

Sentence simplification with core vocabulary

The purpose

Performing sentence simplification beyond lexical substitution.

Abstract

We attempt automatic text simplification with vocabulary restriction on the output side using a machine translation approach based on a simplified corpus that we built. This approach makes it possible to simplify better than existing lexical substitution systems do.

What is “Simplification” ?

A task that reduces the complexity while preserving the main meaning.

What is “Core Vocabulary” ?

2,000 words selected by ourselves (Excluding proper nouns and symbols).

Simplification corpus

- Use of only “core vocabulary”.
- 50k Japanese sentences by hand.

Experiment

- We try automatic simplification using neural machine translation (NMT).
- We compare the machine translation approach with the lexical substitution (LS) system (Kajiwara and Yamamoto 2015).

Result and Discussion

- LS hardly changes the input sentence. Because of this tendency, we think that LS has a high score in grammaticality and meaning preservation, but SARI score is low.
- Although NMT changes the input sentence, the output sentence has sufficient results in grammaticality and meaning preservation.

Conclusions

- Japanese simplification corpus with core vocabulary is built.
- Simplification with NMT model outperforms existing Japanese lexical simplification system.

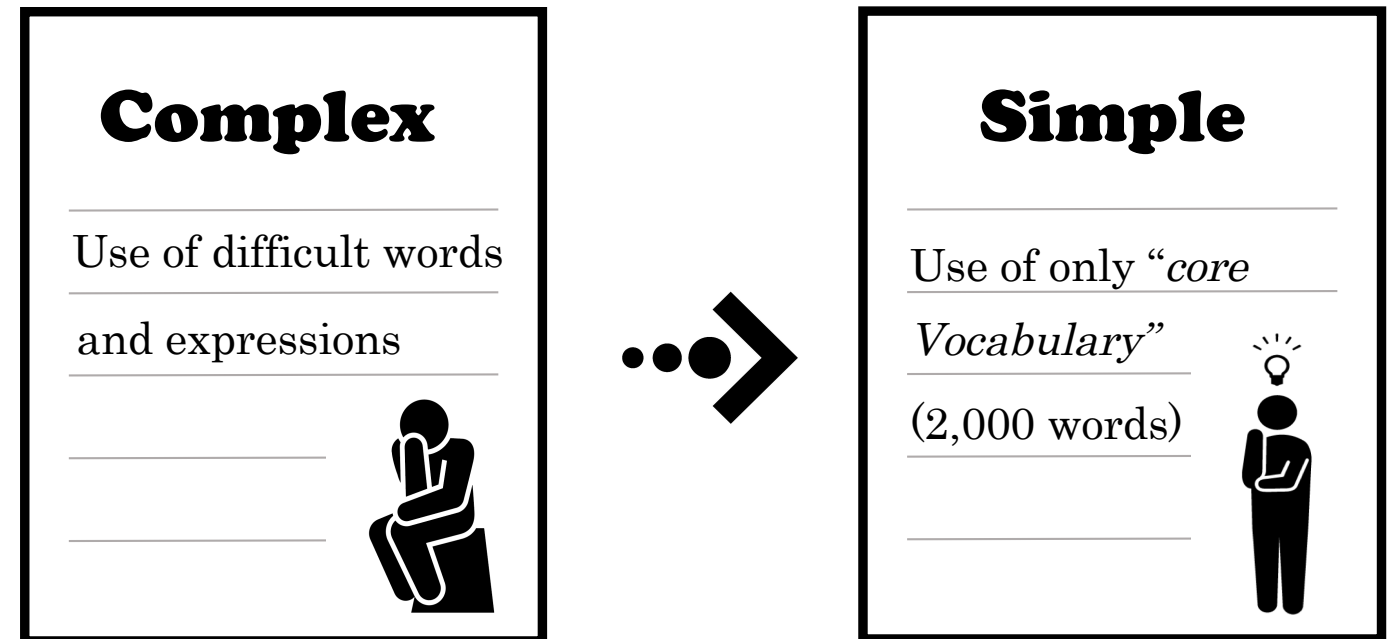


TABLE 1 EVALUATION

| System | SARI * | Simplicity * | Grammaticality * | Meaning * preservation |
|--------|--------------|--------------|------------------|------------------------|
| LS | 0.335 | 0.932 | 3.45 | 3.37 |
| NMT | 0.585 | 0.998 | 3.11 | 3.04 |

- * SARI: An arithmetic average of n-gram precision and the recall of rewrite operations.
- * Simplicity: Proportion that core vocabulary is included.
- * Grammar and Meaning preservation: Scoring 4 stages (1: bad, 4: good) by 5 evaluators.

TABLE 2 EXAMPLE OF OUTPUT

| Input sentence | Output sentence |
|---------------------|----------------------------|
| そこに署名してください。 | LS: そこに署名してください。 |
| | NMT: そこに名前を書いてください。 |
| 彼は大家族を養うために懸命に働いた。 | LS: 彼は大家族を養うために懸命に働いた。 |
| | NMT: 彼は大家族を支えるために頑張って働いた。 |
| 彼は人生に確固とした目的を持っている。 | LS: 彼は人生に強いとした目的を持っている。 |
| | NMT: 彼は人生にしっかりとした目的を持っている。 |