CSSWiki: A Chinese Sentence Simplification Dataset with Linguistic and Content Operations

Fengkai Liu, John S. Y. Lee
Department of Linguistics and Translation
City University of Hong Kong
Hong Kong SAR, China
fengkaliu3-c@my.cityu.edu.hk, jsylee@cityu.edu.hk

Abstract

Sentence Simplification aims to make sentences easier to read and understand. With most effort on corpus development focused on English, the amount of annotated data is limited in Chinese. To address this need, we introduce CSSWiki, an open-source dataset for Chinese sentence simplification based on Wikipedia. This dataset contains 1.6k source sentences paired with their simplified versions. Each sentence pair is annotated with operation tags that distinguish between linguistic and content modifications. We analyze differences in annotation scheme and data statistics between CSSWiki and existing datasets. We then report baseline sentence simplification performance on CSSWiki using zero-shot and few-shot approaches with Large Language Models.

Keywords: Chinese sentence simplification; Corpus creation; Linguistic simplification operations; Content simplification operations

1. Introduction

The Sentence Simplification (SS) task aims to increase the readability of a sentence and make it more accessible for readers, while preserving the original main idea and meaning (Alva-Manchego et al., 2020). Text simplification can assist individuals with reading disabilities (Carroll et al., 1998; Rello et al., 2013; Evans et al., 2014), language learners (Watanabe et al., 2009) and non-native speakers (Paetzold, 2016).

Recent automatic SS systems mostly leverage data-driven and deep learning methods (Feblowitz and Kauchak, 2013; Xu et al., 2016; Aharoni and Goldberg, 2018; Alva-Manchego et al., 2017). Previous studies have primarily focused on English, based on datasets such as ASSET (Alva-Manchego et al., 2020), ASSET$_{ann}$ (Cardon et al., 2022) and Newsela (Xu et al., 2015). Less attention has been given to other languages such as Chinese. To the best of our knowledge, there is only one publicly available dataset for Chinese SS, namely, CSS (Yang et al., 2023).[1] However, it neither directly annotates the text spans that are edited (Cardon et al., 2022), nor explicitly indicates content modifications such as adding clarifications or cutting information (Srikanth and Li, 2021; Rets et al., 2022).

This paper introduces CSSWiki (**C**hinese **S**entence **S**implification from **Wiki**pedia)[2], which not only offers a larger number of source sentences than CSS, but is also the first open-source Chinese SS dataset with annotations that distinguish between linguistic and content modifications. After a literature review (Section 2), we present our annotation scheme (Section 3). We then describe the corpus construction process (Section 4) and then highlight the distinctions between CSSWiki and existing SS datasets (Section 5). Finally, to provide baselines for future research, we report the performance of a number of Large Language Models (LLMs) in zero-shot and few-shot settings on our dataset (Section 6).

[1] https://github.com/maybenotime/CSS

[2] Publicly released at https://github.com/ffliu6/CSSWiki

2. Previous Work

2.1. Sentence Simplification

Sentence simplification (SS) has been a rapidly growing subfield of natural language processing (NLP) in recent years. While earlier approaches mostly relied on rule-based methods (Santaholma, 2007; Zhu et al., 2010), recent studies have shown that data-driven approaches can improve simplification performance (Zhang and Lapata, 2017; Martin et al., 2020). They are supported by large amounts of labeled complex-simple sentence pairs, with most SS datasets focused on English (Alva-Manchego et al., 2017; Zhang and Lapata, 2017; Martin et al., 2020).

Unsupervised methods have also been proposed for low-resource languages (Kajiwara and Komachi, 2018; Palmero Aprosio et al., 2019; Kumar et al., 2020). UNTS (Surya et al., 2019) obtained competitive performance against the supervised models, and MUSS (Martin et al., 2022) outperformed the state-of-the-art supervised methods in French and Spanish. Large Language Models (LLMs) have demonstrated excellent performance

LREC-COLING 2024, pages 4205–4213
20-25 May, 2024.

	CSSWiki	CSS	$ASSET_{ann}$	HSplit	TurkCorpus
# Source Sentences	1,600	383	359	359	2,359
# References per source sentence	3	2	10	4	8
Tokens per reference	40.11	47.29	19.04	25.49	21.29
Content Operations	✓	✗	✗	✗	✗
Linguistic Operations	✓	✗	✓	✗	✗
Operation Tags	✓	✓	✓	✗	✗

Table 1: Comparison between CSSWiki and other sentence simplification datasets.

in a variety of NLP tasks under both zero-shot and few-shot settings (Brown et al., 2020; Chowdhery et al., 2022; Feng et al., 2023).

2.2. Annotation design

The Parallel Wikipedia Simplification (PWKP) corpus (Zhu et al., 2010), harvested from Wikipedia and Simple Wikipedia, does not annotate the transformation operations between a pair of complex and simple sentence. The ASSET dataset (Alva-Manchego et al., 2020) was developed to encompasses a variety of transformations, but also does not provide operation tags. The $ASSET_{ann}$ dataset (Cardon et al., 2022) augments ASSET by annotating a set of fine-grained linguistic operations. Different labels are given to insertions and deletions involving propositions, modifiers, for consistency and other purposes; as well as replacements involving synonyms, hypernyms, hyponyms, singular vs. plural, anaphora, and verbal features. More complex operations, including active vs. passive voice, part-of-speech change, impersonal vs. personal form, and affirmation vs. negation, are also identified. However, it does not differentiate between linguistic and content operations (Section 3).

The annotation scheme of CSS (Yang et al., 2023), the only existing open-source SS dataset for Chinese, consists of four operation tags: (1) lexical simplification, which includes synonym substitution and explanation of an idiom with a short sentence; (2) sentence splitting; (3) compression, i.e. deleting unimportant information; and (4) sentence paraphrasing, which includes reordering and syntactic transformations. Each sentence in the corpus is annotated with one or more of these tags, but the text spans are not specified.

3. Annotation Scheme

Our annotation scheme synthesizes insights from prior studies on manual simplification behaviors (Srikanth and Li, 2021; Rets et al., 2022; Brunato et al., 2022; Cardon et al., 2022). To further refine the taxonomy, we distinguish between "linguistic modification" and "content modification" (Rets et al., 2022). The former, which constitutes "form (lexico-syntactic) modifications", aims to facilitate understanding of the "literal meaning of the text". In contrast, the latter "mainly dealt with the implicit meaning and involved editing the information in the text" (Rets et al., 2022). Correspondingly, our scheme consists of two types of operation tags.

3.1. Linguistic operations

The linguistic operations include Substitution, Insertion, Deletion and Syntactic. Substitution refers to replacements involving synonyms (`SubS`), hypernyms (`SubH`), numbers (`SubN`), and pronouns (`SubP`). Table 2 shows an example of `SubS` in which the noun *dengxianzhe* 'idle people' is replaced with a simpler equivalent, *putongren* 'ordinary people'; as well as an example of Deletion in which the identity of the broadcaster (CCTV) is removed. Syntactic includes merging individual clauses within a sentence (`SynM`), splitting a sentence (`SynS`), and transforming passive voice into active forms (`SynV`).

3.2. Content operations

The two content operations are Elaboration (`Elab`) and Condensation (`Cond`). The example for `Elab` in Table 2 clarifies the idiomatic expression 青黄不接 *qinghuangbujie* 'yellow does not reach green' by expanding it to the implicit meaning, 'The crops are not yet mature, but there is no more food'. The simplified version thus unpacks the meaning of the expression with two new clauses. Given the significant changes in content, beyond form and vocabulary, the revision is not merely a substitution. The source sentence in the example for `Cond` contains references to difficult terms such as 'layout molds' and 'curved surfaces'. The simplified version compresses them into the straightforward term 'usage'. This transformation differs from both substitution and deletion, since it uses a simpler alternative to represent the underlying meaning.

4. Corpus construction

Eight Chinese native speakers in a Masters-level linguistics program were responsible for data col-

Operation	Example (in Chinese)	Translation
Synonym (SubS)	Src: [等闲者 |SubS] 不得入文渊阁。 Simp: [普通人 |SubS] 不得入文渊阁。	[idle people|SubS] could not enter Wenyuan Library [ordinary people|SubS] could not enter Wenyuan Library
Insertion (Ins)	Src: 土地国有 Simp: 土地 [变成 |Ins] 国有	The land was nationalized The land [became|Ins] nationalized
Deletion (Del)	Src: 港台电视转播 [央视的 |Del] 春节联欢晚会， Simp: 港台电视转播春节联欢晚会，	RTHK broadcasts [CCTV's|Del] Spring Festival Gala RTHK broadcasts the Spring Festival Gala
Split (SynS)	Src: [最古老但行之有效的 土电报 |SynS] Simp: [虽然是最古老的土电报，但非常有效 |SynS]	[the oldest but effective native telegraph|SynS] [although it is the oldest native telegraph, it is very effective.|SynS]
Elaboration (Elab)	Src: 早春时节来临后，[青黄不接 |Elab]... Simp: 早春时节来临后，[庄稼还没成熟，旧的粮食也吃完了 |Elab] ...	In early spring, [yellow does not reach green|Elab] In early spring, [the crops are not yet mature, but there is no more food|Elab]
Condensation (Cond)	Src: 这令奥利弗打字机 [在版面模子和曲面打字时 |Cond] 效果更好 Simp: 这令奥利弗打字机 [使用 |Cond] 效果更好	improves the performance [when typing on layout molds and curved surfaces|Cond] improves [usage|Cond] performance

Table 2: Annotation examples including linguistic operations (Synonym, Insertion, Deletion, Split) and content operations (Elaboration, Condensation)

lection and annotation. The annotators were asked to select four texts from Chinese Wikipedia[3] for which they had adequate background knowledge to perform simplification. After manual segmentation of the raw texts into sentences, each selected 200 sentences for simplification.

The annotators were asked to apply an initial annotation taxonomy to a subset of the 200 sentences. After reaching a consensus on refining the taxonomy, they provided sample sentences for each operation to establish the final scheme. They then annotated all sentences in the format shown in Table 2, where the edited text spans and their corresponding operation tags are enclosed in brackets. For each sentence, one other annotator independently performed the simplification. The two then discussed their strategies and produced a third version.

5. Corpus analysis

As shown in Table 1, CSSWiki contains 1,600 source sentences and a total of 4,800 source-simplified sentence pairs. Unlike previous datsets, CSSWiki distinguishes between linguistic and content operations during simplification. This additional information can aid in evaluating controlled SS systems and specific SS sub-tasks.

Table 3 displays the the distribution of the operations applied in the simplification process. As

Category	Freq.	Prop.	Avg. Freq. per sent.
Linguistic operations			
Substitution	**876**	**22.76%**	**0.548**
SubS	825	21.43%	0.516
SubH	15	0.39%	0.009
SubN	19	0.49%	0.012
SubP	17	0.44%	0.011
Insertion	**84**	**2.18%**	**0.052**
Deletion	**2,199**	**57.13%**	**1.374**
Syntactic	**153**	**3.98%**	**0.096**
SynM	132	3.43%	0.083
SynS	11	0.29%	0.007
SynV	10	0.26%	0.006
Content operations			
Content	**537**	**13.95%**	**33.56%**
Elab	152	3.95%	0.095
Cond	385	10.00%	0.241
Total	**3,849**	**100.00%**	**2.406**

Table 3: Distribution of operation tags, and the average frequency of each tag per sentence pair

expected, a considerable number of sentences in CSSWiki underwent simplification through deletion, accounting for 57.13%. On average, each sentence was modified approximately 2.406 times. Insertion constituted the smallest proportion in CSSWiki, comprising only 2.18%. As for content operations, elaborations accounted for 3.95%, which aligns with the findings of Srikanth and Li

[3] https://zh.wikipedia.org/

(2021) .

Following Yang et al. (2023), we analyze a series of surface and syntax-based features, including *Compression level*, *Proportion of words deleted, added and reordered*, *Word deletion only*, *Replace-only Levenshtein distance*, *Lexical complexity score ratio*, and *Dependency tree depth ratio.*[4]

As highlighted in Table 4, there are notable differences between CSSWiki and its peers in the *Compression level* and *Word deletion only* features. The proportion of *Word Deletion Only* is much higher than CSS and ASSET, indicating many extractive compression operations were performed. This is consistent with *Compression level*, with 38.59%. This phenomenon can be attributed to the genre of the source text. Consisting mostly of expository writing, the Chinese Wikipedia texts in our corpus differ significantly from the newspaper materials in CSS. The Chinese Wikipedia content includes more evidence-based descriptions aimed at conveying knowledge and explaining concepts. As a result, our annotators frequently removed descriptive words while preserving sentence meanings, such as adjectives modifying nouns and additional descriptions of previously explained concepts (see further details in Appendix A).

	CSSWiki	CSS	ASSET
Compression	38.59%	9.1%	31.2%
Word Reordered	20.26%	17.6%	28.3%
Word Deletion	26.95%	5.6%	4.5%

Table 4: Comparison between CSSWiki and other SS datasets on simplification features

6. Experiments

Large Language Models (LLMs) are increasingly used across a wide range of natural language processing tasks. To establish baseline results on CSSWiki, we experimented with two publicly available LLMs in zero- and few-shot settings.

6.1. Large Language Models

ChatGLM2-6B[5] As the second generation of the open-source, bilingual Chinese-English chat model ChatGLM-6B[6], this model can consider longer contexts and offer stronger performance. It follows a similar training process as ChatGPT and is specifically optimized for Chinese question-answering and dialogue tasks.

GPT-3.5-turbo-0613[7] This is a snapshot of GPT-3.5-turbo from 13 Jun 2023 and the replaced model for GPT-3.5-turbo-0301.[8]

6.2. Set-up

Recent experiments on zero- and few-shot learning with pre-trained language models have shown excellent performance on text simplification in a low-resource language (Mallinson et al., 2020; Feng et al., 2023). We performed evaluation on the full SS task and four other subtasks: SS with substitution only, linguistic operations only, elaboration only, and condensation only.

Evaluation Metrics. Following Yang et al. (2023), our evaluation metrics include BLEU (Sulem et al., 2018) and SARI (Alva-Manchego et al., 2019). For the latter, we report both SARI_{char} and SARI_{word}, i.e., SARI at the character level and word level, representing the two tokenization methods for Chinese.

Zero-shot. For the full SS task, we used the same prompt in Chinese as Yang et al. (2023) to request sentence simplification. We then tailored the prompt to the four subtasks using slightly different wording. All prompts are listed in Appendix B.

Few-shot. We randomly selected 20% of our dataset as development set. The zero-shot prompt was followed by 5 example sentence pairs, randomly selected from the development set, using the template shown in Appendix B. For the subtasks, the gold output sentence in the example sentence pairs included only the relevant operations; for example, gold outputs in the "linguistic operations only" subtask are derived without any content operations.

6.3. Results

Table 5 presents the experimental results.

Zero-shot. In the zero-shot setting, GPT-3.5-turbo outperformed ChatGLM2 in most tasks and metrics.[9]. The differences were slight, however, perhaps attributable to the similar training approaches employed for both models. Both models performed best in the Elaboration subtask in terms of both SARI and BLEU scores. They performed worst in the "Linguistic operations only" subtask[10], likely due to the more diverse nature of the simplification strategies in this category.

Few-shot. LLMs have been demonstrated to be able to learn from limited samples in various NLP

[4] Extracted with codes provided at https://github.com/maybenotime/CSS

[5] https://github.com/THUDM/ChatGLM2-6B

[6] https://github.com/THUDM/ChatGLM-6B

[7] https://platform.openai.com/docs/models/gpt-3-5

[8] This stable model was chosen to facilitate reproducibility.

[9] Except for BLEU score in the "Linguistic operations only" subtask

[10] Except for BLEU score with ChatGLM2.

Task / Subtask	Setting	ChatGLM2-6B			GPT-3.5-turbo-0613		
		$SARI_{char}$	$SARI_{word}$	BLEU	$SARI_{char}$	$SARI_{word}$	BLEU
Full SS	Zero-shot	63.30	62.39	47.30	64.21	63.12	52.32
	Few-shot	64.73	**63.44**	56.00	**64.77**	63.43	**61.74**
Substitution only	Zero-shot	67.96	65.55	43.65	68.92	66.34	45.44
	Few-shot	68.76	66.58	54.99	**70.48**	**68.16**	**64.81**
Linguistic operations only	Zero-shot	61.98	60.87	47.24	61.53	61.02	25.33
	Few-shot	63.35	61.94	**50.40**	**63.61**	**62.31**	48.91
Elaboration only	Zero-shot	70.16	69.32	48.30	72.30	70.94	54.16
	Few-shot	72.35	70.93	60.29	**73.25**	**71.57**	**63.42**
Condensation only	Zero-shot	66.84	66.77	42.83	69.52	69.28	52.68
	Few-shot	68.97	68.45	53.68	**69.96**	**69.34**	**63.24**

Table 5: Sentence simplification (SS) performance on the full SS task and other subtasks

tasks. Consistent with this trend, for both models, the few-shot setting led to better performance across all tasks and metrics. As in the zero-shot setting, GPT-3.5-turbo slightly outperformed ChatGLM2 in most tasks and metrics.[11] Both models continued to perform worse in the "Linguistic operations only" subtask, suggesting that more samples or fine-tuning might be desirable.

7. Conclusion

We have presented CSSWiki, a publicly available Chinese sentence simplification (SS) dataset, and compared it with other existing datasets. Our contribution is three-fold. First, with 1.6k source sentences taken from Wikipedia, this dataset can facilitate training and evaluation of Chinese SS models. Second, unlike existing datasets, its annotation scheme distinguishes between linguistic and content operations, which enables more fine-grained studies and evaluation on simplification strategies. Finally, using zero- and few-shot approaches with two Large Language Models, we have provided baseline results on the full SS task as well as several subtasks. In future work, we plan to expand the corpus and explore text-level simplification.

Acknowledgement

This work was partly supported by the Language Fund from the Standing Committee on Language Education and Research (project EDB(LE)/P&R/EL/203/14) and by the General Research Fund (project 11207320).

8. Bibliographical References

[11] Again, except for BLEU score in the "Linguistic operations only" subtask.

Roee Aharoni and Yoav Goldberg. 2018. Split and rephrase: Better evaluation and stronger baselines. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 719–724.

Fernando Alva-Manchego, Joachim Bingel, Gustavo Paetzold, Carolina Scarton, and Lucia Specia. 2017. Learning how to simplify from explicit labeling of complex-simplified text pairs. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 295–305, Taipei, Taiwan. Asian Federation of Natural Language Processing.

Fernando Alva-Manchego, Louis Martin, Carolina Scarton, and Lucia Specia. 2019. Easse: Easier automatic sentence simplification evaluation. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, pages 49–54.

Fernando Alva-Manchego, Carolina Scarton, and Lucia Specia. 2020. 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.

Stefan Bott, Horacio Saggion, and Simon Mille. 2012. Text simplification tools for spanish. In *LREC*, pages 1665–1671.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal,

Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

Dominique Brunato, Felice Dell'Orletta, and Giulia Venturi. 2022. Linguistically-based comparison of different approaches to building corpora for text simplification: A case study on italian. *Frontiers in Psychology*, 13:707630.

Dominique Brunato, Felice Dell' Orletta, Giulia Venturi, and Simonetta Montemagni. 2015. Design and annotation of the first italian corpus for text simplification. In *Proceedings of The 9th Linguistic Annotation Workshop*, pages 31–41.

Arnaldo Candido Jr, Erick Galani Maziero, Lucia Specia, Caroline Gasperin, Thiago Pardo, and Sandra Aluisio. 2009. Supporting the adaptation of texts for poor literacy readers: a text simplification editor for brazilian portuguese. In *Proceedings of the Fourth Workshop on Innovative Use of NLP for Building Educational Applications*, pages 34–42.

John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998. Practical simplification of english newspaper text to assist aphasic readers. In *Proceedings of the AAAI-98 Workshop on Integrating Artificial Intelligence and Assistive Technology*, pages 7–10. Association for the Advancement of Artificial Intelligence.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2022. Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311*.

Richard Evans, Constantin Orasan, and Iustin Dornescu. 2014. An evaluation of syntactic simplification rules for people with autism. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)@ EACL*, pages 131–140.

Dan Feblowitz and David Kauchak. 2013. Sentence simplification as tree transduction. In *Proceedings of the second workshop on predicting and improving text readability for target reader populations*, pages 1–10.

Yutao Feng, Jipeng Qiang, Yun Li, Yunhao Yuan, and Yi Zhu. 2023. Sentence simplification via large language models. *arXiv preprint arXiv:2302.11957*.

Núria Gala, Anaïs Tack, Ludivine Javourey Drevet, Thomas François, and Johannes C Ziegler. 2020. Alector: A parallel corpus of simplified french texts with alignments of misreadings by poor and dyslexic readers. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1353–1361.

Tomoyuki Kajiwara and Mamoru Komachi. 2018. Text simplification without simplified corpora. *The Journal of Natural Language Processing*, 25:223–249.

J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.

Dhruv Kumar, Lili Mou, Lukasz Golab, and Olga Vechtomova. 2020. Iterative edit-based unsupervised sentence simplification. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7918–7928.

Jonathan Mallinson, Rico Sennrich, and Mirella Lapata. 2020. Zero-shot crosslingual sentence simplification. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5109–5126.

Louis Martin, Éric de la Clergerie, Benoît Sagot, and Antoine Bordes. 2020. Controllable sentence simplification. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4689–4698, Marseille, France. European Language Resources Association.

Louis Martin, Angela Fan, Éric Villemonte De La Clergerie, Antoine Bordes, and Benoît Sagot. 2022. Muss: Multilingual unsupervised sentence simplification by mining paraphrases. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 1651–1664.

Gustavo Henrique Paetzold. 2016. *Lexical simplification for non-native english speakers*. Ph.D. thesis, University of Sheffield.

Alessio Palmero Aprosio, Sara Tonelli, Marco Turchi, Matteo Negri, and A Di Gangi Mattia. 2019. Neural text simplification in low-resource conditions using weak supervision. In *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation (NeuralGen)*, pages 37–44. Association for Computational Linguistics (ACL).

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In *Pro-*

ceedings of the 40th annual meeting of the Association for Computational Linguistics, pages 311–318.

Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013. Frequent words improve readability and short words improve understandability for people with dyslexia. In *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part IV 14*, pages 203–219. Springer.

Irina Rets, Lluisa Astruc, Tim Coughlan, and Ursula Stickler. 2022. Approaches to simplifying academic texts in english: English teachers' views and practices. *English for Specific Purposes*, 68:31–46.

Horacio Saggion, Sanja Štajner, Stefan Bott, Simon Mille, Luz Rello, and Biljana Drndarevic. 2015. Making it simplext: Implementation and evaluation of a text simplification system for spanish. *ACM Transactions on Accessible Computing (TACCESS)*, 6(4):1–36.

Marianne Santaholma. 2007. Grammar sharing techniques for rule-based multilingual nlp systems. In *Proceedings of the 16th Nordic Conference of Computational Linguistics (NODALIDA 2007)*, pages 253–260.

Neha Srikanth and Junyi Jessy Li. 2021. Elaborative simplification: Content addition and explanation generation in text simplification. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 5123–5137.

Sanja Štajner, Iacer Calixto, and Horacio Saggion. 2015. Automatic text simplification for spanish: Comparative evaluation of various simplification strategies. In *Proceedings of the international conference recent advances in natural language processing*, pages 618–626.

Elior Sulem, Omri Abend, and Ari Rappoport. 2018. BLEU is not suitable for the evaluation of text simplification. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 738–744, Brussels, Belgium. Association for Computational Linguistics.

Renliang Sun, Wei Xu, and Xiaojun Wan. 2023. Teaching the pre-trained model to generate simple texts for text simplification. *arXiv preprint arXiv:2305.12463*.

Sai Surya, Abhijit Mishra, Anirban Laha, Parag Jain, and Karthik Sankaranarayanan. 2019. Unsupervised neural text simplification. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2058–2068.

Sara Tonelli, Alessio Palmero Aprosio, and Francesca Saltori. 2016. Simpitiki: a simplification corpus for italian. In *CLiC-it/EVALITA*, pages 4333–4338.

Laura Vásquez-Rodríguez, Matthew Shardlow, Piotr Przybyła, and Sophia Ananiadou. 2021. The role of text simplification operations in evaluation. In *CEUR Workshop Proceedings*, volume 2944, pages 57–69. CEUR-WS.

Willian Massami Watanabe, Arnaldo Candido Junior, Vinícius Rodriguez Uzêda, Renata Pontin de Mattos Fortes, Thiago Alexandre Salgueiro Pardo, and Sandra Maria Aluísio. 2009. Facilita: reading assistance for low-literacy readers. In *Proceedings of the 27th ACM international conference on Design of communication*, pages 29–36.

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.

Shiping Yang, Renliang Sun, and Xiaojun Wan. 2023. A new dataset and empirical study for sentence simplification in Chinese. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8306–8321, Toronto, Canada. Association for Computational Linguistics.

Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1353–1361.

9. Language Resource References

Fernando Alva-Manchego, Louis Martin, Antoine Bordes, Carolina Scarton, Benoît Sagot, and Lucia Specia. 2020. 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.

Rémi Cardon, Adrien Bibal, Rodrigo Wilkens, David Alfter, Magali Norré, Adeline Müller, Watrin Patrick, and Thomas François. 2022. Linguistic corpus annotation for automatic text simplification evaluation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 1842–1866.

Louis Martin, Éric de la Clergerie, Benoît Sagot, and Antoine Bordes. 2020. Controllable sentence simplification. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4689–4698, Marseille, France. European Language Resources Association.

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.

Shiping Yang, Renliang Sun, and Xiaojun Wan. 2023. A new dataset and empirical study for sentence simplification in Chinese. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8306–8321, Toronto, Canada. Association for Computational Linguistics.

Xingxing Zhang and Mirella Lapata. 2017. Sentence simplification with deep reinforcement learning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 584–594, Copenhagen, Denmark. Association for Computational Linguistics.

Zhemin Zhu, Delphine Bernhard, and Iryna Gurevych. 2010. A monolingual tree-based translation model for sentence simplification. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 1353–1361.

A. Appendix: Corpus analysis

Figure 1 displays the density histograms of the features of CSSWiki except *Word deletion only*. CSSWiki exhibits similar properties as CSS in Lexical complexity score ratio, Proportion of words added, Proportion of words deleted, Proportion of words reordered, Replace-only Levenshtein Distance, and Dependency tree depth ratio.

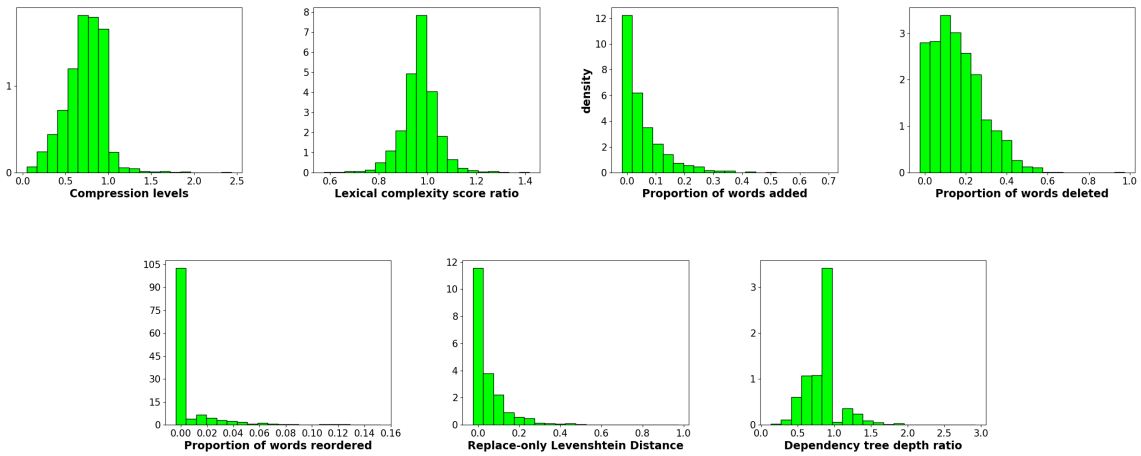

Figure 1: Density of text features in simplifications in CSSWiki

B. Appendix: Experimental Details

Table 6 presents the prompts and the few-shot templates used in our experiments. The subtask prompts replace the underlined characters in the Full SS prompt.

Task	Template
Full SS	*Zero-shot* 请在保留原意的基础上简化 以下句子: 原句：Original Sentence 简化句：Outputs
	Few-shot 请在保留原意的基础上简化 句子, 以下是五个句子简化的示例: 原句：Original Sentence 简化句：Simplified Sentence ..... 原句：Original Sentence 简化句：Outputs
Substitution only	请在保留原意的基础上通过使用更简单的词汇简化
Linguistic operations only	请在保留愿意的基础上通过修改语言简化
Elaboration only	请在保留原意的基础上通过添加注释简化
Condensation only	请在保留原意的基础上通过压缩内容简化

Table 6: Prompt templates for zero-/few-shot in the various tasks