



The
University
Of
Sheffield.

Inria

FACEBOOK AI

Imperial College
London

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

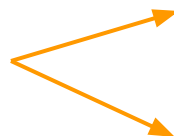
Fernando Alva-Manchego*, Louis Martin*, Antoine Bordes,
Carolina Scarton, Benoît Sagot and Lucia Specia

ACL 2020

What is Sentence Simplification?

To modify the content and structure of a sentence so that it is **easier to understand** while preserving its original meaning

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

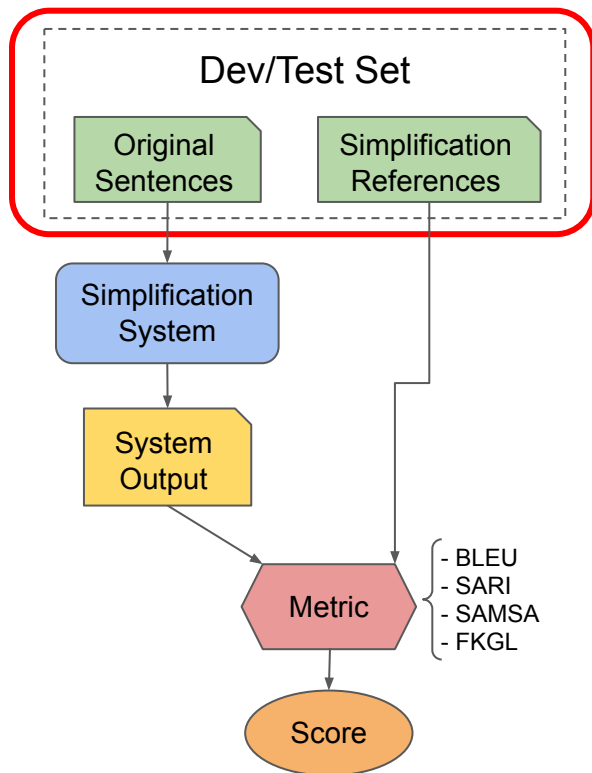


St. Petersburg is the second **biggest** city in Russia.

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

- **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

Automatic Evaluation in Sentence Simplification



- **Multi-reference 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 |

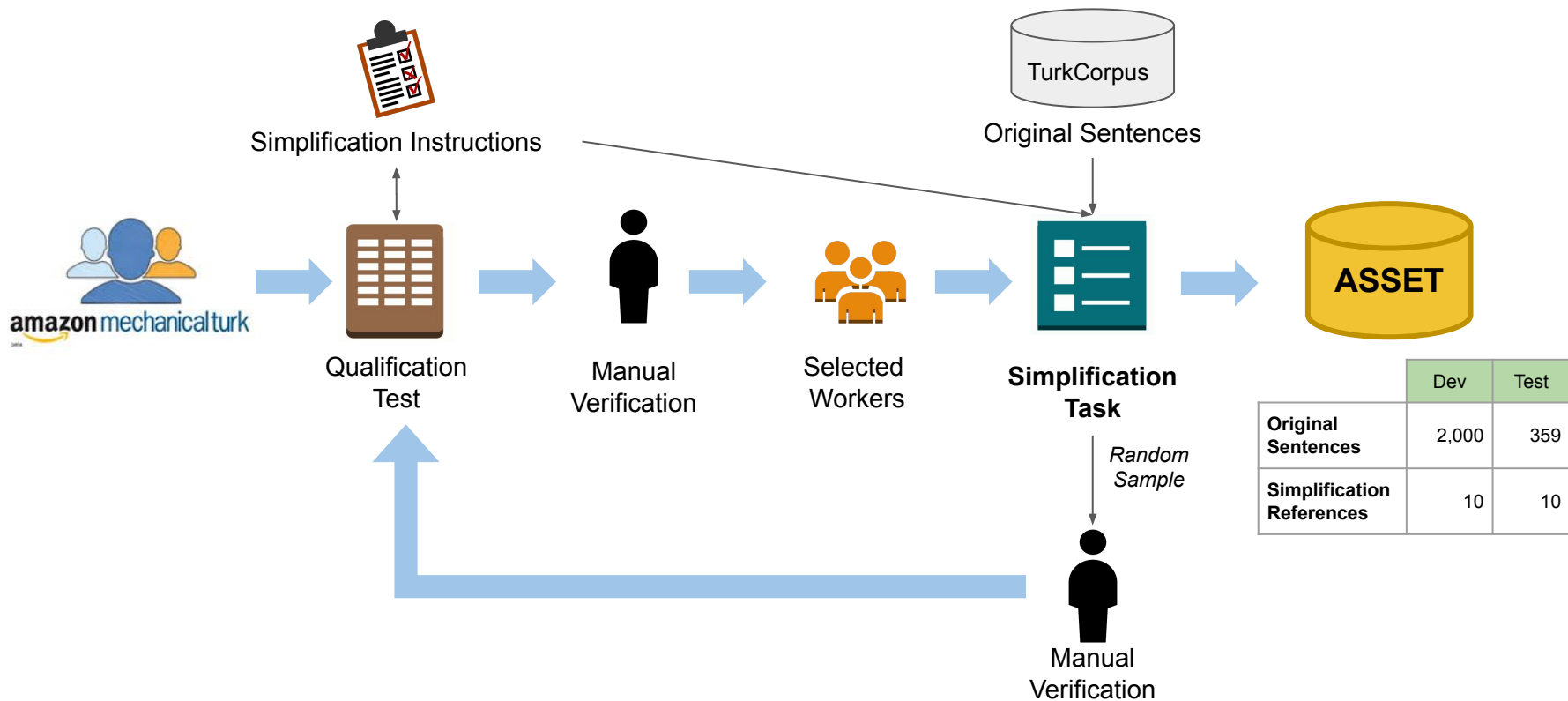
- Human editors perform **multiple operations** simultaneously
 - *Shouldn't automatic systems be expected to do the same?*
 - *Are we evaluating them for that?*

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?



Evaluating Simplifications in ASSET

1. Operations Performed

- *Did we succeed in collecting multi-operation simplifications?*

2. Quality of Simplifications

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

3. As References for Automatic Evaluation

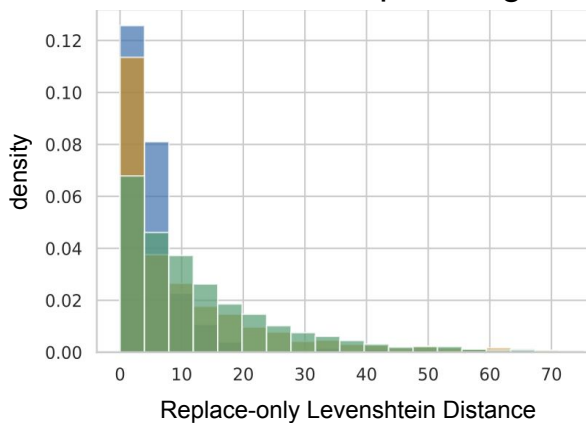
- *Are standard multi-reference automatic evaluation metrics reliable when using multi-operation simplifications?*

1. Operations Performed in ASSET

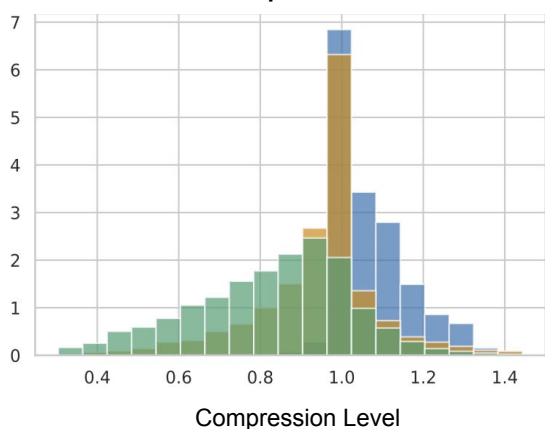
Did we succeed in collecting multi-operation simplifications?



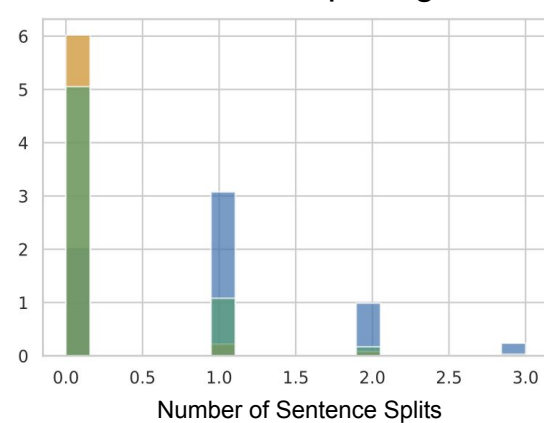
Lexical Paraphrasing



Compression

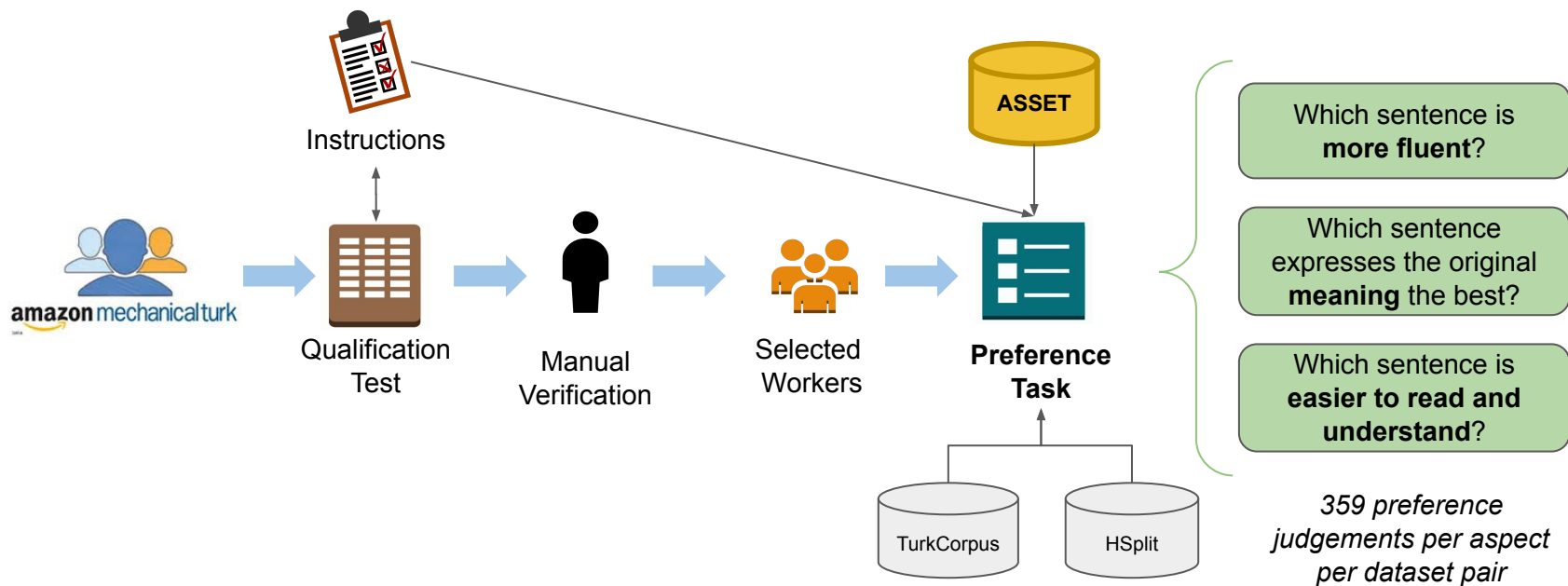


Sentence Splitting



2. Quality of Simplifications in ASSET

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



2. Quality of Simplifications in ASSET

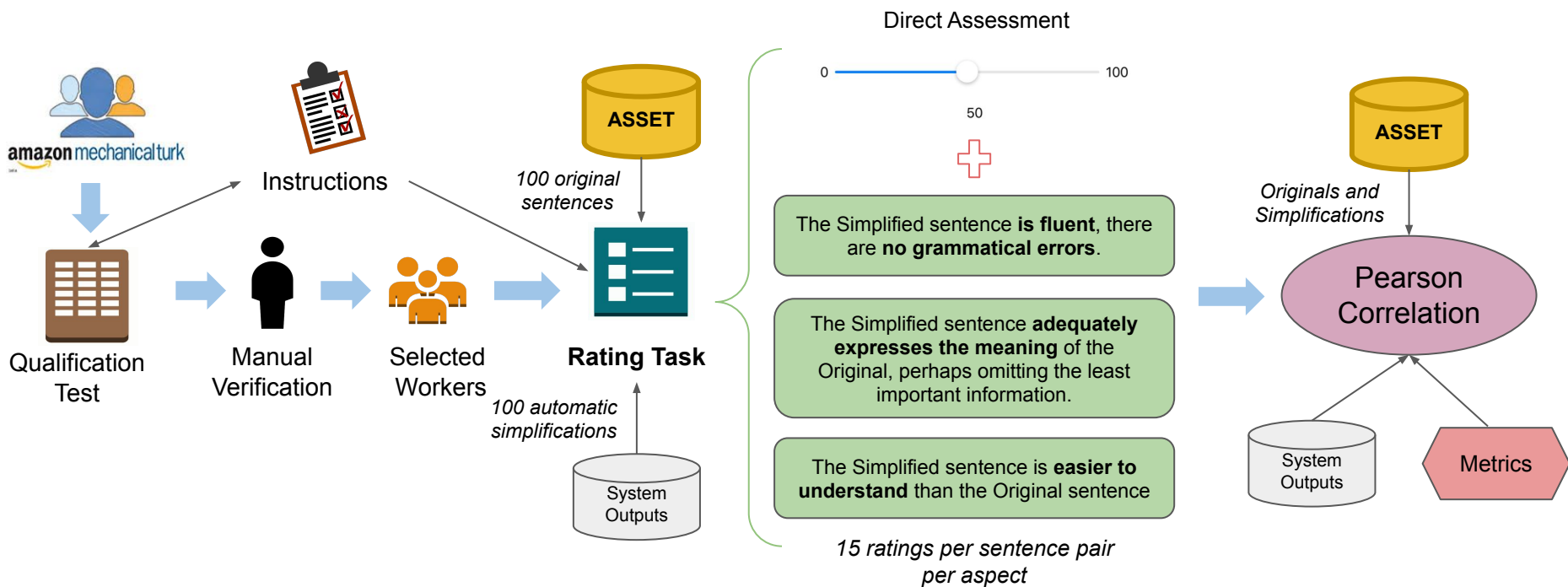
- **ASSET's** simplifications are preferred (or similar) in terms of **fluency** and **simplicity** over TurkCorpus or HSplit
- Simplifications from **TurkCorpus** or **HSplit** are more meaning preserving
 - Compression was not allowed when creating simplifications

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

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

3. ASSET for Automatic Evaluation

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



3. 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!

Summary

- ASSET is a **new multi-reference dataset** for tuning and evaluation of Sentence Simplification in English
- Simplifications in ASSET contain **multiple rewriting transformations**

| | Dev | Test |
|----------------------------------|---------------|--------------|
| Original Sentences | 2,000 | 359 |
| Simplification References | 10 | 10 |
| Total | 20,000 | 3,590 |

- Simplifications in ASSET are judged **simpler than those in other evaluation corpora**
 - Aspects: Fluency, Meaning Preservation, Simplicity

| | Preference Judgements |
|----------------------------|----------------------------------|
| ASSET vs TurkCorpus | 359 sents * 3 aspects = 1,077 |
| ASSET vs HSplit | 359 sents * 3 aspects = 1,077 |
| Total | 2,154 |

Summary

- **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**

| | |
|---------------------------|---|
| Original Sentences | 100 |
| System Outputs | 100 |
| Aspects Evaluated | Fluency Meaning Preservation Simplicity |
| Score Type | 0-100 continuous scale |
| Ratings Collected | 15 per aspect |
| Total | 100 segments * 15 ratings * 3 aspects = 4,500 |

Thank you!

ASSET is publicly available in: <https://github.com/facebookresearch/asset>

Fernando
Alva-Manchego



Louis
Martin



Antoine Bordes



Carolina Scarton



Benoît Sagot



Lucia Specia