The Impact of Collaborative Interfaces on Text Simplification

Connor Ford
cbfa2016@pomona.edu
Pomona College
Claremont, CA

David Kauchak
david.kauchak@pomona.edu
Pomona College
Claremont, CA

ABSTRACT

In this paper we examine the benefits of collaborative text simplification relative to traditional individual simplification. We developed a collaborative web application to gather medical-related sentence simplifications from pairs of crowdsourced workers in four different collaborative environments. We find that chatbox collaboration can improve the quality of the simplification, particularly for fluency. In addition, not only do the number of collaborative actions in the simplification environment positively correlate with the simplicity rating, but also simplification quality increases as users complete more tasks indicating that improved collaboration dynamics benefit simplification.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in collaborative and social computing**; *Synchronous editors.*

KEYWORDS

Collaboration, Text simplification, Health literacy

INTRODUCTION

The goal of text simplification is to transform text into a variant that is easier to understand and more broadly accessible while maintaining the original content. An important application of text simplification is health literacy, where it is critical that patient education materials are readable and accessible [2]. In 2004, 47% of US adults had literacy levels approximately equal to or worse than a high-school level [4]. Further, because medical texts are often much more complex than a high-school literacy level, nearly 50% of American adults may have a difficult time understanding these texts. Not

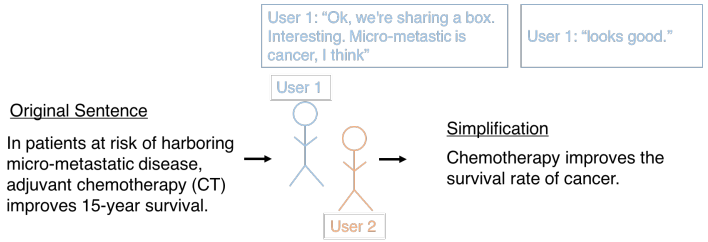

Figure 1: Example collaborative interaction for the Chatbox environment. In this case, User 1 identifies and defines a complex phrase ('micro-metastic') and User 2 constructs a simplification using this knowledge.

only do these medical texts pose an accessibility issue, but they also raise financial concerns. Vernon et al. estimate that low health literacy rates cost the U.S. economy $100 billion annually [11].

Simplification in fields such as health and medicine are traditionally done by a single person, often a medical professional [8]. In this paper we explore a new approach to simplification: simultaneous digital collaborative text simplification. This approach utilizes two or more people simultaneously working together to generate a simplification. Collaborative efforts are shown to have a number of advantages over individual approaches, i.e., in cases where the synergistic benefits of collaboration outweigh the sum of the parts [5, 7, 12]. Whereas previous research has successfully utilized a *sequential* collaborative model [1], in this paper we propose several collaborative simplification approaches for medical-related sentences that utilize real-time, synchronous work.

We study three collaborative environments and compare the quality of these collaborative simplifications with baseline individual human simplifications. These structured environments provide simple collaborative tooling such as a chatbox or text highlighting. We find that collaboration using a chatbox significantly improves the fluency metric. In addition, the number of *collaborative actions* (a numeric proxy for the degree of collaboration) positively correlates with increases in the simplicity metric. We discover that simplification quality improves as pairs of workers simplify more sentences; however, we do not see a similar improvement with the individual cases. This suggests that improvements in real-time collaboration dynamics enhance the quality of the simplification.

EXPERIMENTAL DESIGN

We developed a web application to handle the enqueuing of participants, match-making, simplification processes, and data collection. We crowdsourced participants from Amazon Mechanical Turk[1] which has been shown to be as reliable as in-person evaluations [9]. We assigned workers to one of four

[1]https://www.mturk.com/

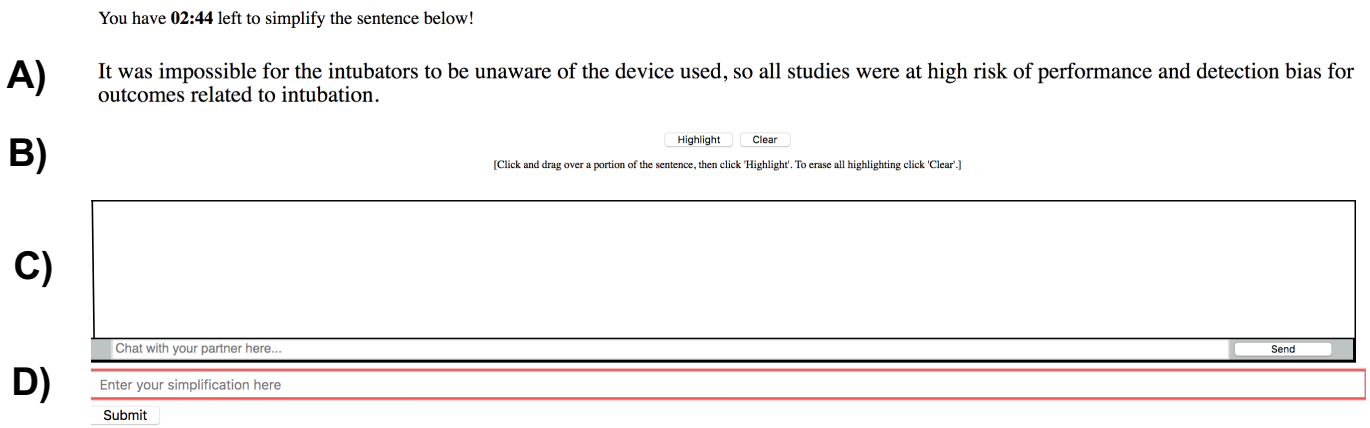

Figure 2: All-access collaboration interface for simplifications. Partners see identical interfaces. We highlight four features: (A) the original sentence, (B) highlighting tooling, (C) chatbox, and (D) submission field.

simplification environments where the functionality of each environment revolves around the use of two features, highlighting and a chatbox. (1) *All-access*: Both users have access to chatbox and highlighting tools. Figure 2 shows the simplification environment for the All-access environment. (2) *Highlighting*: One user is instructed to highlight complex portions of the sentence. The other is instructed to use these highlightings to help construct a simplification (3) *Chatbox*: One user is instructed to provide suggestions and simplifications in the chatbox while the other constructs the simplification. Figure 1 shows an example Chatbox interaction. (4) *Individual*: One user is responsible for the simplification without any collaboration (baseline). For each environment we also experimented with a timing component. Half the cases had no minimum time restriction and the other half only allowed submission after 30 seconds. The functionality of these environments is intentionally simple. We want the participants to be able to focus on the simplification task and collaboration rather than the intricacies of the web tool.

In every environment, the pair (or individual) simplifies a set of five medical-related sentences. These sentences come from a selection of Wikipedia and Cochrane medical articles[2]. For each simplification, we evaluate its quality using three metrics, each on a scale of 1-5: *adequacy*, how much original content the simplification retains; *fluency*, the grammatically of the simplification; and, *simplicity*, how simple the sentence is [3, 6, 10, 13]. For each metric, we solicited three evaluations using Amazon Mechanical Turk and averaged the results.

[2]www.wikipedia.org, www.cochrane.org

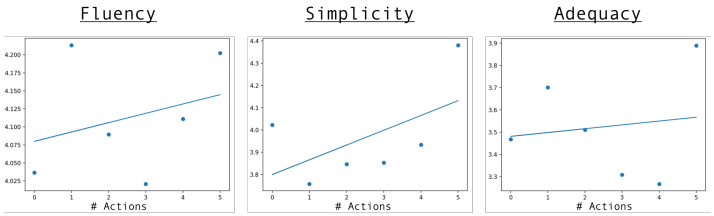

Figure 3: Relationship between the number of collaborative actions and simplification metric ratings. From left to right: $R^2 = 0.29, p = 0.57$; $R^2 = 0.56, p = 0.24$; $R^2 = 0.16, p = 0.76$.

Table 1: Simplification metric ratings by collaborative environment. Values are displayed as a mean, 1(worst)-5(best), followed by a standard deviation. [1] denotes means significantly different from the individual case at the $p < 0.05$ level.

	Adequacy	Fluency	Simplicity
Individual	3.47 (0.73)	4.02 (0.67)	4.02 (0.58)
All-Access	3.36 (0.84)	4.10 (0.50)	3.82 (0.64)
Highlighting	3.57 (0.77)	4.00 (0.75)	3.81 (0.64)
Chatbox	3.62 (0.82)	4.35[1] (0.51)	3.91 (0.63)

RESULTS

We collected over 90 simplifications for each of the four simplification environments from contributions by over 190 unique Mechanical Turk participants. Table 1 shows the results for each of the four environments along the three metrics. While there is some variation in the quality of the simplifications across environments, there is also a high degree of noise. Typically, crowdsourcing simplifications from non-professionals requires some form of pre- or post-processing [1, 14]. However, even with non-professionals, using just the chatbox does significantly improves the fluency over the non-collaborative environment. The variances are notably larger in the collaborative environments than the individual cases for the non-significant results.

To better understand the effect of the different collaborative environments we also measured the number of *collaborative actions*, which we define as the number of highlights, chats, and the number of users directly contributing to final simplification. This value is a proxy for the quality of the collaboration and measures how much interaction happened between the users. While there are only a small number of data points, Figure 3 displays some positive correlation between collaborative actions and metric ratings. This relationship is particularly strong for simplicity.

We also analyzed the change in simplification quality as the pairs worked on more sentences. We find that there exists a positive correlation between how many sentence pairs the partners had worked on already and each of the three simplicity metric ratings. In other words, the pairs got better at simplifying over time. The correlation is particularly strong for the fluency and adequacy metrics ($p < 0.25$ and $R^2 > 0.6$ for both). For the individual simplifications there is no such improvement over the sequence of the simplifications, hinting that the correlation is from more than familiarity with the tool and task. There was no significant difference between the timed and untimed variants for any environments.

While these quantitative results provide several insights, the high degree of noise requires more data and experimentation to be conclusive. Observationally, we find several interesting patterns in the

chatbox logs to supplement the quantitative analysis. Figure 1 illustrates an example of an effective collaboration. In this case, one user, User 1, identifies that the phrase 'micro-metastic' is cancer-related. The other user, User 2, then types out the majority of the simplification with User 1 contributing a few characters. User 1 then sends a chat indicating that the simplification looks good and the pair moves on.

The interaction in Figure 4 demonstrates both an advantage and disadvantage of crowdsourced collaboration. In this case, User 1 does not appear able or willing to construct the simplification. They ask and rely on User 2 to provide the simplification. Consequently, there is no synergistic benefit from the collaboration. Additionally, this collaboration is notably less effective than the collaboration in Figure 1. And yet, the *collaborative actions* proxy gives it a higher score. One positive outcome in this case is that User 2 provides a high quality simplification that is likely better than the simplification User 1 would have offered on an individual assignment.

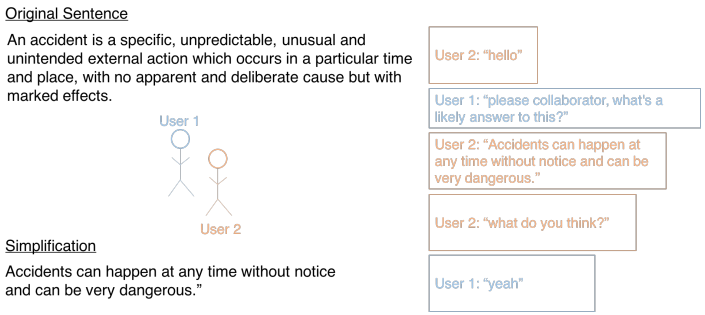

Figure 4: Worker collaboration example for the Chatbox environment.

However, there were many bad collaborative examples. These mostly include collaborators that provide zero insight or support for their partner. Often these individuals would not send messages in the chatbox or highlight portions of text. Because of this behavior, as well as the example in Figure 4, it can be difficult to classify a collaboration as effective or essentially individual.

DISCUSSION

Overall, collaboration can affect and improve the quality of the text simplification, particularly for fluency. The increase in fluency is not entirely surprising. We anticipate that collaborative simplification is better suited for fluency because catching a partner's grammatical mistakes is easier and more natural than suggesting a new simplification. Moreover, as users spend more time collaborating, the

quality of the simplifications improve. This does not happen in the individual case, suggesting that the improvements are from continued collaboration and better collaborative dynamics rather than familiarity with the tool or simplification task.

Despite these encouraging results, the data collected was noisy. In future work, researchers must make careful design considerations when collecting crowdsourced collaborative work. These include establishing effective and positive digital collaborative environments and partnerships before the task begins (as suggested by Kvan [5]), designing productive and easy-to-use collaborative user interfaces, and ensuring workers are both capable and adequately motivated.

REFERENCES

[1] Michael S. Bernstein, Greg Little, Robert C. Miller, Bjorn Hartmann, Mark S. Ackerman, David R. Karger, David Crowell, and Katrina Panovich. 2010. Soylent: A Word Processor with a Crowd Inside. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*.

[2] Nichole Egbert and Kevin M Nanna. 2009. Health literacy: Challenges and strategies. *The Online Journal of Issues in Nursing* 14, 3 (2009).

[3] 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*. 1–10.

[4] David A Kindig, Allison M Panzer, Lynn Nielsen-Bohlman, et al. 2004. *Health literacy: a prescription to end confusion*. National Academies Press.

[5] Thomas Kvan. 2000. Collaborative design: what is it? *Automation in construction* 9, 4 (2000), 409–415.

[6] Max Schwarzer and David Kauchak. 2018. Human Evaluation for Text Simplification: The Simplicity-Adequacy Tradeoff. In *Proceedings of SocalNLP Symposium*.

[7] Chirag Shah and Roberto González-Ibáñez. 2011. Evaluating the Synergic Effect of Collaboration in Information Seeking. In *Proceedings of the International ACM SIGIR Conference on Research and Development in Information Retrieval*.

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

[9] Jon Sprouse. 2011. A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods* 43, 1 (2011), 155–167.

[10] Elior Sulem, Omri Abend, and Ari Rappoport. 2018. BLEU is Not Suitable for the Evaluation of Text Simplification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

[11] John A Vernon, Antonio Trujillo, Sara J Rosenbaum, and Barbara DeBuono. 2007. Low health literacy: Implications for national health policy. (2007).

[12] Judith D Wilson, Nathan Hoskin, and John T Nosek. 1993. The benefits of collaboration for student programmers. *ACM SIGCSE Bulletin* 25, 1 (1993), 160–164.

[13] Sander Wubben, Antal van den Bosch, and Emiel Krahmer. 2012. Sentence Simplification by Monolingual Machine Translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*.

[14] Omar F. Zaidan and Chris Callison-Burch. 2011. Crowdsourcing Translation: Professional Quality from Non-professionals. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*.