JADES: New Text Simplification Dataset in Japanese Targeted at Non-Native Speakers

Akio Hayakawa[1] **Tomoyuki Kajiwara**[2] **Hiroki Ouchi**[1,3] **Taro Watanabe**[1]

[1]Nara Institute of Science and Technology

[2]Ehime University [3]RIKEN AIP

{akio.hayakawa.gv6, hiroki.ouchi, taro}@is.naist.jp

kajiwara@cs.ehime-u.ac.jp

Abstract

The user-dependency of Text Simplification makes its evaluation obscure. A targeted evaluation dataset clarifies the purpose of simplification, though its specification is hard to define. We built JADES (JApanese Dataset for the Evaluation of Simplification), a text simplification dataset targeted at non-native Japanese speakers, according to public vocabulary and grammar profiles. JADES comprises 3,907 complex-simple sentence pairs annotated by an expert. Analysis of JADES shows that wide and multiple rewriting operations were applied through simplification. Furthermore, we analyzed outputs on JADES from several benchmark systems and automatic and manual scores of them. Results of these analyses highlight differences between English and Japanese in operations and evaluations.

1 Introduction

Text Simplification (TS) aims to rewrite texts for easier understanding. Simplified texts can benefit children (Smith et al., 1989), non-native speakers (Paetzold and Specia, 2016), non-specialists (Devaraj et al., 2021; Ivchenko and Grabar, 2022), and people with cognitive disabilities (Rello et al., 2013; Alonzo et al., 2020).

Given the diverse users in various domains, automatic TS has been regarded as an important research area these years (Alva-Manchego et al., 2020b). However, the diversity, in turn, makes the evaluation of TS obscure. As Xu et al. (2015) stated, an appropriate simplification for one type of users will not be appropriate for another. Therefore, the ideal TS system and evaluation is user-dependent, but its specification is difficult to define.

One step to the user-dependent TS could be focusing on a specific population. The validity of the simplification for a specific population can be evaluated using a targeted dataset. Newsela (Xu et al., 2015), available in English and Spanish, can be used in this way when using information about targeted grades on each article. Japanese lacks such a dataset. SNOW (Maruyama and Yamamoto, 2018; Katsuta and Yamamoto, 2018) is a Japanese dataset for TS and limits vocabulary, which comprises the top 2000 words required to understand Japanese. However, this criteria differs from targeting in that no specific populations are considered. For instance, they gave no simplification instructions on grammar to annotators. With a strictly limited vocabulary, this settings causes lengthy expressions. In addition, SNOW is problematic in that its original sentences are already short and simple. Therefore, SNOW may not be suitable for simplifying daily texts such as news articles.

Building a targeted dataset requires criteria for specific populations. In Japanese, Japanese-Language Proficiency Test (JLPT) published vocabulary and grammar profiles for grasping Japanese on each level (Japan-Foundation, 2002). These materials alleviate the difficulties in defining the specification and building a targeted dataset.

In this paper, we introduce a new Japanese TS dataset, **JADES**[1] (**JA**panese **D**ataset for the **E**valuation of **S**implification). JADES is targeted at non-native Japanese speakers capable of everyday communications, following the specification of vocabulary and grammar. JADES comprises 3,907 complex-simple parallel sentence pairs, which an expert of Easy Japanese manually simplifies. Since obtaining manual simplification are costly, JADES is oriented towards tuning and evaluation in size.

We also implemented models as baselines on JADES and rated their outputs automatically and manually. The contributions of this work include: (1) a dataset for TS in Japanese targeted at non-native speakers; (2) analysis of complex-simple text pairs in Japanese; (3) manual scores on simplified sentences.

[1]Our dataset will be available at http://github.com/naist-nlp/jades

Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022), pages 179 - 187
December 8, 2022 ©2022 Association for Computational Linguistics

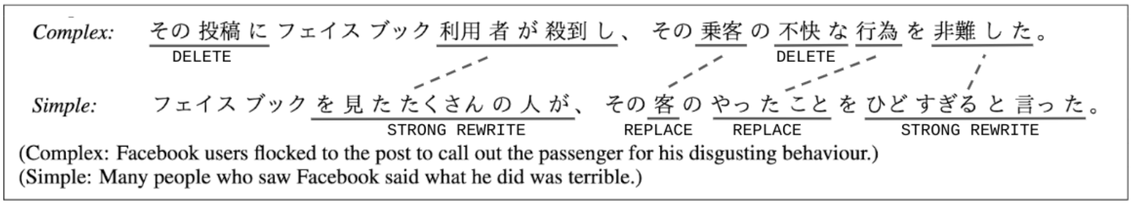

Figure 1: An example sentence pair in JADES with simplification operations.

2 Related Work

2.1 Simplification Dataset for Evaluation

While early works on TS use a subset of a large corpus for evaluation, Xu et al. (2015) pointed out the low quality of automatically aligned sentence pairs. Based on this report, using human-made sentence pairs has become a standard practice for TS evaluation. TurkCorpus (Xu et al., 2016) and ASSET (Alva-Manchego et al., 2020a) are standard datasets for the task comprising multiple reference sentences created by crowdworkers.

Although crowd-sourcing diversifies reference sentences, complex instructions can be difficult for crowdworkers. On the other hand, sentence pairs in Newsela are more valuable in that simplification is done by experts under reliable criteria for multiple levels, though its details are not disclosed. It should be noted that sentence pairs in Newsela dataset are automatically aligned and contains some misalignments.

2.2 Japanese Text Simplification

In addition to works on the typical lexical simplification (Kajiwara and Yamamoto, 2015; Hading et al., 2016), there have been several works for TS in Japanese. Goto et al. (2015) analyzed simplified Japanese news sentences and revealed that more than half of the sentences were reordered through simplification. Kato et al. (2020) focused on simplifying Japanese sentence-ending predicates, which are usually a source of confusion to readers due to their complexity. Each of these works built a dataset for training and evaluation, but unfortunately, they are not publicly available.

For a publicly available corpus, SNOW T15 (Maruyama and Yamamoto, 2018) and SNOW T23 (Katsuta and Yamamoto, 2018) are the largest in size (In the following, we denote them together as SNOW). SNOW extracted 84,400 sentences from Tanaka Corpus[2] and were manually simplified by non-experts. Original sentences in SNOW are mainly from textbooks, and the lengths of those are no more than 16 words, shorter than typical sentences in news and articles.

3 Our New Dataset JADES

We create a new dataset, JADES, for TS in Japanese. JADES contains manually simplified sentences targeted at independent non-native Japanese speakers. JADES comprises 3,907 complex-simple parallel sentence pairs and will help tune and evaluate TS models.

3.1 Simplification Criteria

To build a targeted dataset, we set criteria for the difficulty of simplified sentences. We chose former Level 3 of JLPT as a target level and adopted its vocabulary and grammar profiles as the criteria. Non-native speakers on this level are supposed to understand basic Japanese for everyday communication, which is almost equivalent to CEFR B1. The vocabulary profile contains 1,409 words, but we allowed using named entities and words at the same level as those in the profile. The grammar profile contains basic conjugations and sentence patterns. There is also a profile about Kanji characters, but we ignored it because rewriting Kanji to Hiragana or Katakana can cause misses in tokenization.

Since simplifying sentences based on the strict criteria can require some expertise, we employed an external person with specialized knowledge of Japanese simplification as an annotator. The annotator was asked to simplify sentences according to the criteria and exclude fairly simple sentences with no need for simplification. We asked the annotator to preserve the meaning of sentences on simplification but allowed deletion and addition of words for easier understanding of the main idea of sentences.

[2] http://www.edrdg.org/wiki/index.php/Tanaka_Corpus

	SNOW	JADES
# Sentences	84,400	3,907
# Vocab		
complex	2,610	12,382
simple	1,607	5,633
Avg. # Tokens		
complex	10.89	32.09
simple	11.99	31.51
Avg. Compression Rate	111.03	101.14
% of Identical	25.55	0.00

Table 1: Statistics of SNOW and JADES. Compression rate is calculated by # tokens of a simple sentence / # tokens of a complex sentence.

	N	S	J
REPLACE	38.4	80.0	97.0
SIMPLE REWRITE	26.0	30.0	76.0
DELETE	40.0	3.0	68.0
STRONG REWRITE	11.2	19.0	61.0
ADD	20.0	25.0	31.0
REORDER	11.2	2.0	20.0
SPLIT	17.2	1.0	14.0

Table 2: % of sentences from SNOW (S) and JADES (J) in which each operations was performed. Result of Newsela (N) are extracted from Alva Manchego (2020).

3.2 Data Source

Complex sentences in JADES were originally extracted from the Japanese-English development and test subsets in WMT20 news translation task (Barrault et al., 2020). These subsets include 3,991 sentences, about half of them are originally Japanese, and the rest are manually translated.

Through simplification under the criteria in Section 3.1, we obtained 3,907 complex-simple sentence pairs. We split these pairs into 2,959/448/500 for a train/valid/test subset, respectively. As multiple sentences in this dataset are originally from a single article, we assigned multiple sentences from the same article to the same subset. Meanwhile, the annotator was asked to treat one sentence as independent of the other sentences.

3.3 Analysis of Corpora

Table 1 shows the statistics of sentences in SNOW and JADES. We tokenized sentences with Sudachi (Takaoka et al., 2018) and calculated the vocabulary size, the number of tokens, and compression rates. The compression rates were calculated by dividing the number of tokens of a complex sentence by that of a simple sentence. One major difference between these two is in the number of tokens, which can derive from the difference in the domain of the original text: SNOW is from textbooks, and JADES is from news articles. The difference is also apparent in the vocabulary size as JADES contains broader topics and many named entities. The ratio of identical simplification, namely the exact match between a complex and a simple sentence, indicates that complex sentences in SNOW are fairly simple already.

We also analyzed how sentences were rewritten. The guideline for Easy Japanese[3], which is a guideline for simplifying Japanese texts published by the Japanese government, includes lexical simplification, syntactic simplification, deletion, and splitting, similar to well-discussed simplification operations in English (Xu et al., 2015; Alva Manchego, 2020). We manually identified the simplification operations applied to the original sentence from each randomly picked 100 sentence pair from SNOW and JADES, excluding identical pairs. We considered seven major simplification operations from Alva Manchego (2020), including `DELETE`, `ADD`, `SPLIT`, `REPLACE`, `SIMPLE REWRITE`, `STRONG REWRITE`, and `REORDER`.

The result of manual operation identification in Table 2 indicates that the majority of sentence pairs in JADES have multiple operations. On the other hand, only a few sentence pairs in SNOW have deletion and splitting since sentences are short in length. Compared to Newsela, SNOW and JADES include much more `REPLACE`, which can derive from the vocabulary limitation. On the other hand, JADES include outstanding number of `SIMPLE REWRITE`s and `STRONG REWRITE`s, which implies the large difference in simplicity between sentence pairs. See Appendix A for examples of simplification and operations.

4 Evaluations

We conducted TS in Japanese with several models to investigate the characteristics of our dataset. We also evaluated models automatically and manually.

4.1 Baseline Models

We chose BART (Lewis et al., 2020) and EditNTS (Dong et al., 2019) for model architectures and trained them with sentences from SNOW and JADES.

For BART, we built three models by fine-tuning

[3] https://www.bunka.go.jp/seisaku/kokugo_nihongo/kyoiku/92484001.html

System	Model Name	Train / Fine-tune		Automatic						Manual		
		SNOW	JADES	SNOW			JADES			JADES		
				BLEU	SARI	BS	BLEU	SARI	BS	F	M	S
Reference	`Reference`	-	-	-	-	-	-	-	-	**3.17**	**3.09**	**2.45**
Identical	`Identical`	-	-	57.27	22.58	89.59	29.10	16.27	80.93	-	-	-
BART	`BART-S`	✓	-	**75.85**	**61.06**	**93.25**	26.34	41.21	80.75	-	-	-
	`BART-J`	-	✓	55.58	42.85	88.76	32.50	49.97	82.81	-	-	-
	`BART-SJ`	✓	✓	68.14	59.59	90.18	**36.03**	**58.12**	**83.90**	2.84	2.38	1.95
EditNTS	`EditNTS-S`	✓	-	59.60	47.28	88.97	25.05	36.67	78.38	-	-	-
	`EditNTS-SJ`	✓	✓	51.79	46.86	86.92	30.52	44.30	80.78	2.76	2.36	1.45

Table 3: Automatic and manual evaluation on simplified sentences. In SNOW, multiple references were used for evaluation. F, M, and S stand for Fluency, Meaning Preservation, and Simplicity, respectively.

the Japanese pre-trained model[4]. Two of them were fine-tuned on SNOW and JADES, respectively, and the other was first fine-tuned on SNOW and then fine-tuned again on JADES. For EditNTS, we built two models. One was trained only on SNOW from scratch, and the other was then fine-tuned on JADES. We used the first 80,000 sentence pairs as a training set in SNOW. All models used JADES for their validation.

Subsequently, we generated simplified sentences on the test subset of JADES and SNOW.

In addition to these TS models, we set the identical system, which outputs the input sentences exactly as they are.

4.2 Automatic and Manual Evaluation

Since there are few discussions on suitable automatic metrics for TS in Japanese, we evaluated the outputs of the baseline models with the most commonly used metrics in TS, BLEU (Papineni et al., 2002), SARI (Xu et al., 2016), and BERTScore (Zhang et al., 2020).

In addition to automatic scores, we assessed simplified sentences on JADES by manual scores. We randomly sampled 600 simplified sentences on each of the valid and test subsets, from reference, BART, and EditNTS. We chose the best BART and EditNTS model by automatic evaluation.

Following (Alva-Manchego et al., 2020b), sampled sentences are scored on fluency, meaning preservation, and simplicity. We hired six in-house native Japanese speakers as annotators and asked them to score 300 sentences each from valid and test subsets, respectively. As a result, each sampled sentence was scored by three annotators. Scoring was based on 1-4 Likert scale; see Appendix B for detailed instruction.

[4]`https://github.com/utanaka2000/fairseq/blob/japanese_bart_pretrained_model/JAPANESE_BART_README.md`

	F	M	S
w/ `Reference`			
BLEU	0.308	0.411	0.459
SARI	0.300	0.385	0.493
BERTScore	0.335	0.429	0.474
w/o `Reference`			
BLEU	0.167	0.228	0.177
SARI	0.157	0.142	0.294
BERTScore	0.230	0.275	0.246

Table 4: Correlation between automatic and manual scores on valid/test subsets of JADES.

4.3 Results

Table 3 shows the automatic and manual evaluation of simplified sentences as well as identical outputs. On both SNOW and JADES, fine-tuned BART models are superior to EditNTS models. The best model among BART differs in a test dataset. `BART-SJ` outperforms the other models for JADES and is slightly inferior to `BART-S` for SNOW. This performance implies that two-step fine-tuning works even though the first dataset is rough to some extent.

For manual scores, `BART-SJ` seems able to generate fluent sentences, but lacks the ability to simplify sentences compared to `Reference`. Meanwhile, even `Reference` shows lower scores than expected on simplicity. This result may be because simplification rewritings sometimes euphemize expressions and vocabulary, which are easy to understand for native speakers. Thus, scores by targeted audiences will differ from scores by native Japanese speakers. We calculated Cohen's κ (Cohen, 1960) between each pair of annotators and took the weighted average. κ on fluency, meaning preservation, and simplicity is 0.255, 0.231, and 0.250, respectively. All these values can be assumed as fair agreement (Landis and Koch, 1977).

We also calculated Pearson's correlation coefficients between automatic and manual scores, shown in Table 4. Since `Reference` almost always gains perfect automatic scores, correlation is calculated with and without `Reference` sentences. Although correlations are not much high in all aspects, notably without `Reference`, BERTScore shows the highest correlation in fluency and meaning preservation, while SARI shows the highest in simplicity. The function of SARI differs from Alva-Manchego et al. (2021), which shows that SARI is inferior to BLEU and BERTScore on simplicity for the evaluation of multi-operational simplification.

Focusing on absolute values of automatic scores, `Identical` gains a low score for JADES and a high score for SNOW, which supports that JADES has drastic rewriting. For SARI, although `BART-SJ` and `EditNTS-SJ` show low manual scores, their SARI scores are quite high compared to the fact that state-of-the-art TS models in English, which can outperform even humans, gain just around 45 in SARI (Martin et al., 2020). We found these differences in evaluation between English and Japanese datasets, and will leave them to further research.

5 Conclusion

We have introduced JADES, a new dataset mainly for the evaluation of TS in Japanese. Simplified sentences in JADES are targeted at non-native speakers and made by an expert. This setting may make operations more variant and induce drastic rewriting, as the manual operation identification shows.

We can see the difference between English and Japanese in operations and evaluation metrics, which emphasizes the need for manual datasets in diverse languages.

Since manual scores on automatic TS models are low, TS in Japanese still has room for growth. With manual scores, JADES can also be useful for investigating new evaluation metrics. We believe that JADES facilitates TS in Japanese and its application.

Limitation

JADES has only one reference sentences, which might introduce some biases in simplified sentences since they are created by a single annotator. The heavy workload and quantity of annotation might also impact the overall quality. However, only a few annotators have the expertise to handle such a difficult targeting task. In order to mitigate the current limitations of this work, we are planning to investigate better instructions with detail granularities so that it is easier to expand this task with more annotators. Furthermore, the current dataset is limited in that the qualities are not double-checked by the actual targeted users, and we will leave it as our future studies.

Acknowledgements

The work of H. Ouchi was supported by JSPS KAKENHI grant number 19K20351 and NAIST Foundation.

References

Oliver Alonzo, Matthew Seita, Abraham Glasser, and Matt Huenerfauth. 2020. Automatic text simplification tools for deaf and hard of hearing adults: Benefits of lexical simplification and providing users with autonomy. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–13.

Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. 2020a. ASSET: A dataset for tuning and evaluation of sentence simplification models with multiple rewriting transformations. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4668–4679, Online. Association for Computational Linguistics.

Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2020b. Data-driven sentence simplification: Survey and benchmark. *Computational Linguistics*, 46(1):135–187.

Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2021. The (un)suitability of automatic evaluation metrics for text simplification. *Computational Linguistics*, 47(4):861–889.

Fernando Emilio Alva Manchego. 2020. *Automatic Sentence Simplification with Multiple Rewriting Transformations*. Ph.D. thesis, University of Sheffield.

Loïc Barrault, Magdalena Biesialska, Ondřej Bojar, Marta R. Costa-jussà, Christian Federmann, Yvette Graham, Roman Grundkiewicz, Barry Haddow, Matthias Huck, Eric Joanis, Tom Kocmi, Philipp Koehn, Chi-kiu Lo, Nikola Ljubešić, Christof Monz, Makoto Morishita, Masaaki Nagata, Toshiaki Nakazawa, Santanu Pal, Matt Post, and Marcos Zampieri. 2020. Findings of the 2020 conference on machine translation (WMT20). In *Proceedings of the Fifth Conference on Machine Translation*, pages 1–55, Online. Association for Computational Linguistics.

Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

Ashwin Devaraj, Iain Marshall, Byron Wallace, and Junyi Jessy Li. 2021. Paragraph-level simplification of medical texts. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4972–4984, Online. Association for Computational Linguistics.

Yue Dong, Zichao Li, Mehdi Rezagholizadeh, and Jackie Chi Kit Cheung. 2019. EditNTS: An neural programmer-interpreter model for sentence simplification through explicit editing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3393–3402, Florence, Italy. Association for Computational Linguistics.

Isao Goto, Hideki Tanaka, and Tadashi Kumano. 2015. Japanese news simplification: tak design, data set construction, and analysis of simplified text. In *Proceedings of Machine Translation Summit XV: Papers*, Miami, USA.

Muhaimin Hading, Yuji Matsumoto, and Maki Sakamoto. 2016. Japanese lexical simplification for non-native speakers. In *Proceedings of the 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLPTEA2016)*, pages 92–96, Osaka, Japan. The COLING 2016 Organizing Committee.

Oksana Ivchenko and Natalia Grabar. 2022. Impact of the text simplification on understanding. *Studies in health technology and informatics*.

The Japan-Foundation. 2002. Japanese-language proficiency test: Test content specification (revised edition).

Tomoyuki Kajiwara and Kazuhide Yamamoto. 2015. Evaluation dataset and system for Japanese lexical simplification. In *Proceedings of the ACL-IJCNLP 2015 Student Research Workshop*, pages 35–40, Beijing, China. Association for Computational Linguistics.

Taichi Kato, Rei Miyata, and Satoshi Sato. 2020. BERT-based simplification of Japanese sentence-ending predicates in descriptive text. In *Proceedings of the 13th International Conference on Natural Language Generation*, pages 242–251, Dublin, Ireland. Association for Computational Linguistics.

Akihiro Katsuta and Kazuhide Yamamoto. 2018. Crowdsourced corpus of sentence simplification with core vocabulary. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7871–7880, Online. Association for Computational Linguistics.

Louis Martin, Angela Fan, Éric de la Clergerie, Antoine Bordes, and Benoît Sagot. 2020. Muss: Multilingual unsupervised sentence simplification by mining paraphrases.

Takumi Maruyama and Kazuhide Yamamoto. 2018. Simplified corpus with core vocabulary. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Gustavo Paetzold and Lucia Specia. 2016. Understanding the lexical simplification needs of non-native speakers of English. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 717–727, Osaka, Japan. The COLING 2016 Organizing Committee.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Luz Rello, Ricardo Baeza-Yates, Stefan Bott, and Horacio Saggion. 2013. Simplify or help? text simplification strategies for people with dyslexia. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, pages 1–10.

Suzanne T Smith, Paul Macaruso, Donald Shankweiler, and Stephen Crain. 1989. Syntactic comprehension in young poor readers. *Applied psycholinguistics*, 10(4):429–454.

Kazuma Takaoka, Sorami Hisamoto, Noriko Kawahara, Miho Sakamoto, Yoshitaka Uchida, and Yuji Matsumoto. 2018. Sudachi: a Japanese tokenizer for business. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).

Wei Xu, Chris Callison-Burch, and Courtney Napoles. 2015. Problems in current text simplification research: New data can help. *Transactions of the Association for Computational Linguistics*, 3:283–297.

Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. 2016. Optimizing statistical machine translation for text simplification.

Transactions of the Association for Computational Linguistics, 4:401–415.

Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q. Weinberger, and Yoav Artzi. 2020. Bertscore: Evaluating text generation with bert. In *International Conference on Learning Representations*.

A Simplification Example

Example 1.

Complex: その 投稿 に フェイス ブック 利用 者 が 殺到 し、 その 乗客 の 不快 な 行為 を 非難 し た。

DELETE DELETE

Simple: フェイス ブック を 見 た たくさん の 人 が、 その 客 の やっ た こと を ひど すぎる と 言っ た。

STRONG REWRITE REPLACE REPLACE STRONG REWRITE

(Complex: Facebook users flocked to the post to call out the passenger for his disgusting behaviour.)
(Simple: Many people who saw Facebook said what he did was terrible.)

Example 2.

Complex: 昨年、 67000 名 以上 の アメリカ 人 が 過剰 摂取 に より 死亡 し た。

Simple: 去年、 67000 人 以上 の アメリカ 人 が 悪い 薬 を 飲み すぎ て 死ん だ。

REPLACE REPLACE ADD REPLACE SIMPLE REWRITE REPLACE

(Complex: Last year, more than 67,000 Americans died from an overdose.)
(Simple: Last year, more than 67,000 Americans died from taking too many bad drugs.)

Figure 2: Examples of simplification in JADES.

B Instruction for Manual Scoring

We set 1-4 Likert scale for each of fluency, meaning preservation, and simplicity. Below is a translation of the specific instructions.

Fluency

- (Presenting a simplified sentence) Is the following sentence fluent?
  1. Obviously not fluent
  2. Lack in fluency, and the main idea is hard to understand
  3. Slightly less fluent, but conveys the main idea
  4. Fluent

Meaning Preservation

- (Presenting an original sentence as A and a simplified sentence as B) How much of the meaning of sentence A is retained in sentence B?
  1. Hardly retained (the main idea completely changed)
  2. Not much retained (the main idea changed somewhat)
  3. Largely retained (the main idea is retained)
  4. Almost completely retained

Simplicity

- (Presenting an original sentence as A and a simplified sentence as B) Is sentence B is easier to understand compared to sentence A?
  1. Harder to understand
  2. Almost equal
  3. Slightly easier to understand
  4. Easier to understand