LIS at SimpleText 2025: Enhancing Scientific Text Accessibility with LLMs and Retrieval-Augmented Generation

Anya Amel NAIT DJOUDI*, Sarah NOUALI*,

Mohsine AABID, Ismail BADACHE, Adrian-Gabriel CHIFU and Patrice BELLOT

Aix-Marseille Université, Université de Toulon, CNRS, LIS, Marseille, France

https://www.dei.unipd.it/~faggioli/temp/clef2025/paper_358.pdf

CLEF 2025, September 9–12, 2025, Madrid, Spain

The CLEF SimpleText Initiative

- Task 1: Text Simplification: Simplify scientific text
  - Subtask 1.1 - Sentence-level Scientific Text Simplification
  - **Subtask 1.2 - Document-level Scientific Text Simplification**

The Accessibility Challenge

Exemple 1

We included 68 studies randomising 9132 participants. We are very uncertain whether there is a difference between haloperidol and olanzapine in clinically important change in global state (RR 0.84, 95% CI 0.69 to 1.02; 6 studies, 3078 participants; very low-certainty evidence). We are very uncertain whether there is a difference between haloperidol and olanzapine in relapse (RR 1.42, 95% CI 1.00 to 2.02; 7 studies, 1499 participants; very low-certainty evidence).

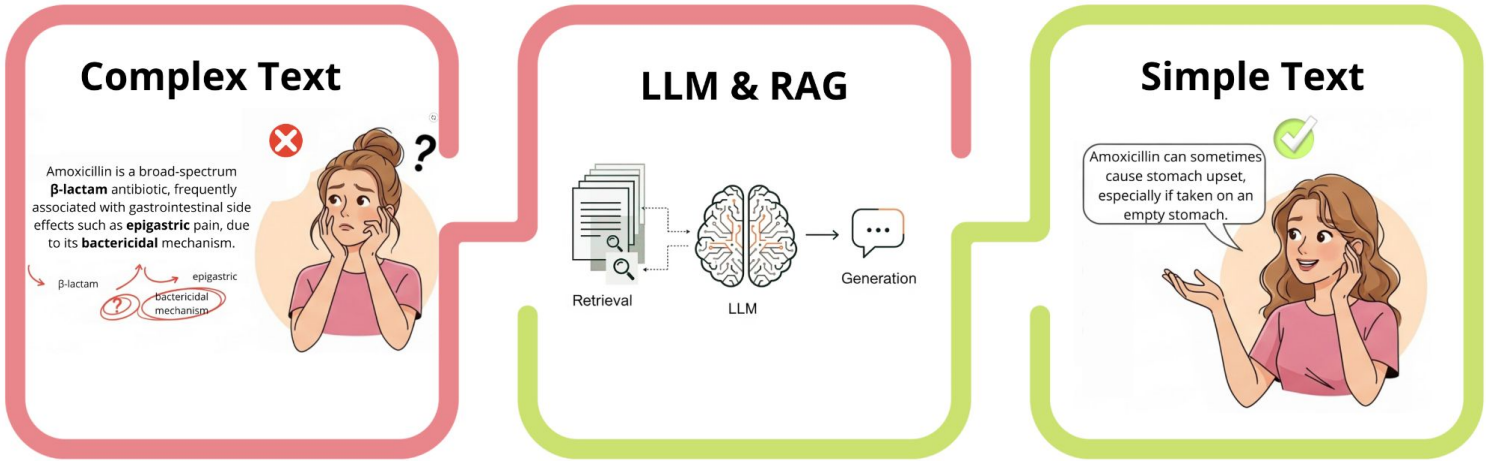
Exemple 2

We identified four eligible RCTs, all of which compared the Lidcombe Program to a wait-list control group. In total, 151 children aged between two and six years participated in the four included studies. In the Lidcombe Program, the parent and their child visit a speech and language therapist (SLT) in a clinic. One study conducted clinic visits by telephone. In each clinic visit, parents were taught how to conduct treatment at home. Two studies took place in Australia, one in New Zealand and one in Germany. Two studies were conducted for nine months, one for 16 weeks and one for 12 weeks. The frequency of clinic visits and practice sessions at home varied within the programme. One study was partially funded by the Rotary Club, Wiesbaden, Germany; and one was funded by the National Health and Medical Research Council of Australia. One study did not report funding sources and another reported that they did not receive any funding for the trial. All four studies reported the outcome of stuttering frequency.

The Core Problem

- Scientific publications are the primary medium for knowledge.
- They are locked behind walls of technical terminology.
- Without domain-specific knowledge even fluent readers struggle to comprehend scientific texts
- Large language models (LLMs), combined with retrieval-augmented generation (RAG), offer a promising approach to simplify scientific texts.
- How to enhance accessibility while preserving the factuality of scientific knowledge ?

Approach (1)


Complex Text
Amoxicillin is a broad-spectrum **β-lactam** antibiotic, frequently associated with gastrointestinal side effects such as **epigastric** pain, due to its **bactericidal** mechanism.
β-lactam
epigastric
bactericidal mechanism
?
LLM & RAG
Retrieval
LLM
Generation
Simple Text
Amoxicillin can sometimes cause stomach upset, especially if taken on an empty stomach.

Approach - RAG Pipeline (2)

Keyword Extraction

We use the KGR prompt to identify domain-specific terms in the complex text

Definition Retrieval

We match extracted keywords with corresponding definitions from our knowledge base - MedSimplify, a custom dictionary containing over 3,000 scientific definitions in layman's terms

Augmented Generation

Our best approach uses the DASP-0 prompt to generate simplified text by augmenting the model's knowledge with the retrieved definitions

KGR Prompt:

I have the following document: {Complex text}
Please give me the keywords that are present in this document and separate them with commas. Make sure you to only return the keywords and say nothing else. For example, don't say: "Here are the keywords present in the document"

MedSimplify

DASP_0 Prompt:

Using these definitions, please simplify the following scientific text for a general audience. Use plain language and explain any complex terms or acronyms. Ensure that all numbers, results, and facts remain exactly the same. Do not paraphrase numerical data or alter the meaning of findings.
DEFINITIONS: {List of definitions}
TEXT: {Complex text}

KGR: Keyword-Guided Retrieval Prompt / **DASP_0:** Definition-Augmented Simplification (Zero-Shot)

Results - Evaluation Phase (1)

Table 2
Overview of best submitted runs and corresponding SARI Scores

Phase	Run	Description	Score
Evaluation	Mistral_DASP_0	Definitions used, zero-shot prompting	**43.5051**
	Med42_DASP_0	definitions used, zero-shot prompting	41.9708
	Gemma2_DASP_1	definitions used, one-shot prompting	42.5695
	Mistral_DASP_1	Mistral 7b with one-shot prompting	42.4718
Post-Competition	Gemma3_DASP_0	Definitions used, zero-shot prompting	41.6236
	Mistral_baseline	No definitions, zero-shot prompting	42.9018
	Mistral_IRP	Mistral 7B with iterative refinement prompt (IRP)	43.1044
	Mistral_FKGL	FKGL-based output selection	**43.5120**

KGR: Keyword-Guided Retrieval Prompt / **DASP_0:** Definition-Augmented Simplification (Zero-Shot) / **DASP_1**: Definition-Augmented Simplification (One-Shot) / **IRP**: Iterative Refinement Prompt/ **FKGL**: Flesch-Kincaid Grade Level

Results - Evaluation Phase (2)

Hypothesis :

- if we use a model that was already trained on medical knowledge it might give us better results

Results - Evaluation Phase (3)

Table 2
Overview of best submitted runs and corresponding SARI Scores

Phase	Run	Description	Score
Evaluation	Mistral_DASP_0	Definitions used, zero-shot prompting	**43.5051**
	Med42_DASP_0	definitions used, zero-shot prompting	41.9708
	Gemma2_DASP_1	definitions used, one-shot prompting	42.5695
	Mistral_DASP_1	Mistral 7b with one-shot prompting	42.4718
Post-Competition	Gemma3_DASP_0	Definitions used, zero-shot prompting	41.6236
	Mistral_baseline	No definitions, zero-shot prompting	42.9018
	Mistral_IRP	Mistral 7B with iterative refinement prompt (IRP)	43.1044
	Mistral_FKGL	FKGL-based output selection	**43.5120**

KGR: Keyword-Guided Retrieval Prompt / **DASP_0:** Definition-Augmented Simplification (Zero-Shot) / **DASP_1**: Definition-Augmented Simplification (One-Shot) / **IRP**: Iterative Refinement Prompt/ **FKGL**: Flesch-Kincaid Grade Level

Results - Evaluation Phase (4)

Table 2
Overview of best submitted runs and corresponding SARI Scores

Phase	Run	Description	Score
Evaluation	Mistral_DASP_0	Definitions used, zero-shot prompting	**43.5051**
	Med42_DASP_0	definitions used, zero-shot prompting	41.9708
	Gemma2_DASP_1	definitions used, one-shot prompting	42.5695
	Mistral_DASP_1	Mistral 7b with one-shot prompting	42.4718
Post-Competition	Gemma3_DASP_0	Definitions used, zero-shot prompting	41.6236
	Mistral_baseline	No definitions, zero-shot prompting	42.9018
	Mistral_IRP	Mistral 7B with iterative refinement prompt (IRP)	43.1044
	Mistral_FKGL	FKGL-based output selection	**43.5120**

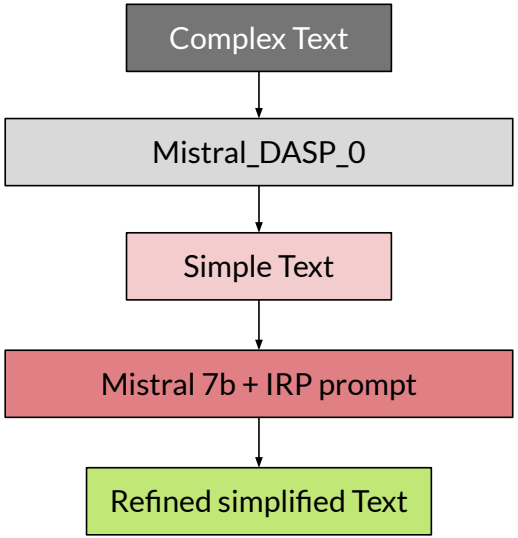
KGR: Keyword-Guided Retrieval Prompt / **DASP_0:** Definition-Augmented Simplification (Zero-Shot) / **DASP_1**: Definition-Augmented Simplification (One-Shot) / **IRP**: Iterative Refinement Prompt/ **FKGL**: Flesch-Kincaid Grade Level

Results- Post Competition Phase

Table 2
Overview of best submitted runs and corresponding SARI Scores

Phase	Run	Description	Score
Evaluation	Mistral_DASP_0	Definitions used, zero-shot prompting	**43.5051**
	Med42_DASP_0	definitions used, zero-shot prompting	41.9708
	Gemma2_DASP_1	definitions used, one-shot prompting	42.5695
	Mistral_DASP_1	Mistral 7b with one-shot prompting	42.4718
Post-Competition	Gemma3_DASP_0	Definitions used, zero-shot prompting	41.6236
	Mistral_baseline	No definitions, zero-shot prompting	42.9018
	Mistral_IRP	Mistral 7B with iterative refinement prompt (IRP)	43.1044
	Mistral_FKGL	FKGL-based output selection	**43.5120**

IRP	Improve the simplified version of the scientific text below to make it clearer and easier for a general audience. Your goal is to maximize the SARI score by simplifying language and structure, while keeping all facts, numbers, and findings exactly the same. Do this step by step. **ORIGINAL TEXT:** {Complex text} **FIRST SIMPLIFIED VERSION:** {Generated simple text} **REFINED VERSION:**



Results- Post Competition Phase

Table 2
Overview of best submitted runs and corresponding SARI Scores

Phase	Run	Description	Score
Evaluation	Mistral_DASP_0	Definitions used, zero-shot prompting	**43.5051**
	Med42_DASP_0	definitions used, zero-shot prompting	41.9708
	Gemma2_DASP_1	definitions used, one-shot prompting	42.5695
	Mistral_DASP_1	Mistral 7b with one-shot prompting	42.4718
Post-Competition	Gemma3_DASP_0	Definitions used, zero-shot prompting	41.6236
	Mistral_baseline	No definitions, zero-shot prompting	42.9018
	Mistral_IRP	Mistral 7B with iterative refinement prompt (IRP)	43.1044
	Mistral_FKGL	FKGL-based output selection	**43.5120**

KGR: Keyword-Guided Retrieval Prompt / **DASP_0:** Definition-Augmented Simplification (Zero-Shot) / **DASP_1**: Definition-Augmented Simplification (One-Shot) / **IRP**: Iterative Refinement Prompt/ **FKGL**: Flesch-Kincaid Grade Level

Example of a Scientific Text Simplified by Mistral_DASP_0

Inputs

Complex Text

Birt-Hogg-Dubé syndrome (BHD) is an autosomal dominant genetic disorder characterised by pulmonary cysts, fibrofolliculomas and renal tumours. The pulmonary cysts may lead to pneumothorax, and in cases of primary, spontaneous pneumothorax the syndrome should be excluded. The renal tumours are frequently malignant, but slow-growing. Screening and family assessment enable discovery of renal cancer at an early stage. The syndrome is underdiagnosed and little known.

List of keywords with their definitions

[***{'pneumothorax': 'free air in the chest cavity'}*** ,
{'malignant': 'cancerous'},
{'screening': 'the process of deciding if someone is qualified to take part in a medical research study or not'}]

Output

Simple Text

Birt-Hogg-Dubé Syndrome (BHD) is an inherited genetic disorder that affects some individuals. It's characterized by three main symptoms: lung sacs filled with air (pulmonary cysts), skin growths called fibrofolliculomas, and kidney tumors.

- If the lung sacs rupture, **a condition called pneumothorax may occur, where air accumulates in the chest cavity**. If this happens without any known injury, it's crucial to consider BHD as a potential cause.
- Most of these kidney tumors are cancerous but grow slowly. Regular check-ups and examining family medical history can help detect these cancers early, which is vital for effective treatment.

Unfortunately, this condition is often underdiagnosed due to its relative rarity.

Conclusion

The RAG approach helped to bridge the accessibility gap in scientific communication.
Our approach:

- Ranked 5th in the CLEF 2025 SimpleText Task 1.2, as evaluated by the SARI metric.
- Demonstrated the adding definition in plain english help improve the quality of the simplified texts.

Future directions include:

- Expanding the glossary with context-sensitive definitions
- Exploring semantic retrieval beyond exact matching

Thank you for your attention

 Anya Amel NAIT DJOUDI https://www.linkedin.com/in/anyantd/  https://github.com/Anyantd/

 Sarah NOUALI https://www.linkedin.com/in/sarahnouali/

 Full paper *https://www.dei.unipd.it/~faggioli/temp/clef2025/paper_358.pdf*

CLEF 2025, September 9–12, 2025, Madrid, Spain

Appendix

Prompt Name	Prompt Text
KGR	I have the following document: {Complex text} Please give me the keywords that are present in this document and separate them with commas. Make sure you to only return the keywords and say nothing else. For example, don't say: "Here are the keywords present in the document"
DASP_0	Using these definitions, please simplify the following scientific text for a general audience. Use plain language and explain any complex terms or acronyms. Ensure that all numbers, results, and facts remain exactly the same. Do not paraphrase numerical data or alter the meaning of findings. **DEFINITIONS:** {List of definitions} **TEXT:** {Complex text}
DASP_1	You are a helpful assistant that simplifies biomedical or scientific texts. **Task:** Using these definitions, simplify the following scientific text for a general audience. Use plain language and explain any complex terms or acronyms. Ensure that all numbers, results, and facts remain exactly the same. Do not paraphrase numerical data or alter the meaning of findings. **Example:** (Example of a pair of complex text and its simplified version) *Definitions: {List of definitions}* *Text:{example of complex text}* *Simplified:{example of simplified text}* Now do the same for the following: **Definitions:** {definition} **Text:**{text} **Simplified:**
IRP	Improve the simplified version of the scientific text below to make it clearer and easier for a general audience. Your goal is to maximize the SARI score by simplifying language and structure, while keeping all facts, numbers, and findings exactly the same. Do this step by step. **ORIGINAL TEXT:** {Complex text} **FIRST SIMPLIFIED VERSION:** {Generated simple text} **REFINED VERSION:**

Prompt for keywords extraction

Prompts for text simplification (evaluation phase)

Prompt for text simplification (post-completion phase)