Target-Level Sentence Simplification as Controlled Paraphrasing

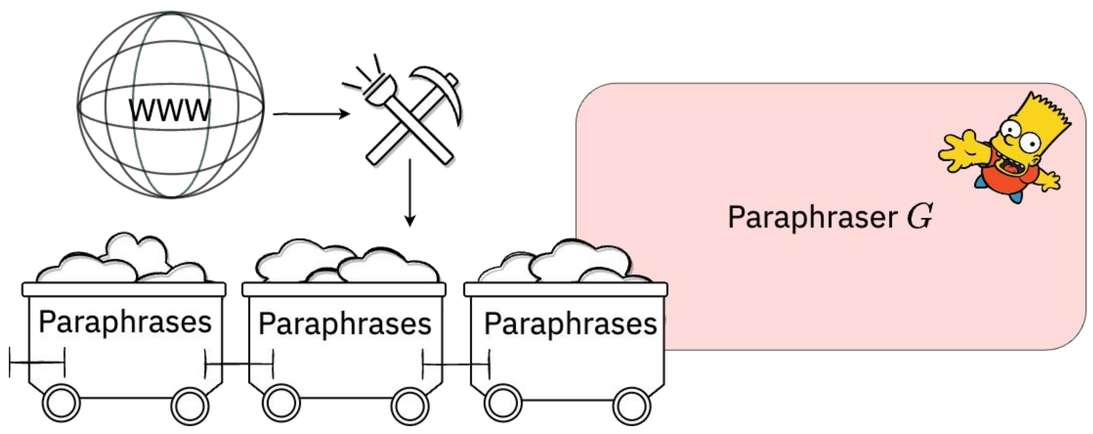

Tannon Kew, Sarah Ebling
Department of Computational Linguistics
University of Zurich

Motivation

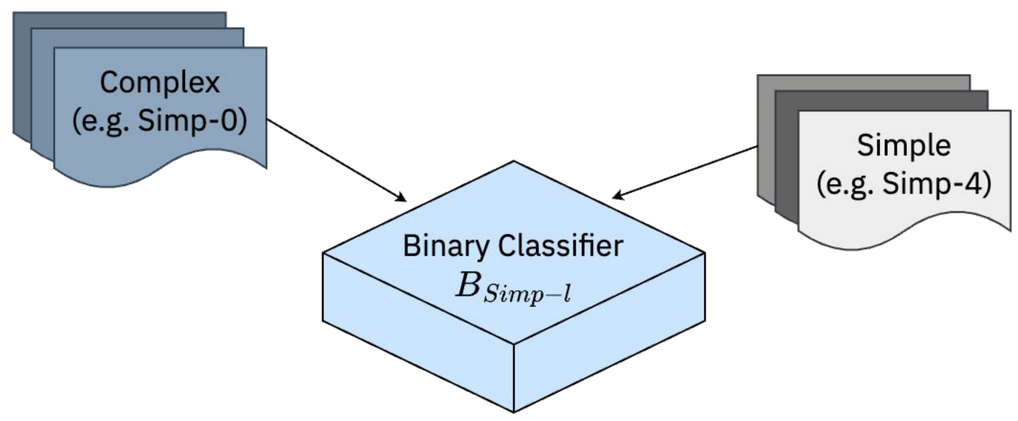
> What is simple for one target audience may not be simple for another.
> Simplification systems should be easily adaptable to serve the needs of different types of readers.[1]
> High quality complex-simple parallel data is scarce. Paraphrases can be mined from the web more easily.[2]
> Can we use controlled decoding techniques to steer paraphrastic generation towards simple text?

Method: paraphrase and control with FUDGE[3]

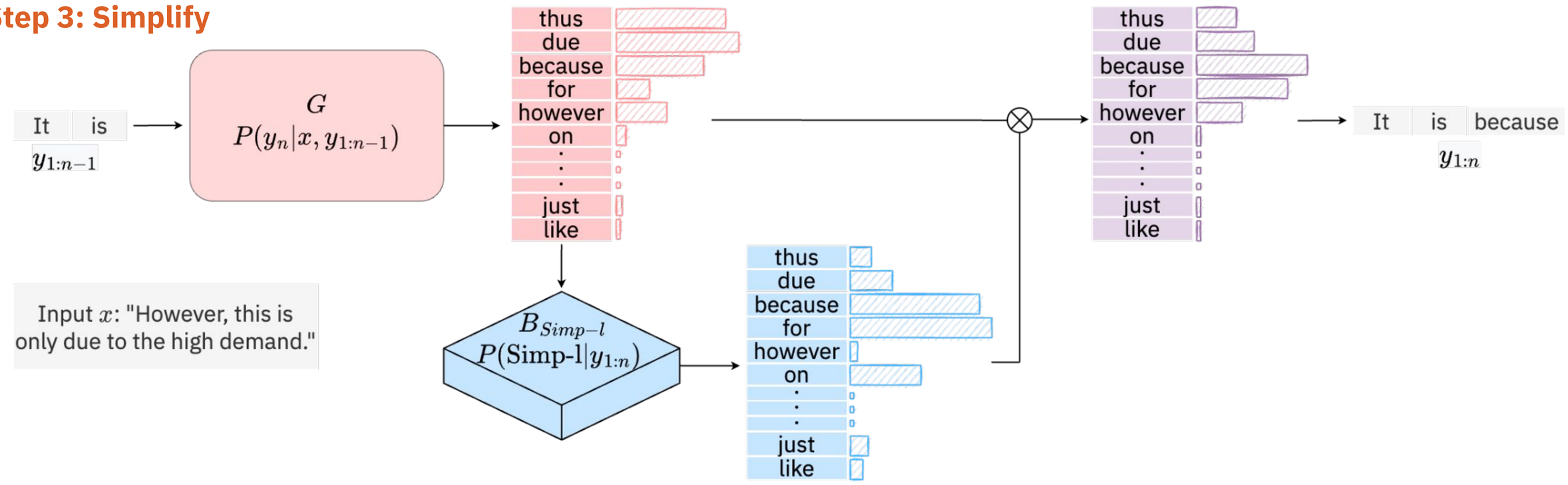
Step 1: Train Generator



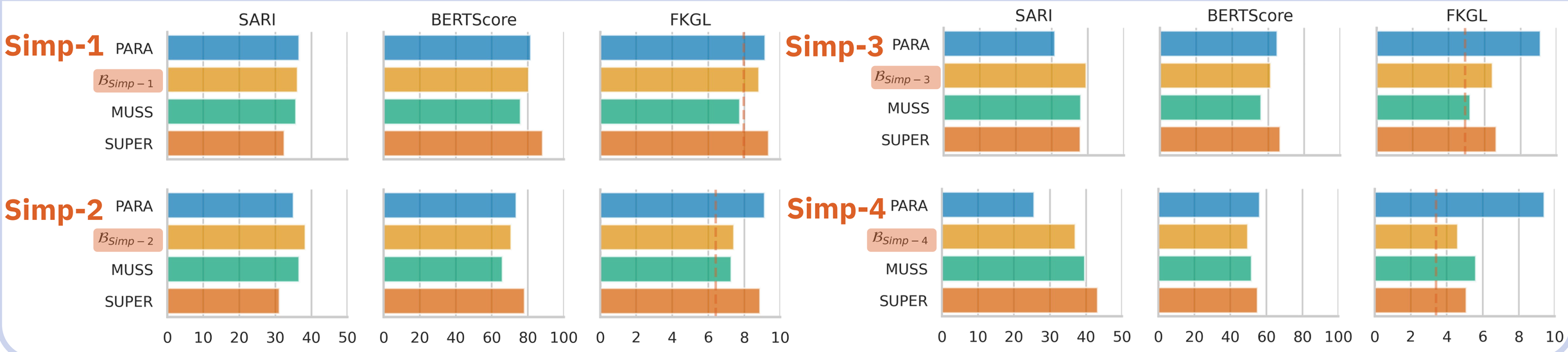
Step 2: Train Discriminator



Step 3: Simplify



Results



Conclusion

> Sentence simplification with **FUDGE** is **modular** and **easily adaptable**, making it possible to accommodate different **target levels** in Newsela (English).
> Discriminator training requires **no parallel aligned** data.
> Simple to use: strength of the discriminator is controlled by a single hyperparameter.
> Automatic metrics show performance is competitive with SOTA methods.

University of Zurich[UZH]

Paper & Code:

[1] Sanja Štajner (2021) Automatic Text Simplification for Social Good: Progress and Challenges. ACL-IJCNLP 2021.
[2] Louis Martin, Angela Fan, Éric de la Clergerie, Antoine Bordes, and Benoît Sagot (2021) MUSS: Multilingual Unsupervised Sentence Simplification by Mining Paraphrases. arXiv:2005.00352
[3] Kevin Yang and Dan Klein (2021) FUDGE: Controlled text generation with future discriminators. NAACL 2021.