On the Ethical Considerations of Text Simplification

Sian Gooding
Department of Computer Science and Technology
University of Cambridge
shg36@cam.ac.uk

Abstract

This paper outlines the ethical implications of text simplification within the framework of assistive systems. We argue that a distinction should be made between the technologies that perform text simplification and the realisation of these in assistive technologies. When using the latter as a motivation for research, it is important that the subsequent ethical implications be carefully considered. We provide guidelines for the framing of text simplification independently of assistive systems, as well as suggesting directions for future research and discussion based on the concerns raised.

1 Introduction

Assistive technology refers to the devices used to support or aid those living with disabilities (Preston, 2003). The intent behind such technologies is to increase independence and maximise societal participation for individuals (Borg et al., 2011).

There are many examples of assistive technology that rely on speech and natural language processing. For instance, sign language translation (Camgoz et al., 2018), pronunciation adaptation for disordered speech (Sriranjani et al., 2015) and synthesised voices for individuals with vocal disabilities (Veaux et al., 2013). *Text simplification* is an area of natural language processing concerned with the simplification of textual information and is often recognised as having assistive applications. Prior research in text simplification posits that such technology may be beneficial for audiences with reading difficulties or a range of disabilities such as dyslexia, aphasia or deafness.

However, currently the algorithms designed for text simplification are considered in isolation from their assistive applications, and there is subsequently little discussion on the ethical implications for the intended users. Text simplification research is often motivated by highlighting the audiences

Table 1: Example of manually simplified sentence from the Newsela Dataset (Xu et al., 2015)

that could benefit from such tools, thereby coupling the technology with the assistive applications. Framing text simplification via the implications for assistive technology means that the ethical considerations cannot be easily separated from the technology used to generate the result. An issue which is commonly acknowledged in the assistive technology literature (Niemeijer et al., 2010).

There are many potential benefits of text simplification embedded in assistive technology, and both for service providers and service users, there are also a number of ethical issues that must be considered. In this paper, we will discuss the ethical considerations that arise from the embedding of text simplification within assistive technologies. Our aim is to encourage the discussion and consideration of these issues, as well as inform the design decisions of future assistive technologies that incorporate text simplification.

2 Background

Complete textual simplification requires many types of transformations which can be grouped into three categories: syntactic, lexical and conceptual (Siddharthan, 2014). Table 1 illustrates a range of simplification operations from these different

50

categories, a description of these is as follows:

Lexical simplification is concerned with reducing the complexity of words within a text (Paetzold and Specia, 2017; Gooding and Kochmar, 2019). In lexical simplification, complex words are identified and replaced with simpler alternatives. We observe an example of lexical simplification with the case of *celebrities* being simplified to *stars*.

Syntactic simplification aims to reduce the grammatical complexity of text by simplifying the syntactical structures. Examples of such transformations include the conversion of text from passive to active voice and dis-embedding relative clauses (Siddharthan, 2006a). In our example, multiple syntactic simplifications have taken place. One such simplification occurs where the subordinated clause '...*undermining a long tradition*' has been split into a separate sentence. Syntactic simplification often requires discourse preserving edits to maintain the coherence and cohesion of simplified text. For instance, the addition of '*They say...*' is necessary to convert the original relative clause into a grammatically correct and coherent sentence.

Finally, conceptual simplification focuses on the simplification of ideas or concepts within text. The example shows how the concept of *philatelist* has been simplified by providing an explanation of the term. This simplification technique is commonly referred to as elaboration, as the meaning of the concept has been elaborated on (Siddharthan, 2006a). Often, this strategy is used in cases where no alternative synonym would suffice, for instance with named entities.

Both syntactic and conceptual simplification contain parallels with the research area of *text summarization* as omitting peripheral or inappropriate information, as well as distilling complex concepts, is relevant for both. However, in simplification these processes can increase the length of the original text, whereas in summarization the goal is to constrain the length of the resulting summary.

In automatic text simplification, the aim is to transform text using the aforementioned operations, to allow individuals with differing comprehension levels access. This requires a fundamental understanding of what factors contribute to text complexity for differing audiences (Gooding et al., 2021b).

Early approaches to automated simplification were largely rule-based systems (Canning et al., 2000; Carroll et al., 1998a; Siddharthan, 2006b), with many prioritising syntactic operations, such as

sentence splitting, deletion or reordering. However, some work combined lexical simplification with syntactic operations (Coster and Kauchak, 2011; Kauchak, 2013; Zhu et al., 2010). In recent years simplification has been viewed as a monolingual translation task (Kauchak, 2013; Zhang and Lapata, 2017; Zhu et al., 2010). These systems perform a number of simplification operations at once by aiming to translate *complex English* to *simple English*. Initial approaches attempt this with phrase-based machine translation (Coster and Kauchak, 2011; Wubben et al., 2012) while subsequent work has focused on neural machine translation techniques (Nisioi et al., 2017; Zhang and Lapata, 2017; Shardlow and Nawaz, 2019; Dong et al., 2019).

3 Risks and Harms

In this section we outline and discuss the potential risks and harms that arise from the integration of text simplification within assistive technology.

3.1 Intended Audience

As with many areas of research, the field of text simplification has converged on a partially boilerplate preamble outlining a set of motivations. Table 2 features extracts taken from a sample of recent text simplification papers. These papers were sampled by searching the ACL anthology for the term *text simplification* and ordering by most recent. We look specifically at sections outlining the audiences said to benefit from text simplification as a whole. Below, we consider the ethical implications of citing such audiences as a motivation for text simplification.

3.1.1 The Homogeneity Effect

As shown in Table 2, the audiences stated to benefit from text simplification are often listed together, namely non-native speakers, children, people with low literacy skills, people with reading disabilities or disabilities generally. Based on this, a reader may be given the impression that general purpose text simplification works adequately for all of the stated groups. Whereas in fact, there is evidence to show that text simplification may not be effective for second language learners (Young, 1999), that alternative strategies to simplification can be most effective for dyslexia (Rello et al., 2013a) and that automated text simplification cannot simplify content to a low enough level for children (De Belder and Moens, 2010).

Audience outline	Datasets	Evaluation	Venue
(1) ...such as children, people with low education, people who have reading disorders or dyslexia, and non-native speakers of the language.	NEWSELA WIKILARGE BIENDATA	Automatic	*ACL 2021*
(2) It provides reading assistance to children (Kajiwara et al., 2013), non-native speakers (Petersen and Ostendorf, 2007; Pellow and Eskenazi, 2014; Paetzold, 2016) and people with reading disabilities (Rello et al., 2013b)	NEWSELA	Automatic + 5 workers	*NAACL 2021*
(3) It can provide convenience for non-native speakers (Petersen and Ostendorf, 2007; Glavaš and Štajner, 2015; Paetzold and Specia, 2016c; Rello et al., 2013b), non-expert readers (Elhadad and Sutaria, 2007; Siddharthan and Katsos, 2010) and children (De Belder and Moens, 2010; Kajiwara et al., 2013)	D-WIKIPEDIA NEWSELA	Automatic + 3 workers	*EMNLP 2021*
(4) ...to children (De Belder and Moens, 2010; Kajiwara et al., 2013), people with language disabilities like aphasia (Carroll et al., 1998b, 1999b; Devlin and Unthank, 2006), dyslexia (Rello et al., 2013a,b), or autism (Evans et al., 2014); non-native (Petersen and Ostendorf, 2007; Paetzold, 2015; Paetzold and Specia, 2016b; Pellow and Eskenazi, 2014) English speakers, and people with low literacy skills or reading ages.	WIKISMALL WIKILARGE	Automatic	*BEA 2021*

Table 2: Examples of paper introductions outlining audiences benefiting from text simplification, alongside the evaluation techniques and venue, specific paper references are included in Appendix A.

Framing the benefit of text simplification as net positive for all groups can have consequences for the development of assistive technology, as merging the audiences serves to diminish the sensitive differences in needs for these groups. Even for specific audiences, such as children, there is a consensus that the homogeneous grouping of reading ability can have detrimental outcomes for learning (Schumm et al., 2000).

A further complication, is that the references commonly used in support of text simplification (for specific audiences) are often more nuanced than stated. For instance, the work of Rello et al. (2013a) is commonly cited as showing the benefit of text simplification for dyslexic readers. However, this paper demonstrates that the most effective strategy to help dyslexic readers with difficult words, is to provide a range of synonyms for the word, and not to simplify the original. Furthermore, the work of Carroll et al. (1998a) and Carroll et al. (1999a) is put forward as evidence for the utility of simplification for individuals with aphasia. However, both of these works outline the proposal for a simplification system targeted for aphasia and propose to evaluate the effectiveness of such a system in future

work. A final example, used in support of text simplification for children is a paper by De Belder and Moens (2010). However, the paper finds that even using lexical and syntatic simplification, it was not possible to reduce the reading difficulty enough for children.

3.1.2 Datasets and Evaluation

Text simplification has many subtleties, as what would be a valid simplification for one reader may not be appropriate for another (Xu et al., 2015). For instance, it has been shown that the factors contributing to word complexity vary depending on the first language and proficiency level of a reader (Gooding et al., 2021b). The subjective nature of text simplification means that system evaluation is difficult. Furthermore, as there is not one 'ground truth' for simplification, the efficacy of automatic evaluation measures is limited. Prior work on the development and evaluation of simplification systems has given little consideration to the target reader population (Xu et al., 2015).

As exemplified in Table 2, current work on text simplification typically relies on automatic evaluation, with the occasional use of human evaluation.

When considering the approach to human evaluation, most work does not specify what "being simpler" entails, and trusts human judges to use their own understanding of the concept (Alva-Manchego et al., 2021). It is also currently not standard practice to include the demographic information of the workers. When using human judgements as a measure of simplification quality, it is important to include relevant information on the demographic background, so that valid conclusions can be drawn about which target population may benefit from the system. Additionally, the concept of what constitutes adequate simplification needs to be precise if the system is aimed for a specialised audience.

The datasets commonly used to train text simplification systems (i.e. Newsela and Simple Wikipedia) have drawbacks such as poor alignment, lack of simplicity and not being tailored for a specific audiences (Xu et al., 2015). In text simplification, it is important to discuss the limitations of the data so that the suitability of such systems for specialised groups is clearly recognised.

In summary, when claiming the benefits of text simplification for specific audiences, it is crucial that the needs of these groups are understood. This is especially the case when emphasising the benefit of such technology for disabled groups. The development of assistive technology is downstream from research, and therefore being clear about the suitability and limitations of the technology for differing audiences helps to avoid poorly suited assistive technology solutions being developed.

3.2 Meaning Distortion

There are multiple genres of text where access is highly important, such as healthcare information or political materials. The benefits of simplifying such content have been shown, for example simplifying text in health care improves understanding of information regardless of health literacy level (Kim and Kim, 2015). Furthermore, the complexity of language matters for voters' perceptions of political parties and their positions (Bischof and Senninger, 2018).

The benefit of text simplification in such cases is apparent, as is the need to ensure the meaning of such text is preserved and that no errors are introduced. A drawback to current automated text simplification systems is that the subtleties of meaning intended by the author may be diluted, if not lost altogether (Chandrasekar et al., 1996). For exam-

ple, Shardlow and Nawaz (2019) found that fully automated approaches omitted 30% of critical information when used to simplify clinical texts. For these types of domains, instead of fully-automated approaches, interactive text simplification tools are better suited to generate more efficient and higher quality simplifications (Kloehn et al., 2018).

The link between factual correctness and natural language generation has been considered for multiple domains such as summarization (Cao et al., 2018), data to document generation (Wiseman et al., 2017) and dialog generation (Shuster et al., 2021). However, this is a relatively underexplored area for text simplification and is currently not incorporated into the evaluation of such systems.

Encouraging further discussion on this limitation of text simplification is necessary. Especially when we consider the downstream applications of assistive technology for critical consumer information.

3.3 Paternalism

There are choices made in the design characteristics of assistive technology that can affect the degree of independence, privacy and participation that are possible (Lenker et al., 2013). These decisions have real world impact for the users and thus warrant careful consideration.

The process of text simplification involves an understanding of *what* is difficult and *how* best to simplify it. There are two approaches when deciding *what* should be simplified. The first, involves including the reader in the loop – either implicitly or explicitly. Relying on user signal to identify areas for simplification has its own set of ethical concerns which are discussed in § 3.4. The second approach, is performing general level simplification with end-to-end systems. In fact, the majority of current work in text simplification is now data-driven and performs simplification in a 'black-box' fashion (Sikka and Mago, 2020). One of the concerns for such systems, is that they learn operations based on the simplification choices made in the data they are trained with. As outlined in Section 3.1.2, most of the data used to train text simplification systems is not audience specific (Xu et al., 2015).

Integrating general purpose simplification systems into assistive technologies has a range of potential problems. For instance, it raises the issue of "paternalism" which is the interference of a state or individual in relation to another person (Martin

et al., 2007). The relationship between paternalism and assistive technology is widely acknowledged, as design decisions made on behalf of a user can be problematic if they override the autonomy of the individual (Martin et al., 2007, 2010). In the case of text simplification, not allowing the individual the choice of what they would want simplified restricts their autonomy.

Furthermore, assistive technologies should contribute to growth and independence for individuals. The goals of text simplification are to make textual information accessible to a range of different audiences. However, the question of whether such systems should support learning is rarely discussed. One concern with text simplification within assistive systems, is that it would prevent the exposure to new terms and concepts thereby encouraging learning stagnation.

In summary, the design decisions pertaining to what content is simplified have ethical implications for the user. Removing the individual from the decision process can reduce the person's autonomy, and not allowing exposure to new and unfamiliar terms limits learning opportunities, subsequently reducing the user's independence.

3.4 Privacy and Security

As outlined in Section 3.3, an effective approach for text simplification in assistive technology is to include the user in the decision of what is simplified. Prior research has shown that eye-tracking (Berzak et al., 2018) and scroll-based interactions (Gooding et al., 2021a) correlate with text understanding. As such, these implicit techniques can be used to gain an insight into what the reader is finding difficult. The reader can also be explicitly asked to select text that they would like to be simplified, for instance by selecting words that are difficult for them (Devlin and Unthank, 2006; Paetzold and Specia, 2016a).

Adaptive text simplification is advantageous and provides autonomy and learning opportunities for the user. However, the information about the areas a user finds difficult is highly sensitive, and there is a responsibility to ensure that such information is stored securely.

To protect the privacy of the user, the aim of the assistive technology and the way it is used by service providers or care organisations must be clear. Moreover, how personal data will be handled must be described explicitly in a privacy statement and communicated to the user (Martin et al., 2010).

The above is a clear example of how viewing text simplification through the paradigm of assistive technology yields more nuanced ethical considerations. We believe it would be beneficial to encourage the discourse on such aspects in the text simplification literature.

4 Going Forward

We suggest that papers focusing on general purpose text simplification should de-couple the motivations from specific audiences with disabilities. An example of a general purpose motivation by Nisioi et al. (2017) is as follows:

Automated text simplification (ATS) systems are meant to transform original texts into different (simpler) variants which would be understood by wider audiences and more successfully processed by various NLP tools.

Alternatively, if discussing the different groups of users who may benefit from text simplification, being clear about the specific strategies that work for these audiences is critical. Additionally, it is worth acknowledging that when framing a system using a target demographic, it is appropriate that the system is tested with that target group. For human evaluation generally, it is highly beneficial to report the demographic statistics, as this allows an insight into which types of audiences the system may work well for.

Finally, it is important to be forthright about the current limitations of both the data and evaluation techniques used in automatic text simplification. Whilst great improvements are being made in this area, these systems are still far from perfect and this needs to be taken into account when judging the suitability of systems for assistive technology.

5 Conclusion

Assistive technologies can dramatically affect the lives of those who rely on them, and it is important to understand the potential ethical concerns – especially as such technologies can impact vulnerable populations. In this paper, we discuss a set of potential issues that arise from the embedding of text simplification within assistive technologies.

Our aim in this work is to encourage further discussion on how the design decisions of text simplification algorithms can have the potential to impact future users of assistive technology.

References

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.

Yevgeni Berzak, Boris Katz, and Roger Levy. 2018. Assessing language proficiency from eye movements in reading. 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 1986–1996, New Orleans, Louisiana. Association for Computational Linguistics.

Daniel Bischof and Roman Senninger. 2018. Simple politics for the people? complexity in campaign messages and political knowledge. *European Journal of Political Research*, 57(2):473–495.

Johan Borg, Stig Larsson, and Per-Olof Östergren. 2011. The right to assistive technology: For whom, for what, and by whom? *Disability & Society*, 26(2):151–167.

Necati Cihan Camgoz, Simon Hadfield, Oscar Koller, Hermann Ney, and Richard Bowden. 2018. Neural sign language translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7784–7793.

Yvonne Canning, John Tait, Jackie Archibald, and Ros Crawley. 2000. Cohesive generation of syntactically simplified newspaper text. In *Proceedings of the Third International Workshop on Text, Speech and Dialogue (TDS '00)*, pages 145–150.

Ziqiang Cao, Furu Wei, Wenjie Li, and Sujian Li. 2018. Faithful to the original: Fact aware neural abstractive summarization. In *thirty-second AAAI conference on artificial intelligence*.

John Carroll, Gido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998a. Practical simplification of English newspaper text to assist aphasic readers. In *Proceedings of AAAI Workshop on Integrating AI and Assistive Technology*, pages 7–10.

John Carroll, Guido Minnen, Yvonne Canning, Siobhan Devlin, and John Tait. 1998b. 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. Citeseer.

John Carroll, Guido Minnen, Darren Pearce, Yvonne Canning, Siobhan Devlin, and John Tait. 1999a. Simplifying Text for Language-Impaired Readers. In *Proceedings of the 9th Conference of the European Chapter of the ACL (EACL'99)*, pages 269–270, Bergen, Norway.

John A Carroll, Guido Minnen, Darren Pearce, Yvonne Canning, Siobhan Devlin, and John Tait. 1999b. Simplifying text for language-impaired readers. In *Ninth Conference of the European Chapter of the Association for Computational Linguistics*, pages 269–270.

Raman Chandrasekar, Christine Doran, and Srinivas Bangalore. 1996. Motivations and methods for text simplification. In *COLING 1996 Volume 2: The 16th International Conference on Computational Linguistics*.

William Coster and David Kauchak. 2011. Learning to Simplify Sentences Using Wikipedia. In *Proceedings of the Workshop on Monolingual Text-To-Text Generation*, pages 1–9.

Jan De Belder and Marie-Francine Moens. 2010. Text simplification for children. In *Proceedings of the SIGIR workshop on accessible search systems*, pages 19–26. ACM.

Siobhan Devlin and Gary Unthank. 2006. Helping aphasic people process online information. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, pages 225–226.

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.

Noemie Elhadad and Komal Sutaria. 2007. Mining a lexicon of technical terms and lay equivalents. In *Biological, translational, and clinical language processing*, pages 49–56.

Richard Evans, Constantin Orasan, and Iustin Dornescu. 2014. An evaluation of syntactic simplification rules for people with autism. Association for Computational Linguistics.

Cristina Garbacea, Mengtian Guo, Samuel Carton, and Qiaozhu Mei. 2021. Explainable prediction of text complexity: The missing preliminaries for text simplification. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1086–1097, Online. Association for Computational Linguistics.

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)*, volume 2, pages 63–68.

Sian Gooding, Yevgeni Berzak, Tony Mak, and Matt Sharifi. 2021a. Predicting text readability from scrolling interactions. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, pages 380–390, Online. 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.

Sian Gooding, Ekaterina Kochmar, Seid Muhie Yimam, and Chris Biemann. 2021b. Word complexity is in the eye of the beholder. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4439–4449, Online. Association for Computational Linguistics.

Tomoyuki Kajiwara, Hiroshi Matsumoto, and Kazuhide Yamamoto. 2013. Selecting proper lexical paraphrase for children. In *Proceedings of the 25th Conference on Computational Linguistics and Speech Processing (ROCLING 2013)*, pages 59–73.

David Kauchak. 2013. Improving Text Simplification Language Modeling Using Unsimplified Text Data. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1537–1546.

Eun Jin Kim and Su Hyun Kim. 2015. Simplification improves understanding of informed consent information in clinical trials regardless of health literacy level. *Clinical Trials*, 12(3):232–236.

Nicholas Kloehn, Gondy Leroy, David Kauchak, Yang Gu, Sonia Colina, Nicole P Yuan, Debra Revere, et al. 2018. Improving consumer understanding of medical text: Development and validation of a new subsimplify algorithm to automatically generate term explanations in english and spanish. *Journal of medical Internet research*, 20(8):e10779.

James A Lenker, Frances Harris, Mary Taugher, and Roger O Smith. 2013. Consumer perspectives on assistive technology outcomes. *Disability and Rehabilitation: Assistive Technology*, 8(5):373–380.

Mounica Maddela, Fernando Alva-Manchego, and Wei Xu. 2021. Controllable text simplification with explicit paraphrasing. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 3536–3553, Online. Association for Computational Linguistics.

Suzanne Martin, Johan E Bengtsson, and Rose-Marie Dröes. 2010. Assistive technologies and issues relating to privacy, ethics and security. In *Supporting people with dementia using pervasive health technologies*, pages 63–76. Springer.

Suzanne Martin, Colm Cunningham, and Chris Nugent. 2007. Ethical considerations for integrating technology. *Alzheimer's Care Today*, 8(3):251–258.

Alistair R Niemeijer, Brenda JM Frederiks, Ingrid I Riphagen, Johan Legemaate, Jan A Eefsting, and Cees MPM Hertogh. 2010. Ethical and practical concerns of surveillance technologies in residential care for people with dementia or intellectual disabilities: an overview of the literature. *International Psychogeriatrics*, 22(7):1129–1142.

Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P Dinu. 2017. Exploring neural text simplification models. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 85–91.

Kostiantyn Omelianchuk, Vipul Raheja, and Oleksandr Skurzhanskyi. 2021. Text Simplification by Tagging. In *Proceedings of the 16th Workshop on Innovative Use of NLP for Building Educational Applications*, pages 11–25, Online. Association for Computational Linguistics.

Gustavo Paetzold. 2015. Reliable lexical simplification for non-native speakers. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Student Research Workshop*, pages 9–16.

Gustavo Paetzold and Lucia Specia. 2016a. Anita: An intelligent text adaptation tool. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 79–83.

Gustavo Paetzold and Lucia Specia. 2016b. 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.

Gustavo Paetzold and Lucia Specia. 2016c. Unsupervised lexical simplification for non-native speakers. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Gustavo H Paetzold and Lucia Specia. 2017. A survey on lexical simplification. *Journal of Artificial Intelligence Research*, 60:549–593.

Gustavo Henrique Paetzold. 2016. *Lexical Simplification for Non-Native English Speakers*. Ph.D. thesis, University of Sheffield.

David Pellow and Maxine Eskenazi. 2014. An open corpus of everyday documents for simplification tasks. In *Proceedings of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations (PITR)*, pages 84–93.

Sarah E Petersen and Mari Ostendorf. 2007. Text simplification for language learners: a corpus analysis. In *Workshop on speech and language technology in education*. Citeseer.

Karen Preston. 2003. Assistive technologies: Principles and practice. *Rehabilitation Nursing*, 28(2):64.

Luz Rello, Ricardo Baeza-Yates, Stefan Bott, and Horacio Saggion. 2013a. 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.

Luz Rello, Ricardo Baeza-Yates, and Horacio Saggion. 2013b. The impact of lexical simplification by verbal paraphrases for people with and without dyslexia. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 501–512. Springer.

Jeanne Shay Schumm, Sally W Moody, and Sharon Vaughn. 2000. Grouping for reading instruction: Does one size fit all? *Journal of learning disabilities*, 33(5):477–488.

Matthew Shardlow and Raheel Nawaz. 2019. Neural text simplification of clinical letters with a domain specific phrase table.

Kurt Shuster, Spencer Poff, Moya Chen, Douwe Kiela, and Jason Weston. 2021. Retrieval augmentation reduces hallucination in conversation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3784–3803, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Advaith Siddharthan. 2006a. Syntactic simplification and text cohesion. *Research on Language and Computation*, 4(1):77–109.

Advaith Siddharthan. 2006b. Syntactic Simplification and Text Cohesion. *Research on Language and Computation*, 4:77–109.

Advaith Siddharthan. 2014. A survey of research on text simplification. *ITL-International Journal of Applied Linguistics*, 165(2):259–298.

Advaith Siddharthan and Napoleon Katsos. 2010. Reformulating discourse connectives for non-expert readers. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 1002–1010.

Punardeep Sikka and Vijay Mago. 2020. A survey on text simplification. *arXiv preprint arXiv:2008.08612*.

R Sriranjani, Srinivasan Umesh, and M Ramasubba Reddy. 2015. Pronunciation adaptation for disordered speech recognition using state-specific vectors of phone-cluster adaptive training. In *Proceedings of SLPAT 2015: 6th Workshop on Speech and Language Processing for Assistive Technologies*, pages 72–78.

Renliang Sun, Hanqi Jin, and Xiaojun Wan. 2021. Document-level text simplification: Dataset, criteria and baseline. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 7997–8013, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Christophe Veaux, Junichi Yamagishi, and Simon King. 2013. Towards personalised synthesised voices for individuals with vocal disabilities: Voice banking and reconstruction. In *Proceedings of the Fourth Workshop on Speech and Language Processing for Assistive Technologies*, pages 107–111.

Sam Wiseman, Stuart Shieber, and Alexander Rush. 2017. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2253–2263, Copenhagen, Denmark. Association for Computational Linguistics.

Sander Wubben, Antal Van Den Bosch, and Emiel Krahmer. 2012. Sentence simplification by monolingual machine translation. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 1015–1024. Association for Computational Linguistics.

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 (TACL)*, 3:283–297.

Dolly N Young. 1999. Linguistic simplification of sl reading material: Effective instructional practice? *The Modern Language Journal*, 83(3):350–366.

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.

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

Paper references from Table 2:

1. Explainable Prediction of Text Complexity: The Missing Preliminaries for Text Simplification by Garbacea et al. (2021)

2. by Controllable Text Simplification with Explicit Paraphrasing by Maddela et al. (2021)

3. Document-Level Text Simplification: Dataset, Criteria and Baseline by Sun et al. (2021)

4. Text Simplification by Tagging by Omelianchuk et al. (2021)