Med-EASi: Finely Annotated Dataset and Models for Controllable Simplification of Medical Texts

Chandrayee Basu*, **Rosni Vasu+**, **Michihiro Yasunaga***, **Qian Yang$**

*Stanford University,
+University of Zurich,
$Cornell University

THE 37TH AAAI CONFERENCE ON ARTIFICIAL INTELLIGENCE
FEBRUARY 7-14, 2023 - WASHINGTON, DC, USA
WALTER E. WASHINGTON CONVENTION CENTER

AAAI-23

Vision

Controllability to simplify communication between care team and patient

Image source: https://www.grocerresourceministries.org/blog/let-us-talk-health-literacy
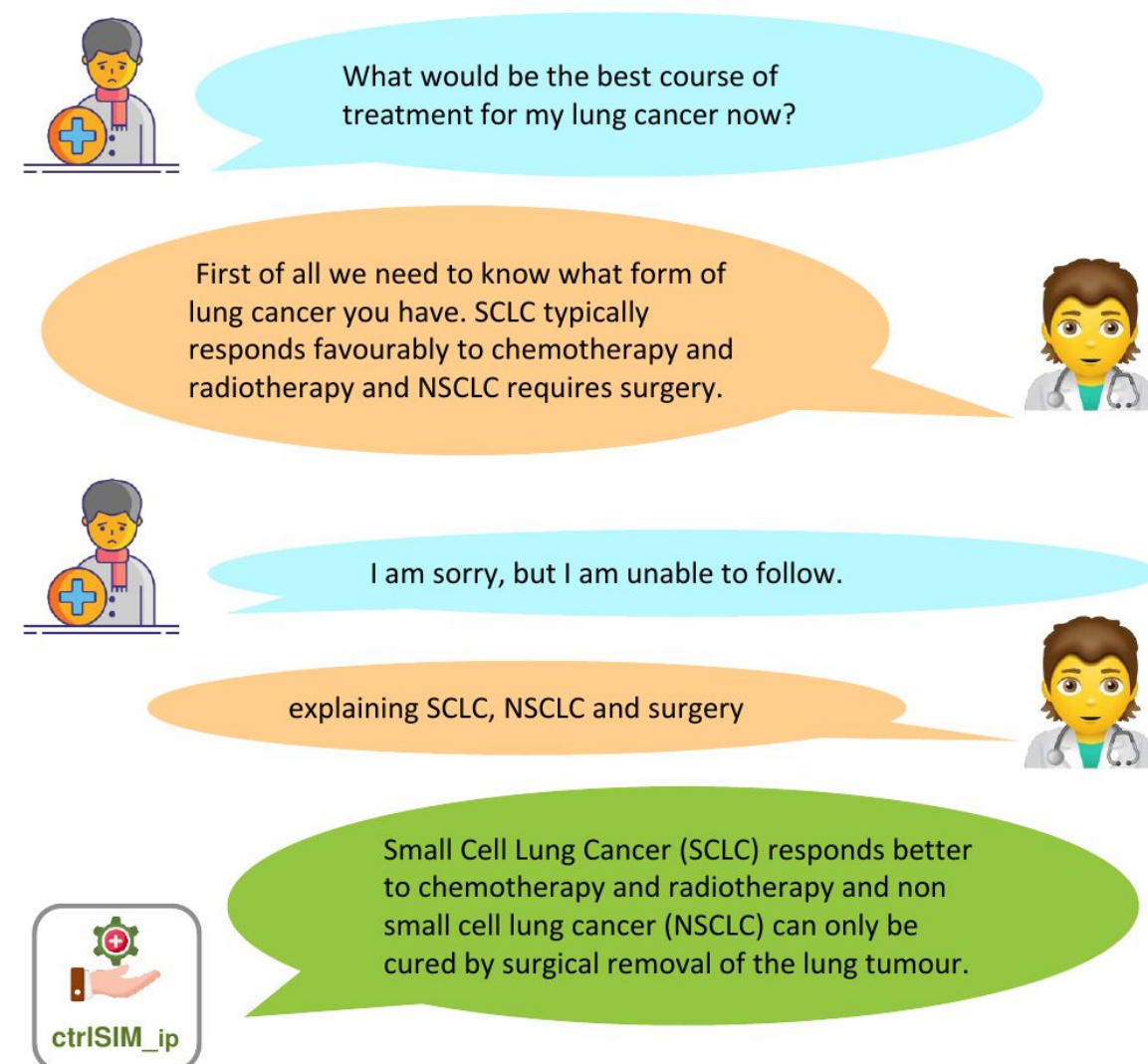
Large parallel corpora
Wikismall:108K pairs

Multi-references manually generated
ASSET: 10, Turkcorpus: 8, Newsela: 4

Reference generation follows human style of editing
ASSET

Controllability over simplicity
Newsela

Controllability over transformation type
CUT, EditNTS, ACCESS, with explicit paraphrasing

General text simplification — Medical Text simplification

Small parallel corpora
(Simpwiki:6K pairs)

Large non-parallel corpora
(MSD:100K+)

Single reference per medical text

References written by professional translators (MSD)
drawback automatic alignment

No controllability over simplicity
grade levels vary by pairs

No control over transformation type
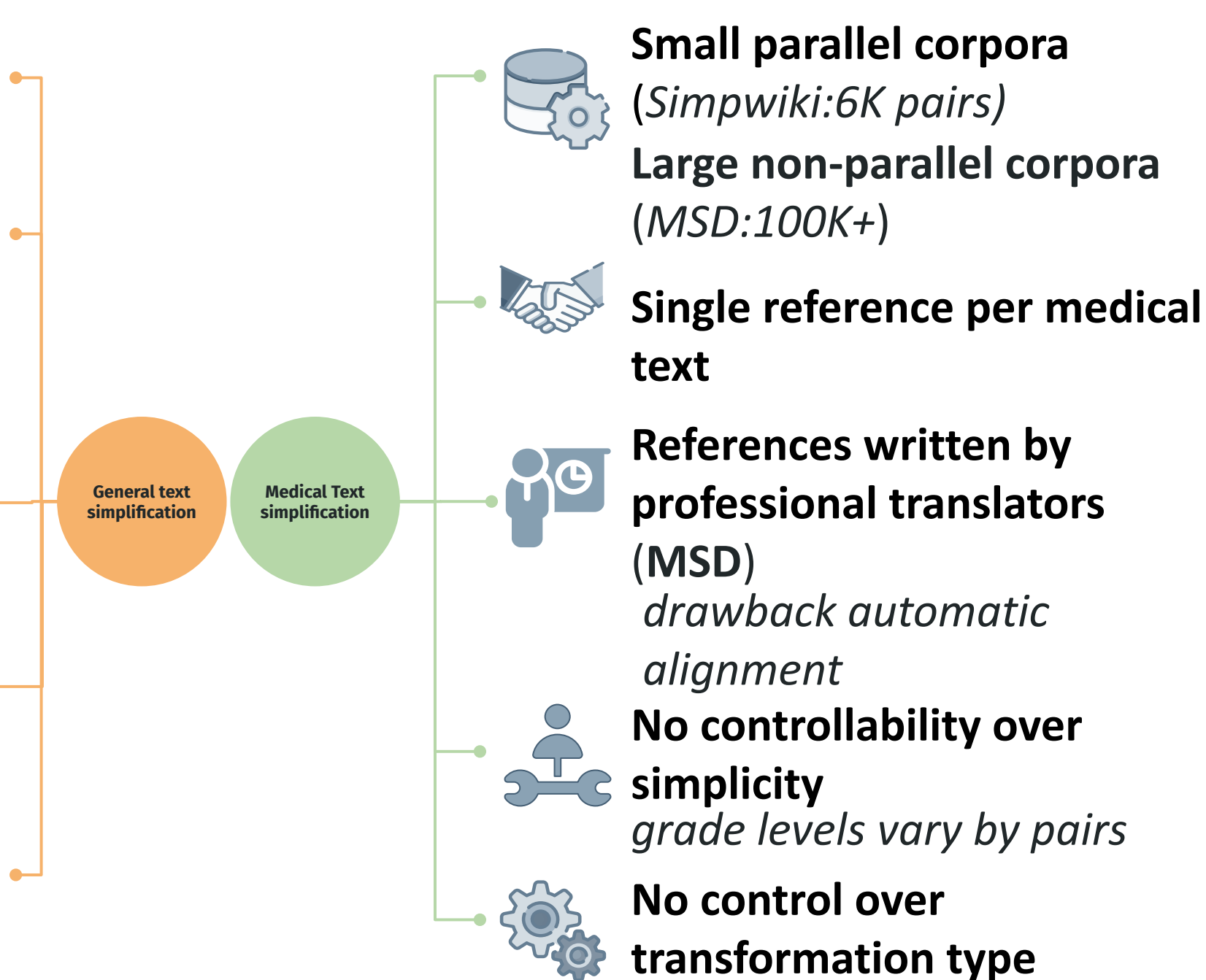
Textual Transformations in Simplification

Rewrite the complex text into the simple text while retaining the meaning of the actual text.

Involves four kinds of textual transformations: Deletion, Insertion, Replacement, and Elaborations

Expert: NSCLC is treated with surgery, whereas SCLC usually responds better to chemotherapy and radiotherapy. Worldwide in 2012, lung cancer occurred in 1.8 million people and resulted in 1.6 million deaths.

Layman: Non-small cell lung cancer is more commonly treated with surgical removal of the lung tumor, and Small cell lung cancer responds well to chemotherapy and radiotherapy.

Our Contributions

MSD + SIMPWIKI (1500 pairs)= Med-EASi
- 1979 expert-simple text pairs
- 4478 unique UMLS concepts

Dataset

	Med-EASi
Levenshtein similarity	0.689±0.22
Fraction of added, deleted and kept words	9.347±11.57, 10.574 ± 12.38, 12.792 ± 10.01
Compression ratio	1.025 ± 0.59
Grammatical Acceptability	expert texts: 98.989% layman texts: 98.737%
Kincaid Flesch Grade Level	expert texts (12.47 ± 5.28) layman texts (10.491 ± 4.98)

- Layman crowd-workers recruited from Toloka AI (crowdsourcing platform) and medical experts recruited from Upwork (freelancing platform).

References

Cao, Yixin, et al. "Expertise Style Transfer: A New Task Towards Better Communication between Experts and Laymen." *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020.
Van den Bercken, Laurens, Robert-Jan Sips, and Christoph Lofi. "Evaluating neural text simplification in the medical domain." The World Wide Web Conference. 2019.
Tafjord, Oyvind, and Peter Clark. "General-purpose question-answering with macaw." arXiv preprint arXiv:2109.02593 (2021).

Expert-Layman-AI Annotation

AI assisted layman annotation → Result aggregation

Text pair with Levenshtein similarity > 0.7 for annotation

Difflib annotation: Hypoglycemia, <ins>also known as low blood sugar<ins>, is the fall of blood sugar levels below normal.

Coref-res annotation: <elab>Hypoglycemia<by>, also known as low blood sugar</elab>, is the fall of blood sugar levels below normal.

Answers from 3 annotators

Dawid-Swene aggregation confidence score < 90 %

Expert annotation

Text pair for annotation

Expert answer: <elab>Hypoglycemia<by>Hypoglycemia, also known as low blood sugar</elab>, is the fall of blood sugar levels below normal.

Multi-Angle Controllable Simplification Model

Controllable Models

- **Human interpretable controllability:** a user should be able to indicate the complex contents that must be deleted, replaced, or elaborated.
- End-to-end model for all types of simplification edits.
- **Multi-angle training:** Following MACAW (Tafjord and Clark 2021), our approach use **multi-angle training**, where each component of the input and the output is a **slot** and the input-output combination is called an **angle**.
- **Model backbone:** Flexible input-output structure of T5 using a combination of infilling and prompting
- **Model Versions:** control-free versions (SIM and SIM_{ip}) and controllable versions (ctrlSIM and $ctrlSIM_{ip}$)

requested slots/instructions   input slots

$replace$; $simple$; $expert$ = Ankles, knees, elbows, and wrists are usually involved. ; $replace\_in$ = [involved]

output slots

$replace$ = [involved <by> affected] ; $simple$ = Ankles, knees, elbows, and wrists are usually affected.

Evaluation

- ctrlSIM model with angle ERi → RS achieves highest SARI score of 0.49
- $ctrlSIM_{ip}$: average SARI score vary by angles

Automatic Evaluation

models	SARI	ADD	DEL	KEEP	FKGL	ROUGE-1
ctrlSIM	39.63	8.99	70.67	39.2	10.55	0.41
ctrlSIMP$_{ip}$	40.2	8.51	70.07	42.04	11.09	0.43

Human Evaluation

- Fluency: mostly fluent
- Grammaticality: 40% grammatically correct model output
- Output annotation quality
  - 22% of the time model performed what user have asked for
  - Model generated mostly matches the replaced content
  - Elaborated contents are mostly relevant and somewhat satisfactory

Next Step

Improve elaboration using additional contexts (definitions and description of complex medical terms)