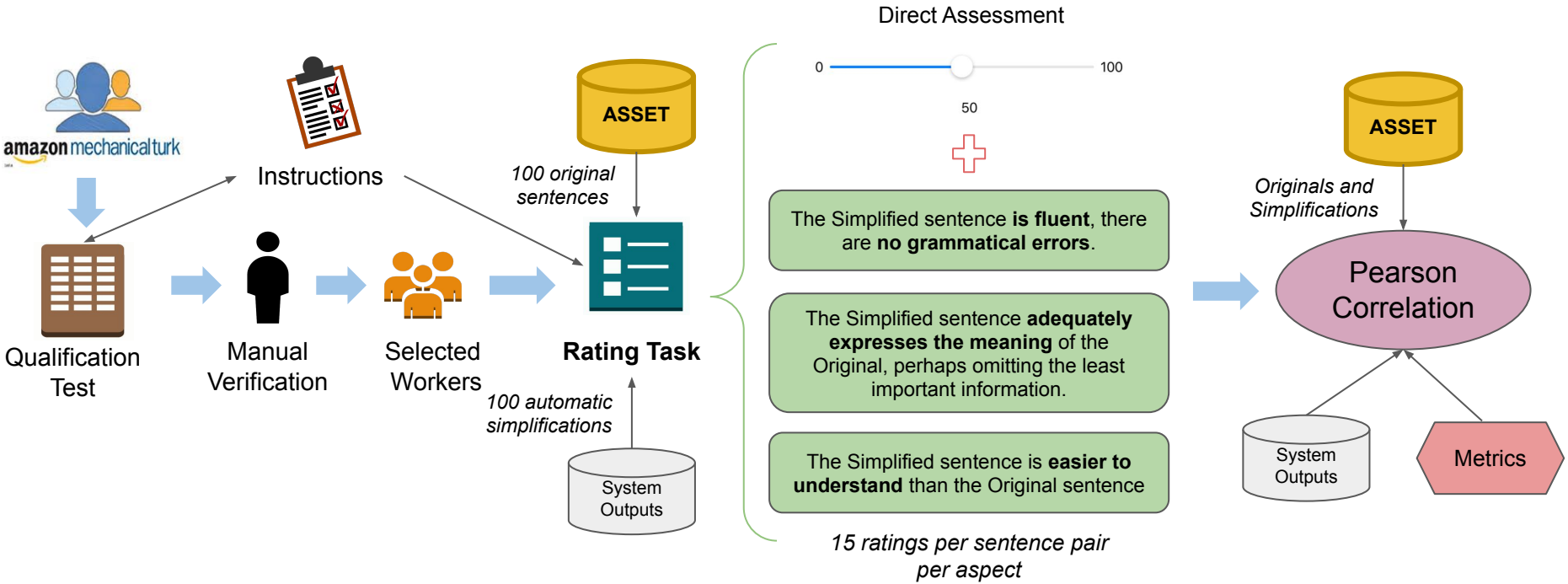
- **ASSET**'s simplifications are preferred (or similar) in terms of **fluency** and **simplicity** over TurkCorpus or HSplit
- Simplifications from **TurkCorpus** or **HSplit are more meaning preserving**
  - Compression was not allowed when creating simplifications

	Fluency	Meaning	Simplicity
ASSET	**38.4%***	23.7%	**41.2%***
TurkCorpus	22.8%	**37.9%***	20.1%
Similar	*38.7%*	*38.4%*	*38.7%*
ASSET	**53.5%***	17.0%	**59.0%***
HSplit	19.5%	**51.5%***	14.8%
Similar	*27.0%*	*31.5%*	*26.2%*

Percentages of judges who preferred simplifications in ASSET or TurkCorpus/HSplit

ASSET for Automatic Evaluation

Are standard multi-reference automatic evaluation metrics reliable when using multi-operation simplifications?



ASSET for Automatic Evaluation

BLEU (Papineni et al., 2002)**:**

- Strong **correlation with Meaning Preservation** using simplifications from ASSET or TurkCorpus
- Some **correlation with Fluency** judgements, but that is **not always the case for Simplicity**.
  - In line with previous work that has shown that **BLEU is not a good estimate for simplicity**

Metric	References	Fluency	Meaning	Simplicity
BLEU	ASSET	**0.42***	**0.61***	**0.31***
	TurkCorpus	**0.35***	**0.59***	0.18
SARI	ASSET	0.16	0.13	**0.28***
	TurkCorpus	0.14	0.10	0.17

Pearson correlation of human ratings with automatic metrics on automatic simplifications. (*) $p < 0.05$

SARI (Xu et al., 2016)**:**

- **Low correlation** with all criteria and significant only for **simplicity** with ASSET's references

SARI may not be suitable to evaluate simplicity in **multi-operation** simplifications

Better metrics are needed!

Takeaways

- ASSET, a **new multi-reference dataset** for evaluation of Sentence Simplification in English
  - Simplifications contain **multiple rewriting transformations**
  - Simplifications are judged **simpler than those in other evaluation corpora**
- **Multi-reference automatic evaluation** metrics show **low correlation** for human judgements of **Simplicity** when using multi-operation simplifications
  - **New metrics are required** for automatic evaluation of simplifications with multiple rewriting operations

Controllable Text Simplification with Explicit Paraphrasing

Mounica Maddela[1], Fernando Alva-Manchego[2], Wei Xu[1]
[1]School of Interactive Computing, Georgia Institute of Technology
[2]Department of Computer Science, University of Sheffield

NAACL 2021

https://github.com/mounicam/controllable_simplification

Previously: Automatic Text Simplification

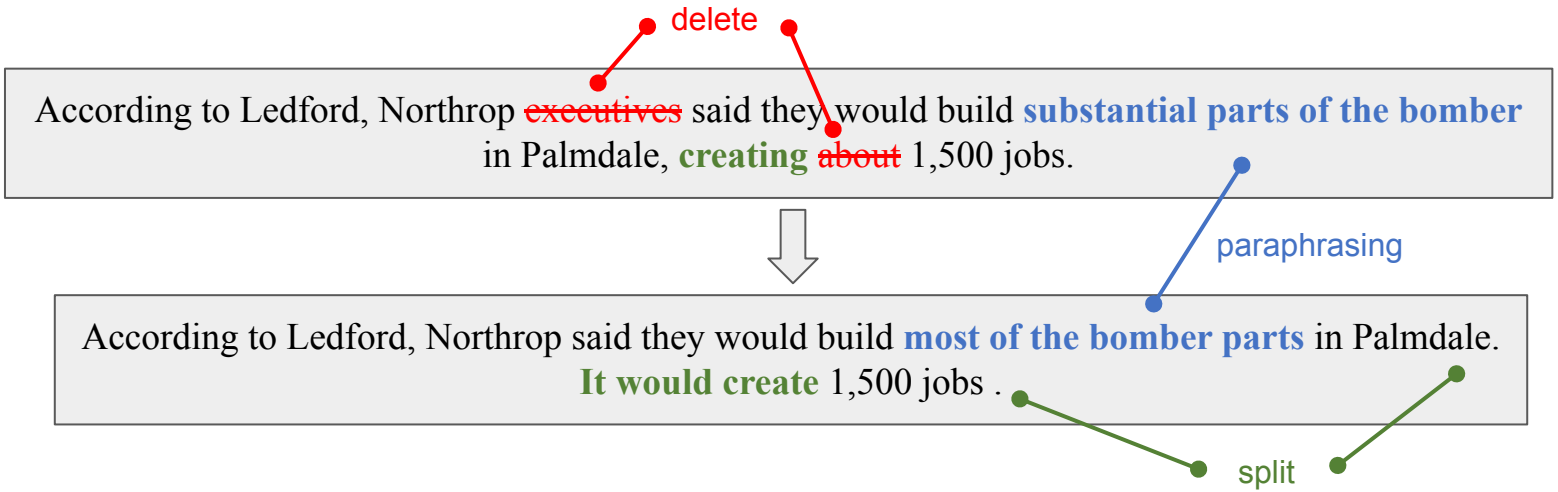
- **Rewrite** complex text **into simpler language** while retaining the original meaning
- Often involves **three operations**: delete, split and paraphrase

According to Ledford, Northrop executives said they would build substantial parts of the bomber in Palmdale, creating about 1,500 jobs.

According to Ledford, Northrop said they would build most of the bomber parts in Palmdale. It would create 1,500 jobs .

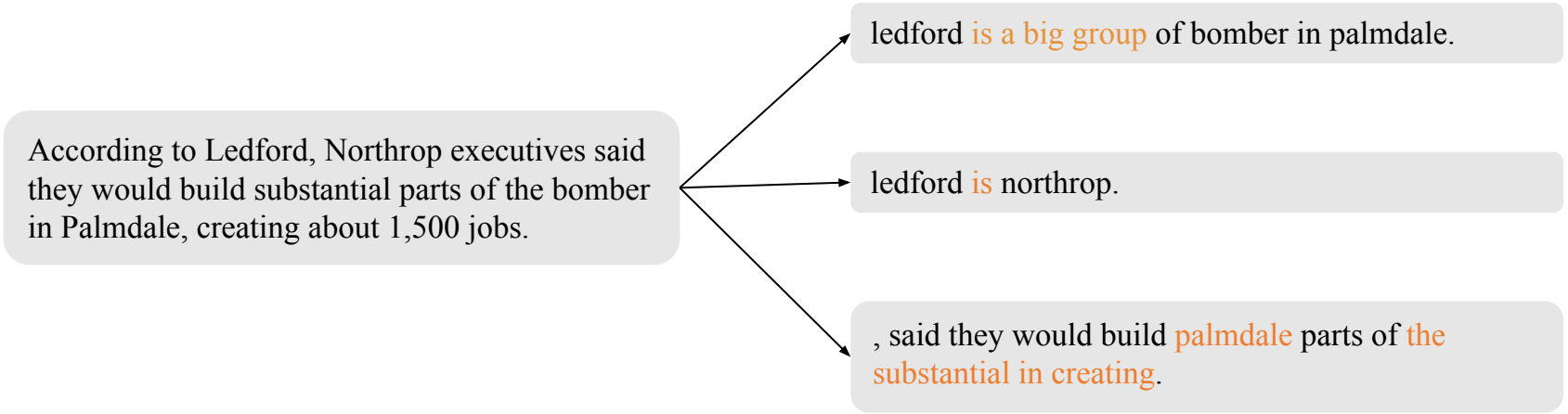
Previously: Automatic Text Simplification

- **Rewrite** complex text **into simpler language** while retaining the original meaning
- Often involves **three operations**: delete, split and paraphrase



Problems with SotA Simplification Systems

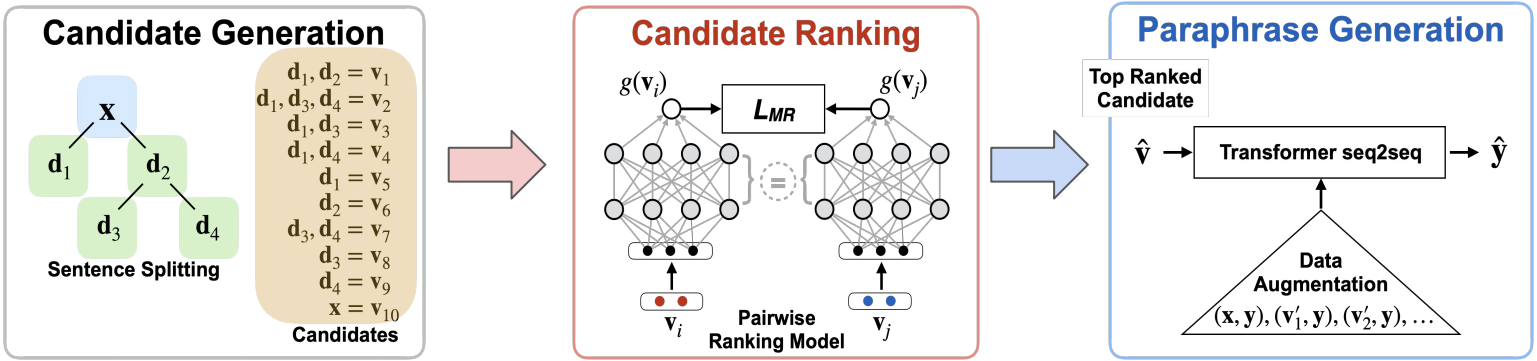
- Perform **mostly deletion**;



- **Lack controllability** in terms of edit operations

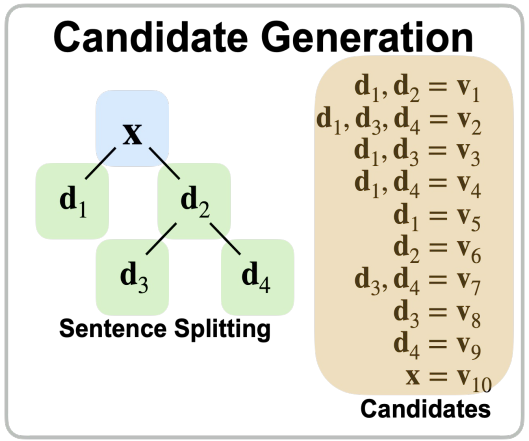
Our Work - Controllable Text Generation

- **Control over 3 edit operations** - deletion, splitting and paraphrasing
- Incorporate **linguistic rules with neural** generation models
- **New setup to evaluate** models' capability over these edit operations



Step 1: Candidate Generation

- **Split + Delete**: Rule-based method (Niklaus et al., 2019) + a seq2seq model

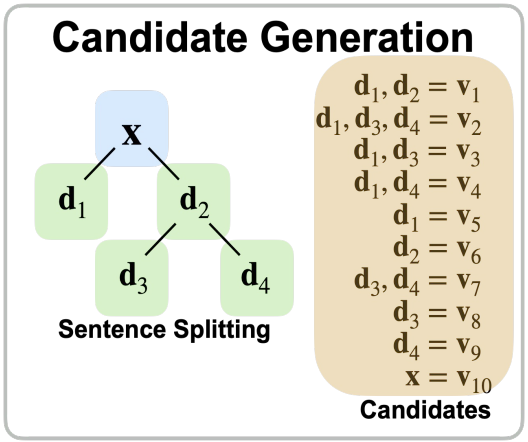


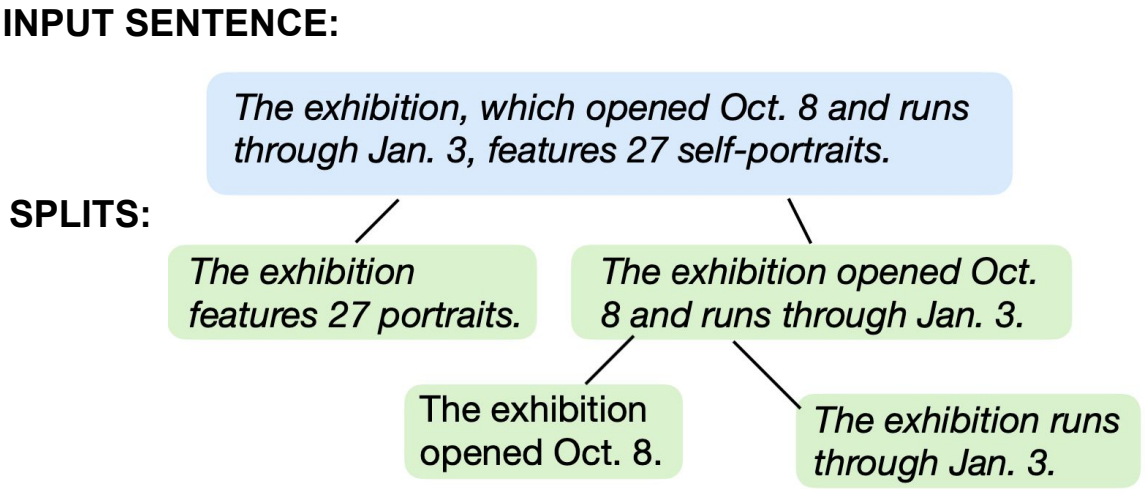
- 35 hand-crafted grammar rules for English based on Stanford's parser
- successfully split 92% of sentences with >= 20 words and make only 6.8% errors.

Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2019. Transforming Complex Sentences into a Semantic Hierarchy. In *ACL 2019*.

Step 1: Candidate Generation

- **Split + Delete**: Rule-based method (Niklaus et al., 2019) + a seq2seq model

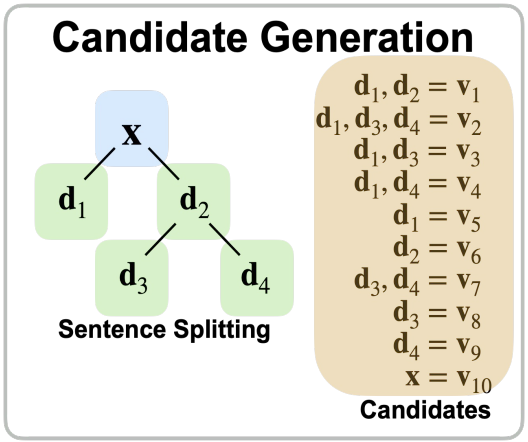




Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2019. Transforming Complex Sentences into a Semantic Hierarchy. In *ACL 2019*.

Step 1: Candidate Generation

- **Split + Delete**: Rule-based method (Niklaus et al., 2019) + a seq2seq model



COMBINE SPLITS AS CANDIDATES:

The exhibition features 27 portraits. The exhibition opened Oct. 8 and runs through Jan. 3.

The exhibition opened Oct. 8 and runs through Jan. 3.

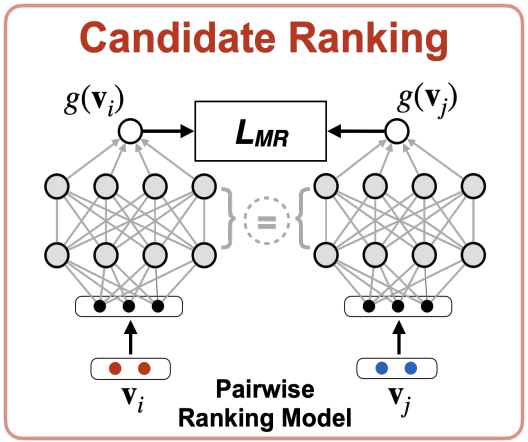
The exhibition features 27 portraits.

The exhibition opened Oct. 8. The exhibition runs through Jan. 3. *.. (and more)*

Christina Niklaus, Matthias Cetto, André Freitas, and Siegfried Handschuh. 2019. Transforming Complex Sentences into a Semantic Hierarchy. In *ACL 2019*.

Step 2: Candidate Ranking

- **Rank all candidate outputs** after (splitting and deletion)



"Gold" Scoring Function

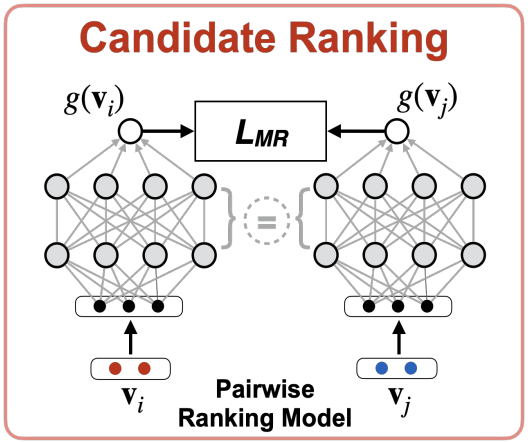
target compression ratio

$$g^*(\mathbf{v}_i, \mathbf{y}) = e^{-\lambda||\phi_{\mathbf{v}_i} - \phi_{\mathbf{y}}||} \times BERTScore(\mathbf{v}_i, \mathbf{y})$$

candidate reference

Step 2: Candidate Ranking

- **Rank all candidate outputs** after (splitting and deletion)



Loss function:

$$L_{MR} = \frac{1}{m}\sum_{k=1}^{m}\frac{1}{n_k^2}\sum_{i=1}^{n_k}\sum_{j=1,i\neq j}^{n_k}\max(0, 1 - l_{ij}^k d_{ij}^k)$$

$$d_{ij}^k = g(\mathbf{v}_i^k) - g(\mathbf{v}_j^k)$$

$$l_{ij}^k = sign\left(g^*(\mathbf{v}_i^k, \mathbf{y}^k) - g^*(\mathbf{v}_j^k, \mathbf{y}^k)\right)$$

Ranker score

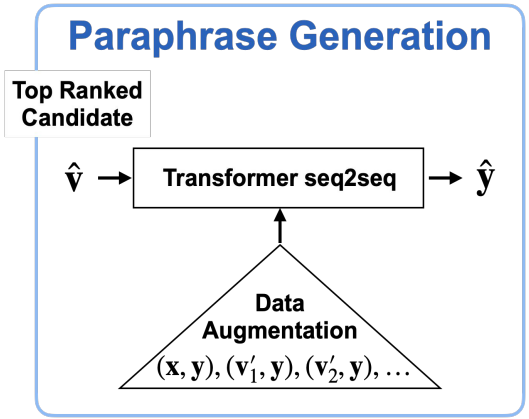
Length-penalized BERTScore

Features:

number of words in v_i and x, compression ratio of v_i with respect to x, Jaccard similarity between v_i and x, the rules applied on x to obtain v_i, and the number of rule applications.

Step 3: Paraphrase Generation

- Paraphrase **top-ranked** candidate
  - **Data Augmentation** with additional training data that focuses on lexical paraphrasing
  - **Copy-control token** as a soft constraint to control paraphrasing
  - **Auxiliary task** (if a word should be copied)



Example Output

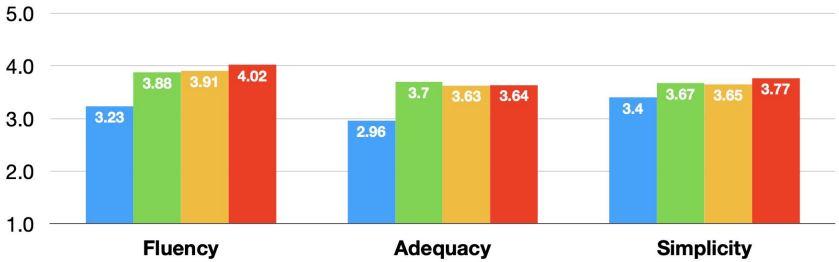
Input: Since 2010, project researchers have uncovered documents in Portugal that have revealed who owned the ship

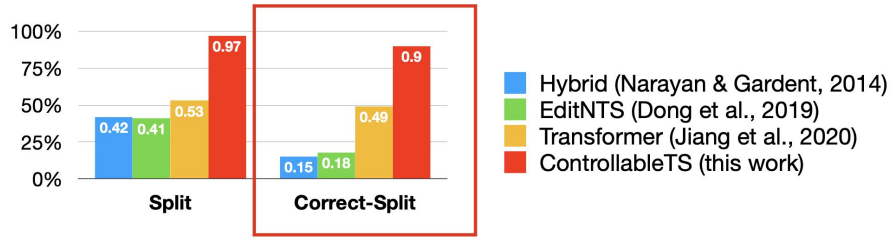
Reference: Scientists have found documents in Portugal. They have also found out who owned the ship.

Hybrid-NG	since 2010, project researchers have uncovered documents in portugal that have **about** who **owns** the ship.
LSTM	since 2010, **scientists** have uncovered documents in portugal that have revealed who owned the ship.
Transformer	**they discovered** that the ship **had been important**.
EditNTS	since 2010, project researchers have uncovered documents in portugal**. have** revealed who owned the ship
Our Model (*split, cp = 0.6*)	**scientists** have **found** a **secret deal**. **they have discovered** who owned the ship.
Our Model (*split, cp = 0.7*)	**scientists** have **found** documents in portugal**. they have also found out** who owned the ship.
Our Model (*split, cp = 0.8*)	**scientists** have **found** a documents in portugal**. they have discovered** who owned the ship.

Experiments - Human Evaluation

- Trained on Newsela-Auto (Jiang et al., 2020)
- 259,778 train + 32,689 valid + 33,391 test pairs
- Human evaluation on 100 random simplifications

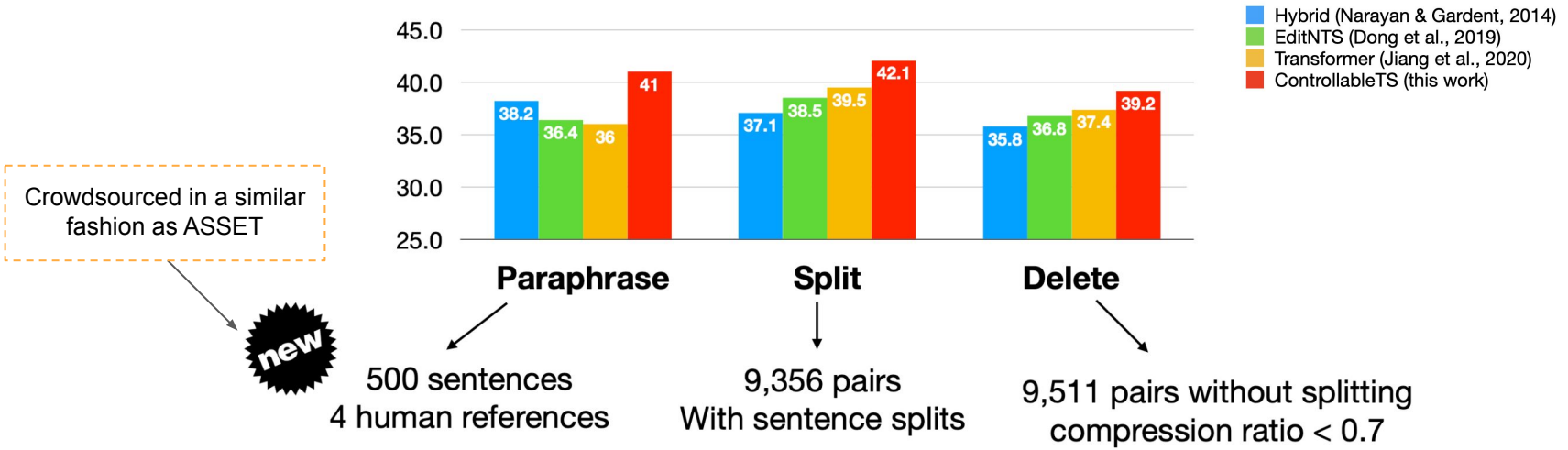




Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. 2020. Neural CRF Sentence Alignment Model for Text Simplification. In *ACL 2020*.

Edit-Focused Automatic Evaluation

- Evaluation on sections of the Newsela-Auto test set (Jiang et al., 2020)
- We report SARI, the main automatic metric for Sentence Simplification



Chao Jiang, Mounica Maddela, Wuwei Lan, Yang Zhong, and Wei Xu. 2020. Neural CRF Sentence Alignment Model for Text Simplification. In *ACL 2020*.

Takeaways

- Novel **hybrid approach** for Sentence Simplification that offers **control over 3 types of edit operations**
- Inject **linguistic knowledge** into neural models
- **New dataset** to evaluate lexical paraphrasing
- **Evaluation setup** for edir operations

Conclusions

- Text Simplification is a text generation tasks whose goal is to **rewrite a complex** sentence into an **easier to understand** version
- Types of Rewriting: delete, paraphrase, split, etc.
- In this talk:
  - A **new dataset for evaluation** with multi-operation simplification references
  - A new **hybrid approach** for **controllable** simplification
- Challenges:
  - Develop new metrics that evaluate multi-operation capabilities
  - Design models that perform more variety of operations (e.g. elaboration)

Thanks!

Fernando Alva Manchego

f.alva@sheffield.ac.uk
@feralvam