# Identifying and Simplifying Non-consumer Terminology in Biomedical Abstracts

**Bill Xia**, Dr. Brian Ondov, Dr. Dina Demner-Fushman
Lister Hill National Center for Biomedical Communication, National Library of Medicine

## INTRODUCTION

**Motivation**

- Medical literacy informs healthcare decision-making.
- Education level is linked to health disparities.
- Biomedical knowledge remains inaccessible due to jargon.

**Tools**

- Rules-based models are simple to implement but limited in capabilities.
- Advances in Deep Learning (LLMs) have unlocked new methods for accurate adaption of scientific texts.

**Task**

- **Identification** – Identify non-consumer terms.
- **Generalization** – Replace terms with a more general category.

  *"ring suture" → "eye procedure"*
- **Omission** – Remove irrelevant / overly technical terms.

## METHODS

**Identification**

- **Baseline** – MetaMapLite for identification, filtered by term frequency.
- **LLM** – BERT fine-tuned for a Named Entity Recognition (NER) task.

```
"Patients received a ring suture"
 O        O        O B    I
```

**Generalization**

- **Baseline** – UMLS parent terms used to find generalized categories.
- **LLM** – BART fine-tuned for a sequence-to-sequence (seq2seq) task.

*"Patients received a ring suture."*
↓
*"eye procedure"*

**Omission**

- **LLM** – BART fine-tuned for a seq2seq task with a T5-based grammar correction model (T5-GCM).

## RESULTS

### Identification

| Model | Avg F1 | ∪ F1 | ∩ F1 | Pyramid |
|---|---|---|---|---|
| Baseline | 0.2097 | 0.2487 | 0.1497 | 0.2916 |
| BERT-L | 0.3530 | 0.4260 | 0.2515 | 0.4891 |
| BioBERT-L | 0.3058 | 0.3898 | 0.2071 | 0.3938 |
| XLM RoBERTa-L | 0.3745 | 0.4596 | 0.2578 | 0.5147 |
| DeBERTa-L | 0.4317 | 0.5255 | 0.2976 | 0.6014 |

**Table 1.** Performance of each identifier model.

| Sentence | Model | Expert Terms |
|---|---|---|
| "Ring sutures induced cataract more frequently than other procedures." | Baseline | sutures, cataract |
| | BERT-L | Ring sutures, cataract |
| | BioBERT-L | Ring sutures, cataract |
| | XLM RoBERTa-L | Ring sutures, cataract |
| | DeBERTa-L | Ring, cataract |

**Table 2.** Example input sentence and terms identified by each identifier model.

### Generalization

| Model | ROUGE-1 | ROUGE-2 | ROUGE-L | BLEU | SARI | BERT-Score |
|---|---|---|---|---|---|---|
| Baseline | 0.8925 | 0.8458 | 0.8920 | 0.8535 | 65.15 | 0.9728 |
| BART | 0.9233 | 0.8802 | 0.9231 | 0.8760 | 66.66 | 0.9813 |
| BioBART | 0.9444 | 0.9108 | 0.9442 | 0.9138 | 78.31 | 0.9886 |

**Table 3.** Performance of each generalizer model.

| Sentence | Model | Generalizations |
|---|---|---|
| "Patients were treated with a [ring suture]." | Baseline | `None found` |
| | BART | surgery |
| | BioBART | ring suture |

**Table 4.** Example input sentence and simplification generated by each generalizer model.

### Omission

| Model | ROUGE-1 | ROUGE-2 | ROUGE-L | BLEU | SARI | BERT-Score |
|---|---|---|---|---|---|---|
| BART | 0.9191 | 0.8517 | 0.9199 | 0.8198 | 66.21 | 0.9609 |
| w/ T5-GCM | 0.8123 | 0.7412 | 0.8077 | 0.7156 | 53.49 | 0.9609 |

**Table 5.** Performance of each omission model.

| Sentence | Model | Adaption |
|---|---|---|
| "Patients were treated with a [ring suture]." | BART | "Patients were treated with a ." |
| | BART w/ T5-GCM | "Patients were treated with a ." |

**Table 6.** Example input sentence and simplification generated by each omission model.

## CONCLUSION

**Identification**

- LLMs offer performance significantly exceeding that of rules-based models.
- Domain-specific pretraining seems to harm performance.

**Generalization**

- Rule-based generative models rely on rigid vocabulary, so the baseline couldn't find generalizations for all terms.

**Omission**

- Even with the T5-GCM, the model often failed to produce coherent simplifications.

**Other Takeaways**

- Automated generative evaluation metrics fail to capture nuances of text simplification tasks.
  - Simplicity
  - Information accuracy
  - Grammatical correctness

## FUTURE WORK

**Further Evaluation**

- Human evaluation allows for a more nuanced evaluation of generative AI output.

**End-to-End Simplification**

- After identifying terms, a model should be able to predict the most appropriate method of simplification.
- Models need to be combined into a consumer-friendly application for the public to use them.

**Alternative Methods**

- Retrieval-Augmented Generation (RAG) allows models to generate text with greater accuracy and more up-to-date knowledge.
- Effective omission may require precise prompt engineering or better ways of maintaining grammatical correctness.

NIH U.S. National Library of Medicine