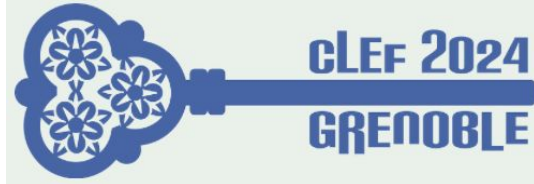




Universität
Zürich^{UZH}

Department of Computational Linguistics



SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding

Andrianos Michail, Pascal Severin Andermatt, Tobias Fankhauser

09/09/2024



Andrianos Michail

PhD Candidate
Department of Computational
Linguistics

Under the Impresso
II project:

Speak to me about
it!



Pascal Severin Andermatt

PhD Candidate
Department of Informatics



Tobias Fankhauser

MSc Artificial Intelligence

Special thanks to Simon Clematide



SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding

Presentation Layout:

1. Scientific Text Simplification
2. Minimum Bayes Risk Decoding
3. Multi-Prompt
4. Beyond Single Prompt Minimum Bayes Risk Decoding
5. Why Lens?
6. Complete Schematic
7. Results
8. Minimum Bayes Risk Decisions Analysis

SimpleText Best of Labs in CLEF-2023: **Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding**

Rewriting a text in language such that it is accessible by more readers. A good simplification must fulfill the following criteria:

- preserve original meaning
- fluency
- easier to comprehend than the original

Example:

| | |
|---------------------------|---|
| Source: | Furthermore, the same approach is applicable to similar malware analysis scenarios. |
| Simplification(1): | The same method can be used to analyze other types of malware. |
| Simplification(2): | The same method can be used to analyze other types of malicious software, such as harmful code designed to steal sensitive information or disrupt normal computer functions, to identify its characteristics, behavior, and potential impact, and understand and counter its threats. |

SimpleText Best of Labs in CLEF-2023: **Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding**

Dataset of Simpletext 2023/2024 Task 3/3.1:

Training set of 648 parallel training sentences:

Used to domain adapt **Alpaca-LoRA 7B** models on four different prompts (2023)

Test Set: 577 Sentences

Simplifications candidates generated through:

1. Domain-adapted **Alpaca-LoRA 7B** models (2023)
2. Off-the-shelf **LLama 3 8B** (2024)

SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using Multi-Prompt **Minimum Bayes Risk Decoding**

Source input for text generation
Ex. MT, Text Simp.

Candidate Generation method
Ex. beam search

Task Dependent Evaluation metric

Shared Utility (higher is better)
Ex. Score

Source Name the best things about a fruit salad:

Cand1

Juicy

evaluation metric(source=Source, hyp=Cand1, references=[Cand2...4])

0.33

Cand2

Refreshing

....

0.33

Cand3

Healthy

....

0.33

Cand4

Juicy, refreshing and healthy

evaluation metric(source=Source, hyp=Cand4, references=[Cand1...3])

1.00

→ **Best Candidate**

SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using **Multi-Prompt** Minimum Bayes Risk Decoding

Candidates given into Minimum Bayes Risk come from the same prompt on the same model weights, typically generated through beam search.

What we do differently is:

- Each candidate is generated by a different prompt
- We generate candidates through also prepending intermediate definitions of complex terms in the prompt so that they can be included in the simplification (*Through ID*)

| Prompt Name | Prompt | Intuition |
|------------------------------------|--|--|
| P1: General (+ Through ID variant) | Simplify the following scientific sentence to make it more understandable for a general audience | This prompt aims to rephrase complex scientific content into a format that is easier for a general audience to grasp, enhancing overall accessibility and comprehension. |
| P2: 5Y (+ Through ID variant) | P2: 5Y & Simplify the following scientific sentence. Explain it as if you were talking to a 5-year-old, using simple words and concepts: | Break down complex words into their most fundamental elements, using very simple language and concepts that even a child could understand. |

SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using **Multi-Prompt Minimum Bayes Risk Decoding**

Why Lens? (2023)

Technically, any simplification metric can be used as the utility of Minimum Bayes Risk. However, the ones that by default can aggregate multiple “reference” candidates are more straightforward to apply:

A Learnable Evaluation Metric for Text Simplification[3]

- **LENS** best correlates with humans - trained on different kinds of text simplifications.
- **SARI** fails to capture simplifications that deviate lexically (*first two examples*)

^ : Deletion
V...V : Paraphrase
|| : Splitting

Simplifications:

From GPT-3.5 davinci-003 (zero shot):

More than 200 ^Vbrand-new, ^Vcolored^V photographs ^Vhave been taken^V by British architectural photographer Morley von Sternberg for this book. ^{||} ^VAdditionally,^V older images ^Vhave been used^V to ^Vshow^V buildings that have ^Vchanged^V or ^Vbeen destroyed^V.

SARI 0.410 **BERTScore** 0.919 **Human** 🗑️ 85.67 **LENS** 🔍 1.022

From T5-3B:

The book ^Vhas^V more than 200 newly commissioned ^Vcolour^V photographs by the British architectural photographer Morley von Sternberg. ^{||} ^VThere are also^V older images where needed to ^Vshow^V buildings that have been significantly altered or demolished.

SARI 0.454 **BERTScore** 0.932 **Human** 🗑️ 74.33 **LENS** 🔍 0.731

From MUSS (Martin et al., 2022):

The book is illustrated with more than 200 newly commissioned color photographs by the British architectural photographer Morley von Sternberg, as well as older images [^] that have ^Vnot been used a lot since the book was first published^V.

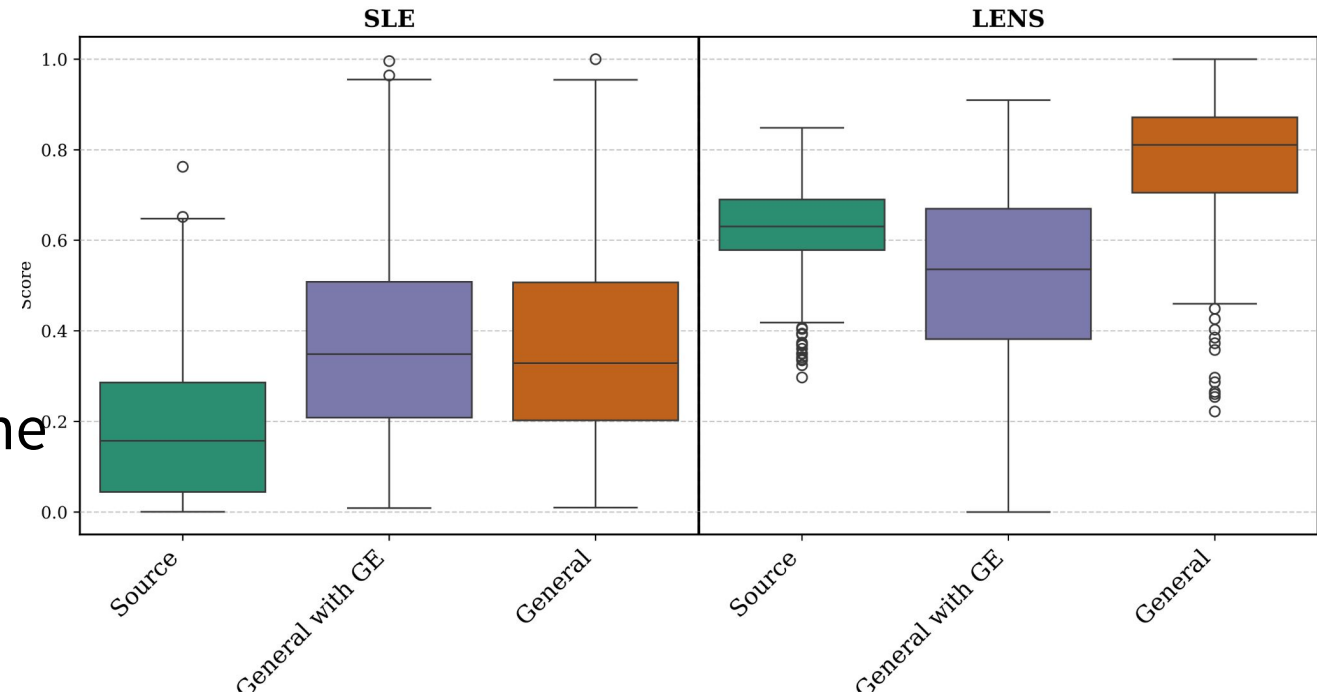
SARI 0.462 **BERTScore** 0.907 **Human** 🗑️ 46.67 **LENS** 🔍 0.476

M. Maddela, Y. Dou, D. Heineman, W. Xu, Lens: A learnable evaluation metric for text simplification (2023). arXiv:2212.09739.

SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using **Multi-Prompt Minimum Bayes Risk Decoding**

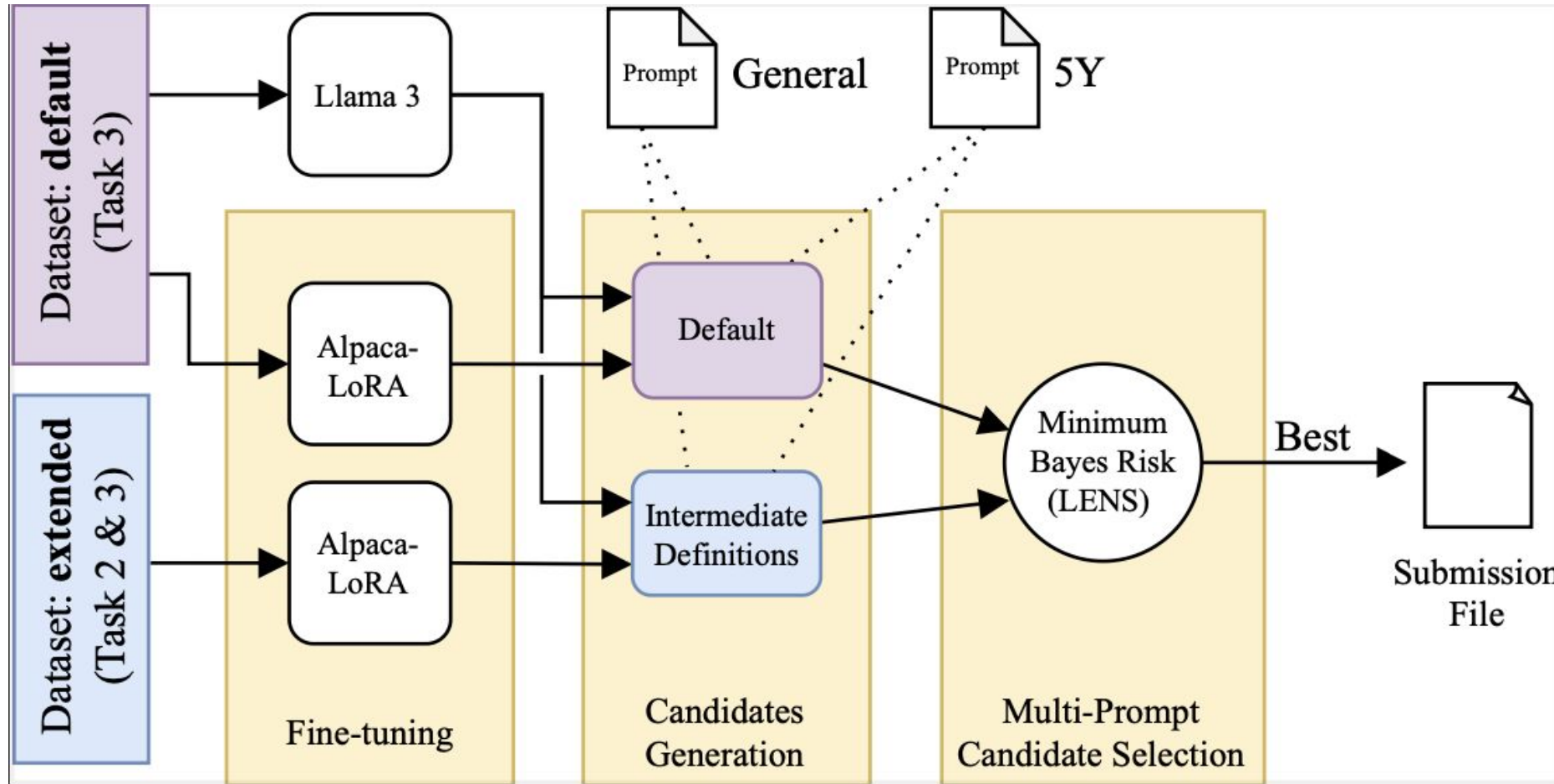
Why Lens? (2024)

Since 2023, Simplicity Level Estimate(SLE)[**1**] claims to have higher correlation with human annotators. We augment simplification candidates with **Grammar errors (GE)** and observe how the metric behaves.



-> We stick with LENS

SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using **Multi-Prompt Minimum Bayes Risk Decoding**



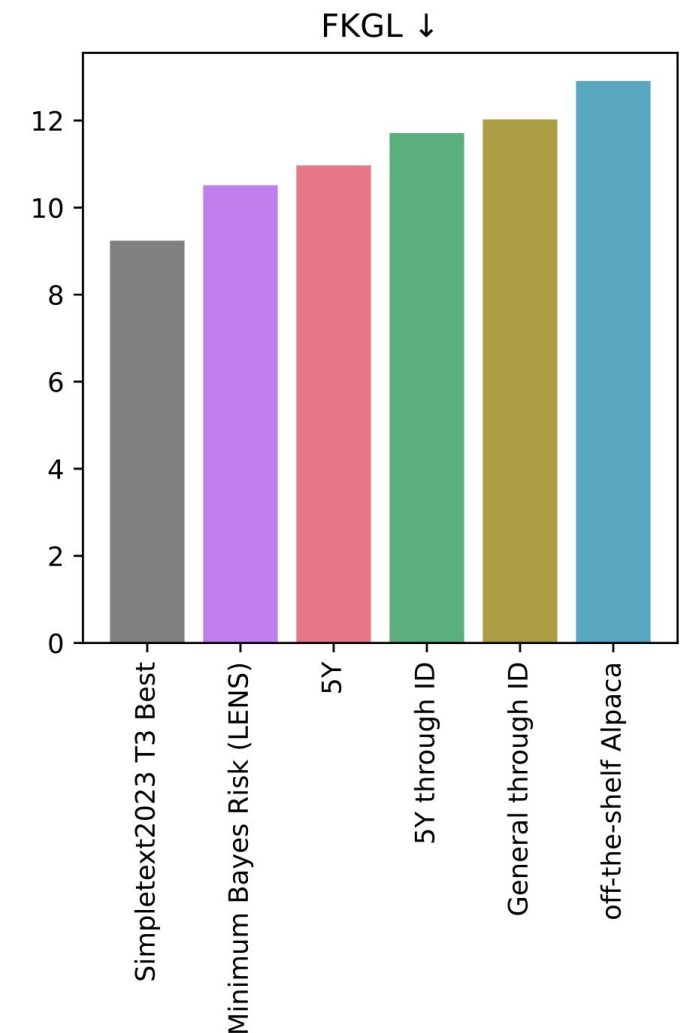
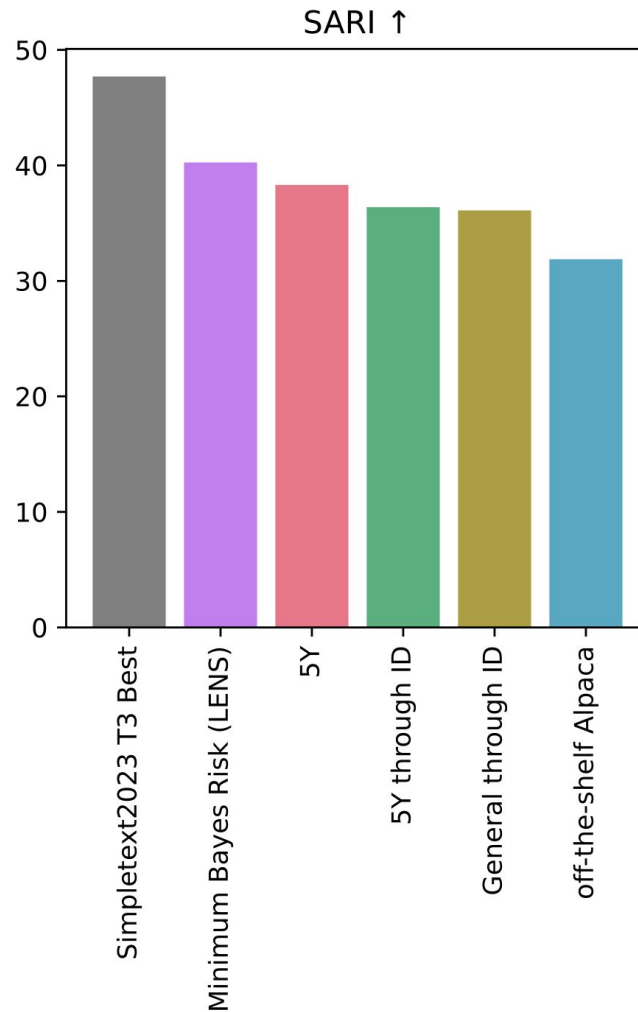
SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding

Results 2023 (Alpaca LoRA):

👍 Domain adaptation improved performance

👍 Different prompts exhibited different performance, ranging per metric

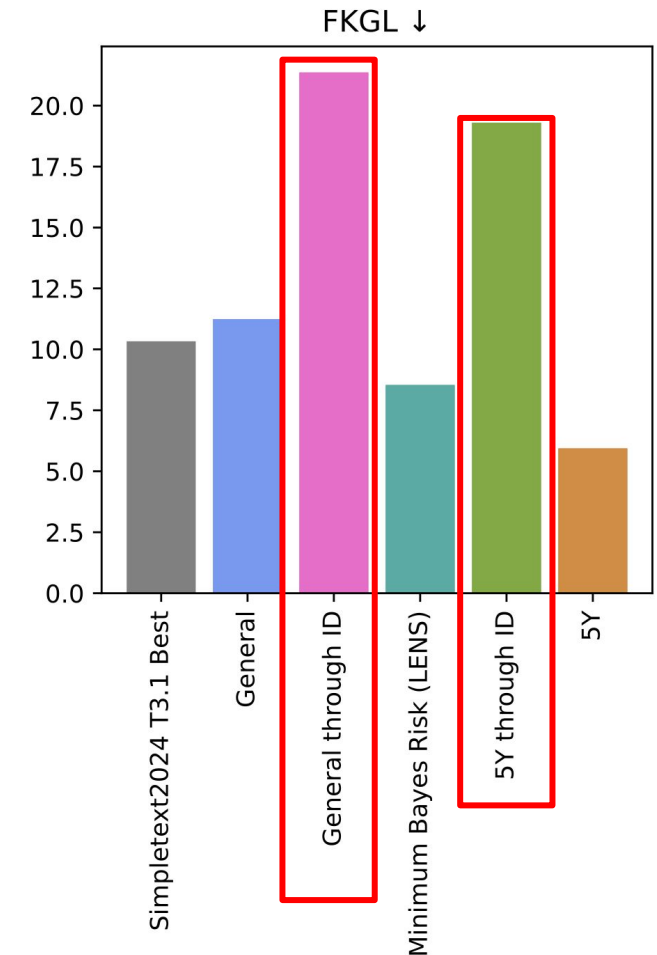
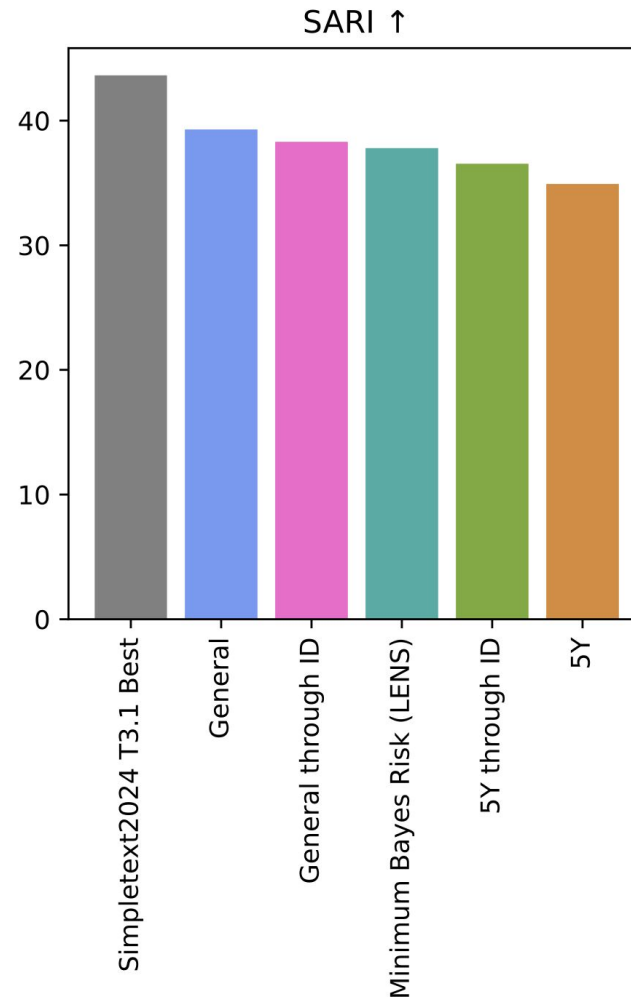
👍 Minimum Bayes Risk improves performance over the best prompt



SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding

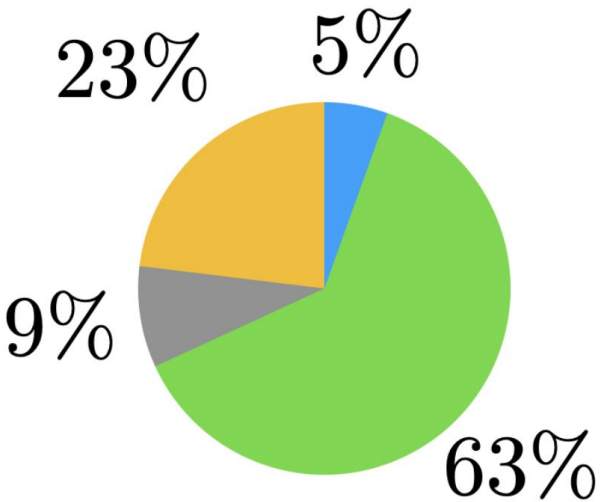
Results 2024 (LLama3):

- 👍 Different prompts exhibit different performances
- 👎 5Y prompt seems to oversimplify text (very low FKGL)
- 👎 Minimum Bayes Risk performs worse than the best prompt.

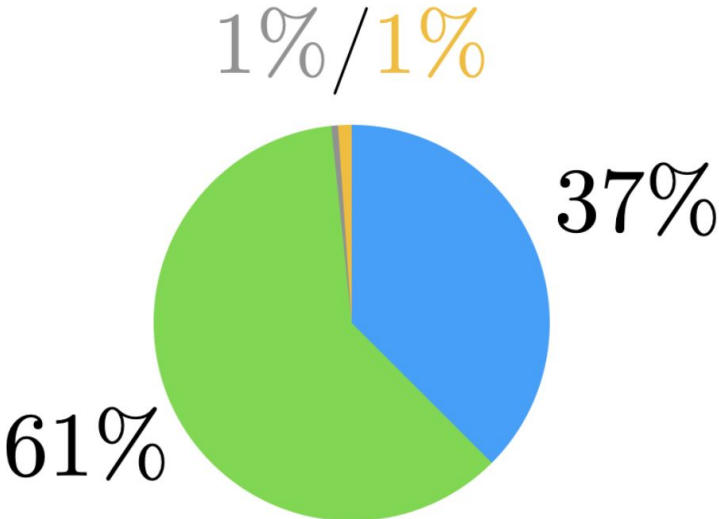


SimpleText Best of Labs in CLEF-2023: Scientific Text Simplification Using Multi-Prompt Minimum Bayes Risk Decoding

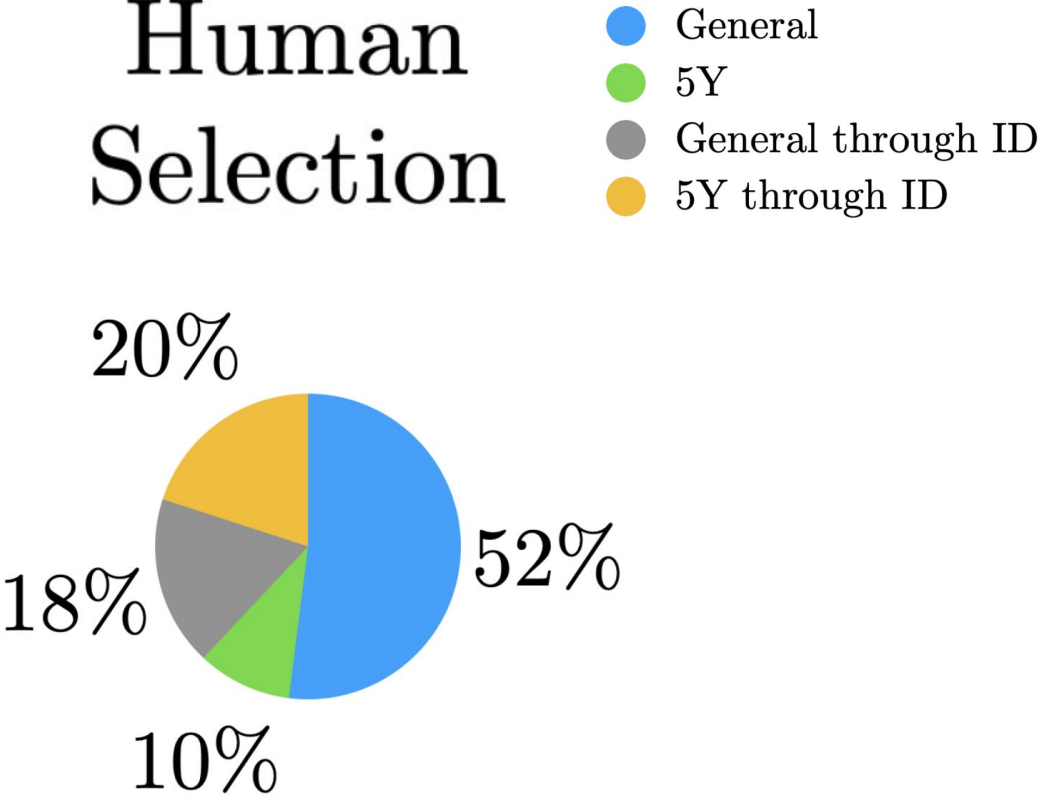
Domain Adapted
Alpaca LoRA



off-the-shelf
Llama3



Human
Selection



- General
- 5Y
- General through ID
- 5Y through ID

Thank you for your attention.

Questions?

Bibliography

- [1] Cripwell, L., Legrand, J., & Gardent, C. (2023, December). *Simplicity Level Estimate (SLE): A learned reference-less metric for sentence simplification*. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 12053-12059).
- [2] Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., ... & Ganapathy, R. (2024). *The llama 3 herd of models*. *arXiv preprint arXiv:2407.21783*.
- [3] Maddela, M., Dou, Y., Heineman, D., & Xu, W. (2023, July). *LENS: A learnable evaluation metric for text simplification*. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 16383-16408).