CILS at TSAR-2022 Shared Task: Investigating the Applicability of Lexical Substitution Methods for Lexical Simplification

Sandaru Seneviratne[1], **Elena Daskalaki**[1], **Hanna Suominen**[1,2]
[1]The Australian National University (ANU) / Canberra, ACT, Australia
[2]University of Turku / Turku, Finland
{sandaru.seneviratne, eleni.daskalaki, hanna.suominen}@anu.edu.au

Abstract

Lexical simplification — which aims to simplify complex text through the replacement of difficult words using simpler alternatives while maintaining the meaning of the given text — is popular as a way of improving text accessibility for both people and computers. First, lexical simplification through substitution can improve the understandability of complex text for, for example, non-native speakers, second language learners, and people with low literacy. Second, its usefulness has been demonstrated in many natural language processing problems like data augmentation, paraphrase generation, or word sense induction. In this paper, we investigated the applicability of existing unsupervised lexical substitution methods based on pre-trained contextual embedding models and WordNet, which incorporate *C*ontext *I*nformation, for *L*exical *S*implification (CILS). Although the performance of this CILS approach has been outstanding in lexical substitution tasks, its usefulness was limited at the TSAR-2022 shared task on lexical simplification. Consequently, a minimally supervised approach with careful tuning to a given simplification task may work better than unsupervised methods. Our investigation also encouraged further work on evaluating the simplicity of potential candidates and incorporating them into the lexical simplification methods.

1 Introduction

Lexical simplification — which aims to simplify complex words and phrases in text while maintaining the meaning of the original text — is an important natural language processing (NLP) problem to improve the understandability of text for, for example, non-native speakers, second language learners, and people with low literacy skills (Gooding and Kochmar, 2019). Due to its importance in achieving complete text simplification with simpler and easier-to-read content, lexical simplification has received rising interest over the years.

Shardlow (2014) has introduced a lexical simplification pipeline, which consists of several sub-problems, including, for instance, complex word identification, simpler substitution generation, substitution selection, and substitution ranking. Out of these four sub-problems, the latter three entirely focus on the generation of relevant and simpler substitutes for better understandability.

Underpinned by Shardlow (2014) among others, over the years researchers have introduced a wide range of methods for simpler substitution generation for the complex words identified in text. Earlier approaches to substitution generation have relied on rule-based methods and lexical resources like WordNet (Miller, 1995) or paraphrase databases (Pavlick and Callison-Burch, 2016). Lexical substitution research has advanced to the use of word embedding models (e.g., word2vec (Mikolov et al., 2013), Glove (Pennington et al., 2014), Embeddings from Language Models (ELMo) (Peters et al., 2018)) to remove the requirement of lexical resources and obtain potential candidates through the cosine similarity of word embeddings.

The introduction of Transformers (Vaswani et al., 2017) has resulted in advanced contextual word and sentence embedding models like Bidirectional Encoder Representations from Transformers (BERT) (Devlin et al., 2019), robustly optimised BERT (RoBERTa) (Liu et al., 2019), and XLNet (Yang et al., 2019) which have been extensively used for NLP, including, but not limited to, lexical simplification and lexical substitution. These models have been useful in generating potential candidates for simplification given a target word and the context, taking the meaning preservation aspect into account; researchers have introduced methods and frameworks for lexical simplification, some of which rely entirely on contextual embeddings (Qiang et al., 2021) whereas some others incorporate lexical resources alongside contextual embedding models (Gooding and Kochmar, 2019).

Proceedings of the Workshop on Text Simplification, Accessibility, and Readability (TSAR-2022), pages 207 - 212

Lexical substitution can be identified as a broader problem, which aims to generate alternative substitutes for a target word (McCarthy and Navigli, 2007) whereas lexical simplification specifically focuses on generating simpler substitutes (Shardlow, 2014). Although not identical, the two problems are coupled; both aim to generate substitutes for an identified target word. Hence, also the proposed, studied, and adopted solution techniques have similarities.

In our work, we investigated the applicability of lexical substitution methods for simpler substitution generation in lexical simplification. We applied the *CILex* solution proposed in our previous work (Seneviratne et al., 2022) on *C*ontext *I*nformation for *Lex*ical substitution to lexical simplification[1]. The objective of the research was to evaluate the usefulness of existing substitution methods in a given text simplification task and to identify how these methods can be improved.

2 Related Work

Researchers have used different techniques to achieve lexical simplification; this problem aims to simplify complex content in text while maintaining the meaning, for better understandability.

Earliest lexical simplification approaches relied on rule-based methods and lexical resources (Devlin, 1998) where a set of rules was defined to extract simpler substitutes from lexical resources like WordNet (Miller, 1995) and rank them based on a simplicity metric. Extending beyond these linguistic databases, researchers also used parallel corpora that consisted of complex and simpler sentences to identify simpler substitutes for a target word (Biran et al., 2011; Yatskar et al., 2010). However, given both these approaches were dependent on linguistic databases and parallel corpora, they had limitations with respect to the availability and coverage of simpler alternatives.

To address the limitations of lexical resources, researchers adopted word embedding models for lexical simplification (Glavaš and Štajner, 2015). Further improving on the word embedding models, researchers introduced context-aware lexical simplification methods (Paetzold and Specia, 2016; Gooding and Kochmar, 2019) which also incorporated linguistic features and information from lexical resources. The introduction of Transformer-based language models like BERT, RoBERTa, and XLNet resulted in widely adopting them for downstream NLP problems. To illustrate, Qiang et al. (2020) introduced a recursive simplification method called LSBert based on BERT.

Similar techniques have been used for lexical substitution, which is a broader problem aiming to generate alternative words for a given target word. Early methods, which relied on rule-based systems and lexical resources, have evolved to the methods that use word embeddings (Melamud et al., 2016), contextual embeddings (Zhou et al., 2019; Arefyev et al., 2020) and methods that incorporate additional information from lexical resources (Michalopoulos et al., 2022).

Given the similarities in lexical substitution and simplification problems, we investigated the applicability of the CILex lexical substitution solution also for lexical simplification. This investigation was part of the Text Simplification, Accessibility, and Readability (TSAR) shared task on lexical simplification in 2022 (Saggion et al., 2022).

3 Experiments

3.1 Method

We used the CILex solution proposed in our previous work (Seneviratne et al., 2022) for our experiments, which focused on lexical substitution methods. We based our experiments on pre-trained contextual word embedding models, contextual sentence embedding models, and WordNet. We then defined several metrics to obtain the final set of relevant substitutes and rank them to filter out the most suitable substitutes.

The initial set of substitutes was obtained using the combination of i) a model prediction score $P(w|c)$ computed using the XLNet model given the context c and target word x with any word w in the vocabulary of XLNet and ii) an embedding similarity score $P(w|x)$ by computing the inner product of the embedding of the target word and the embedding of the respective word $(embedding_x \cdot embedding_w^\top)$. This followed the approach by Arefyev et al. (2020).

For each word in the XLNet vocabulary, these scores were combined to obtain $S_{\texttt{XLNet}}$ score with α and β being parameters that can be fine-tuned:

$$S_{\texttt{XLNet}} = \alpha P(w|c) + \beta P(w|x). \tag{1}$$

The scoring was then used to rank all the words to filter out the top 20 words.

[1]The implementation is available at `https://github.com/sandaruSen/CILex` under the MIT license.

For the filtered-out set of potential candidates, we computed a sentence similarity score

$$S_{\texttt{sent}} = \cos(s,\, s') \tag{2}$$

using the cosine similarity between the original and updated sentences $(s,\, s')$ obtained by replacing the target word using each potential candidate (Michalopoulos et al., 2022).

We also used the gloss sentence similarity score $S_{\texttt{gloss}}$ proposed by Michalopoulos et al. (2022) which integrated additional context information from WordNet and BERT (*bert-large-uncased*). We computed the score as follows: first, we obtained lists of potential definitions for target words and possible substitutes from WordNet. Second, for each target word and substitute, we formulated the most suitable definition by computing the cosine similarity between the given sentence and the definition. Third, for each substitute, we calculated the gloss sentence similarity score

$$S_{\texttt{gloss}} = \cos(d_t,\, d_w) \tag{3}$$

using the cosine similarity between the most suitable definition embedding of the target word d_t and the most suitable definition embedding of the substitute d_w.

Similarly to $S_{\texttt{gloss}}$, we computed

$$S_{\texttt{wordnet}} = \cos(d_t,\, s') \tag{4}$$

where lists of potential definitions were obtained only for the possible candidates and the cosine similarity was computed using the updated sentence and the most suitable definition of each substitute.

Additionally, we computed the validation score $S_{\texttt{val}}$ (Zhou et al., 2019) using the cosine similarities of the BERT-based contextual embeddings (*bert-large-uncased*) of the top four layers of every token in the original sentence and the modified sentence was used.

Using these scores, we defined three CILex solutions as

$$\textit{CILex\_1} = \gamma S_{\texttt{XLNet}} + \delta S_{\texttt{sent}}, \tag{5}$$

$$\begin{aligned}\textit{CILex\_2} = \gamma S_{\texttt{XLNet}} &+ \delta S_{\texttt{sent}} \\ &+ \theta S_{\texttt{wordnet}} + \omega S_{\texttt{val}}, \text{ and}\end{aligned} \tag{6}$$

$$\begin{aligned}\textit{CILex\_3} = \gamma S_{\texttt{XLNet}} &+ \delta S_{\texttt{sent}} \\ &+ \theta S_{\texttt{gloss}} + \omega S_{\texttt{val}}\end{aligned} \tag{7}$$

by interpolating them together using γ and δ as the weights for $S_{\texttt{XLNet}}$ and $S_{\texttt{sent}}$ scores, respectively, for all three CILex solutions. For *CILex_2* and *CILex_3*, ω was used as the weight for $S_{\texttt{val}}$ while θ was used as the weight for $S_{\texttt{wordnet}}$ and $S_{\texttt{gloss}}$ in *CILex_2* and *CILex_3*, respectively. The CILex solutions were specifically proposed for lexical substitution which is a broader problem compared to lexical simplification.

3.2 Datasets

We tested the *CILex* solution on the trial and testing datasets of the English dataset provided at the TSAR-2022 shared task (Štajner et al., 2022). The English dataset was created by manually selecting 400 instances from the 2018 Complex Word Identification Shared task dataset. This set of instances was further filtered based on the quality of the annotations provided by the annotators to obtain the final set of 386 instances with their average number of unique simpler substitutes per instance provided by the annotators being 10.55. The dataset consisted of 10 trial instances and 373 instances in the testing dataset.

3.3 Evaluation metrics

We based our evaluation on the metrics used in the TSAR-2022 shared task (Saggion et al., 2022). $MeanAveragePrecision@K$ (*MAP@K*) score with $K \in \{1,\, 3,\, 5,\, 10\}$ evaluated if the predicted substitutes by the system were relevant and if they were ranked in the top positions. $Potential@K$ and $Accuracy@K$ metrics evaluated the percentage of instances for which at least one of the substitutions predicted was present in the set of gold annotations and the ratio of instances where at least one of the K top predicted candidates matched the most frequently suggested synonym(s) in the gold list of annotated candidates, respectively.

3.4 Experimental Setup

Following Arefyev et al. (2020), we used the XLNet model (Yang et al., 2019) to obtain the initial set of substitutes, RoBERTa (*stsb-roberta-large*) model (Reimers et al., 2019) to obtain the sentence similarity score, and BERT model (*bert-large-uncased*) to obtain the WordNet similarity, gloss sentence similarity, and validation scores. We used the same hyper-parameters introduced in our previous work (Seneviratne et al., 2022) for our experiments without further tuning to the TSAR-2022 shared task. We conducted our experiments

Method	ACC@1	ACC@1@Top1	ACC@3@Top1	MAP@3	MAP@10	Potential@3	Potential@10
LSBert	0.5978	0.3029	0.5308	0.4079	0.1755	0.823	0.9463
CILex_3	0.386	0.1957	0.3083	0.2603	0.1267	0.5656	0.638
CILex_2	0.3806	0.1903	0.3083	0.2597	0.1262	0.563	0.6434
CILex_1	0.3753	0.201	0.3109	0.2555	0.1235	0.5361	0.63
TUNER	0.3404	0.142	0.1823	0.1706	0.0546	0.4343	0.445

Table 1: Results of our proposed three CILex solutions and the LSBert and TUNER baselines for the test subset of the English dataset provided at the TSAR-2022 shared task.

on a RTX 3090 graphics card with 24 GB memory and CUDA 11.4.

3.5 Results

The proposed three solutions outperformed the TUNER-baseline (Table 1). However, they did not perform as well as the LSBert baseline.

Although the performance of our CILex approach has been outstanding in lexical substitution tasks, its usefulness was limited at the TSAR-2022 shared task on lexical simplification. Remembering that lexical substitution and simplification problems are not identical and that also text datasets and their respective annotations have their unique characteristics, a minimally supervised approach with some careful tuning to this specific simplification task could have worked better at TSAR-2022 than our unsupervised lexical substitution methods with pre-trained models.

4 Discussion

In this paper, we have adopted our previous work on lexical substitution for the TSAR-2022 shared task on lexical simplification and experimented with the use of different methods that provide context information. The three methods used for our experiments have not performed as well as the LSBert baseline. However, they have outperformed TUNER-baseline at TSAR-2022 (Štajner et al., 2022) and our approach has excelled in lexical substitution as evidenced by its evaluation using two annotated lexical substitution datasets that are widely used for the broader problem (Seneviratne et al., 2022).

The observed performance difference for lexical simplification and lexical substitution can be explained by the problem differences. The methods used in our experiments were developed to target lexical substitution, which can be identified as a substantially broader problem than lexical simplification. Lexical substitution generally focuses on generating similar rather than simpler substitutes — the narrower focus of lexical simplification. In order to tackle this issue of our lexical substitution approach for lexical simplification, identifying metrics, which can evaluate the simplicity of the potential candidates and using them to rank the potential candidates can be done.

5 Conclusion

We have applied our previous work on lexical substitution for lexical simplification, focusing on the added value of context information for the lexical simplification problem. The results from our methods indicate, that even though the proposed approach has performed well in lexical substitution more broadly, their usefulness in the narrower lexical simplification problem at the TSAR-2022 shared task was limited; a minimally supervised approach with some careful tuning to a given simplification task may have worked better at TSAR-2022 than unsupervised lexical substitution methods with pre-trained models.

Our investigation encourages further work on evaluating the simplicity of potential substitution candidates and incorporating them into lexical substitution methods. This approach should extend these broader methods to lexical simplification by targeting the more specific constraints for substitutes in the narrower text simplification problem.

Acknowledgement

This research was funded by and has been delivered in partnership with Our Health in Our Hands (OHIOH), a strategic initiative of the ANU, which aims to transform health care by developing new personalized health technologies and solutions in collaboration with patients, clinicians and healthcare providers. We gratefully acknowledge the funding from the ANU School of Computing for the first author's PhD studies.

References

Nikolay Arefyev, Boris Sheludko, Alexander Podolskiy, and Alexander Panchenko. 2020. Always keep your target in mind: Studying semantics and improving performance of neural lexical substitution. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1242–1255, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Or Biran, Samuel Brody, and Noémie Elhadad. 2011. Putting it simply: a context-aware approach to lexical simplification. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 496–501, Portland, Oregon, USA. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Siobhan Devlin. 1998. The use of a psycholinguistic database in the simplification of text for aphasic readers. *Linguistic Databases*.

Goran Glavaš and Sanja Štajner. 2015. Simplifying lexical simplification: Do we need simplified corpora? In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 63–68, Beijing, China. Association for Computational Linguistics.

Sian Gooding and Ekaterina Kochmar. 2019. Recursive context-aware lexical simplification. 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)*, pages 4853–4863, Hong Kong, China. Association for Computational Linguistics.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized BERT pretraining approach. *arXiv preprint arXiv:1907.11692*.

Diana McCarthy and Roberto Navigli. 2007. SemEval-2007 task 10: English lexical substitution task. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 48–53. Association for Computational Linguistics.

Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning generic context embedding with bidirectional LSTM. In *Proceedings of The 20th SIGNLL Conference on Computational Natural Language Learning*, pages 51–61, Berlin, Germany. Association for Computational Linguistics.

George Michalopoulos, Ian McKillop, Alexander Wong, and Helen Chen. 2022. LexSubCon: Integrating knowledge from lexical resources into contextual embeddings for lexical substitution. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1226–1236, Dublin, Ireland. Association for Computational Linguistics.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013*.

George A Miller. 1995. WordNet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.

Gustavo Paetzold and Lucia Specia. 2016. Benchmarking lexical simplification systems. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3074–3080, Portorož, Slovenia. European Language Resources Association (ELRA).

Ellie Pavlick and Chris Callison-Burch. 2016. Simple PPDB: A paraphrase database for simplification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 143–148, Berlin, Germany. Association for Computational Linguistics.

Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543, Doha, Qatar. Association for Computational Linguistics.

Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. 2018. Deep contextualized word representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana. Association for Computational Linguistics.

Jipeng Qiang, Yun Li, Yi Zhu, Yunhao Yuan, Yang Shi, and Xindong Wu. 2021. LSBert: Lexical simplification based on BERT. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:3064–3076.

Jipeng Qiang, Yun Li, Yi Zhu, Yunhao Yuan, and Xindong Wu. 2020. Lexical simplification with pretrained encoders. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 8649–8656.

Nils Reimers, Iryna Gurevych, Nils Reimers, Iryna Gurevych, Nandan Thakur, Nils Reimers, Johannes Daxenberger, Iryna Gurevych, Nils Reimers, Iryna Gurevych, et al. 2019. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*, pages 671–688. Association for Computational Linguistics.

Horacio Saggion, Sanja Štajner, Daniel Ferrés, Kim Cheng Sheang, Matthew Shardlow, Kai North, and Marcos Zampieri. 2022. Findings of the TSAR-2022 shared task on multilingual lexical simplification. In *Proceedings of TSAR workshop held in conjunction with EMNLP 2022*. Association for Computational Linguistics.

Sandaru Seneviratne, Elena Daskalaki, Artem Lenskiy, and Hanna Suominen. 2022. CILex: An investigation of context information for lexical substitution methods. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 4124–4135, Gyeongju, Republic of Korea. International Committee on Computational Linguistics.

Matthew Shardlow. 2014. A survey of automated text simplification. *International Journal of Advanced Computer Science and Applications*, 4(1):58–70.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30:5998–6008.

Sanja Štajner, Daniel Ferrés, Matthew Shardlow, Kai North, Marcos Zampieri, and Horacio Saggion. 2022. Lexical simplification benchmarks for English, Portuguese, and Spanish. *Frontiers in Artificial Intelligence*, 5.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. XLNet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*.

Mark Yatskar, Bo Pang, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2010. For the sake of simplicity: Unsupervised extraction of lexical simplifications from Wikipedia. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 365–368, Los Angeles, California. Association for Computational Linguistics.

Wangchunshu Zhou, Tao Ge, Ke Xu, Furu Wei, and Ming Zhou. 2019. BERT-based lexical substitution. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3368–3373, Florence, Italy. Association for Computational Linguistics.