

Considering Human Interaction and Variability in Automatic Text Simplification

Jenia Kim

HU University of Applied Sciences
Utrecht, Netherlands



Stefan Leijnen

HU University of Applied Sciences
Utrecht, Netherlands



Lisa Beinborn

University of Göttingen
Göttingen, Germany



Introduction

- Real-world use case: simplification technology for municipalities in the Netherlands.
- Specific needs of the content creators, readers, organizations.
- What kind of simplification technology is needed to address these needs?

Problem

- 1 in 6 adults in the Netherlands has limited literacy.
- Recent analysis of 240 texts (payments, healthcare) from 70 government organizations → not understandable enough.

Writers' Challenges

- Guidelines are difficult to implement in practice
- Not enough expertise in the organization
- Diverse, heterogeneous audience
- Diverse contexts

Requirements

Needs

- Writers remain in control and are responsible for the final output.
- No "one-fits-all" solution.
- Technology is a source of expertise, from which writers can learn.

Co-creation system

- assists the human writer, without taking over the task;
- provides diverse outputs, tailored for specific target audiences;
- explains the suggestions that it outputs.

Simplification technology

- Modularity
- Variability
- Explainability

Human-AI Co-Creation

Individual: Human agency, control, ownership, meaningfulness.

Organization: Employees' expertise.

Society: Allocation of responsibility authorities vs. citizens.

Personalized Simplification

CEFR B1-level



Low literate native speakers



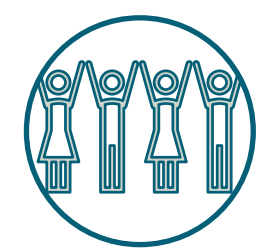
Non-native speakers



Intellectual disabilities



Dyslexia



...

Example

NL: Voor kinderen met een medisch hoog risico op ernstige ziekte door corona kan de behandelend (kinder)arts adviseren om in het najaar een injectie te halen.

EN: For children with a high medical risk for serious illness due to corona, the treating (pediatric) doctor may advise them to get an injection in the fall.

Word choice

Audience: native speakers with limited literacy
Suggestion: *injectie* → *prik*; *najaar* → *herfst*
Explanation: more common in everyday language

Audience: non-native speakers, English proficient
Suggestion: *arts* → *dokter*
Explanation: similar to the English *doctor*

Sentence length and structure

Audience: everybody
Suggestion: split into two sentences
Explanation: one message per sentence

Audience: everybody
Suggestion: delete *medisch*, *behandelend*, *(kinder)*
Explanation: not crucial for the main message

Audience: everybody
Suggestion: reorder sentence parts
Explanation: common word order *subject-verb-object*