university of groningen / faculty of arts

GroNLP

Reading Time Prediction for Dutch Text Simplification in the PAGINA Project

Sijbren van Vaals, Rik van Noord, Malvina Nissim
University of Groningen
CLIN35

12 September 2025

university of groningen / faculty of arts

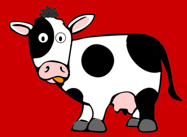
The PAGINA Team

Our beautiful team and partners:

- RuG: University
- DvhN (Mediahuis): Newspaper
- 8D: Research design and gamification
- AI Hub: Development of AI-applications

paginaproject.nl

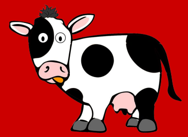
The PAGINA Project



- Accessibility of Dutch news, with a specific focus on low literacy
- Oct 2024 - Sep 2028
- Main goal: Bringing journalism closer to the public

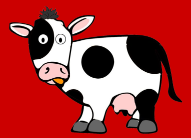
- Text simplification: Difficulty, comprehension, readability
- Perspective and frame: How can we make texts more interesting?

Motivation



Background of the project:

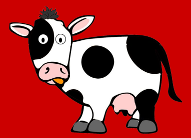
- 2.5 million Dutch citizens struggle with reading, numeracy, and digital devices (Rijksoverheid, 2019)
- Many citizens feel disconnected from news media, especially young people
- This disconnect threatens democratic participation
- Regional journalism is particularly vulnerable



Dataset

Dataset from all Mediahuis Noord titles with:

- News articles: Title and body
- Metadata: Topic, newspaper source
- Engagement metrics: Total nb. of views and total reading time (sec)



Dataset

DAGBLAD VAN HET NOORDEN

Dataset from all Mediahuis Noord titles with:

- News articles: Title and body
- Metadata: Topic, newspaper source
- Engagement metrics: Total nb. of views and total reading time (sec)

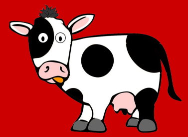
Reading time:

- Captures interest and attention (skimming)
- Approximates complexity and understandability
- Sets the stage for multiple research directions

Possibilities

The dataset offers several possibilities, such as:

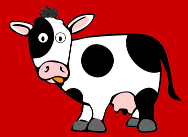
- How do linguistic complexity and length influence reading time?



Possibilities

The dataset offers several possibilities, such as:

- How do linguistic complexity and length influence reading time?
- Are metadata features (sentiment, topic, source) predictive of reading time?

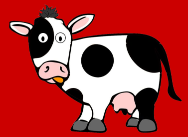


Possibilities

The dataset offers several possibilities, such as:

- How do linguistic complexity and length influence reading time?
- Are metadata features (sentiment, topic, source) predictive of reading time?
- Are readability metrics informative in a real-world context?

Possibilities



The dataset offers several possibilities, such as:

- How do linguistic complexity and length influence reading time?
- Are metadata features (sentiment, topic, source) predictive of reading time?
- Are readability metrics informative in a real-world context?
- What is the extent to which LLMs can effectively predict reading time?

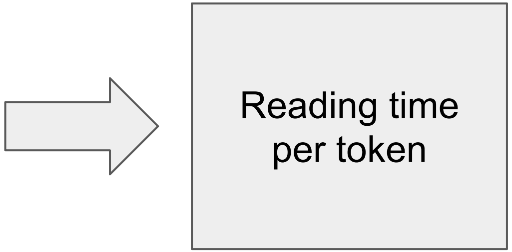
Possibilities

The dataset offers several possibilities, such as:

- How do linguistic complexity and length influence reading time?
- Are metadata features (sentiment, topic, source) predictive of reading time?
- Are readability metrics informative in a real-world context?
- What is the extent to which LLMs can effectively predict reading time?
- Can we develop an effective reading time predictor to approximate complexity?

Experimental Setup

Systematic assessment of blocks of features:



Experimental Setup

Systematic assessment of blocks of features:

Text profiling
Profiling-UD
T-Scan
Lingualyzer

Reading time per token

Experimental Setup

Systematic assessment of blocks of features:

Text profiling	**Read. metrics**
Profiling-UD	Flesch-Douma
T-Scan	Brouwer's Index
Lingualyzer	LiNT

Reading time per token

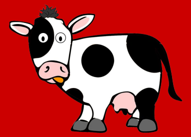
Experimental Setup

Systematic assessment of blocks of features:

Text profiling	**Read. metrics**	**LLM**
Profiling-UD	Flesch-Douma	Next-word prediction (surprisal)
T-Scan	Brouwer's Index	Direct assessment
Lingualyzer	LiNT	

→ Reading time per token

Experimental Setup



Systematic assessment of blocks of features:

Text profiling	**Read. metrics**	**LLM**	**Metadata**
Profiling-UD	Flesch-Douma	Next-word prediction (surprisal)	Topic
T-Scan	Brouwer’s Index	Direct assessment	Source
Lingualyzer	LiNT		Sentiment

→ Reading time per token

Experimental Setup

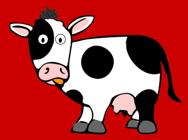
Systematic assessment of blocks of features:

Text profiling	**Read. metrics**	**LLM**	**Metadata**
Profiling-UD	Flesch-Douma	Next-word prediction (surprisal)	Topic
T-Scan	Brouwer's Index	**Direct assessment**	Source
Lingualyzer	LiNT		Sentiment

→ Reading time per token

Can we develop an effective reading time predictor to approximate complexity?

Our Experiment



Reading time prediction:

- Assumption: people read faster through simple(r) texts
- Human-centered evaluation, based on actual human data
- Reading time correlates with comprehension (Levy 2008; Wang et al., 2024) and complexity (Singh et al., 2016; Hollenstein et al., 2022)

Idea:

Useful for **evaluating** simplified texts: lower predicted reading time implies a text is easier to read.

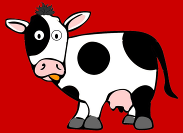
Feature Extraction

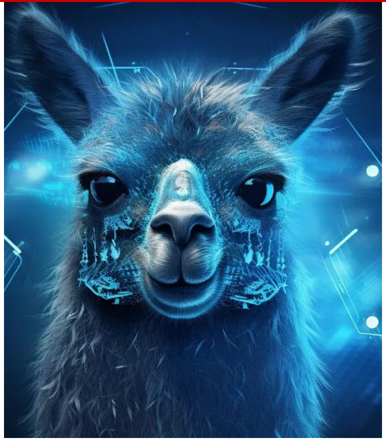
- Get as many features from different linguistic layers as possible
- Profiling-UD pipeline (Brunato et al., 2020)
- Add more uncovered features and readability metrics
- Perform PCA to account for dependent features

Model Selection

- Random Forest (linear regressor)

- Generative AI models:
  - GPT-4o (ChatGPT)
  - Fietje-2-chat
  - Llama-3-8b-instruct

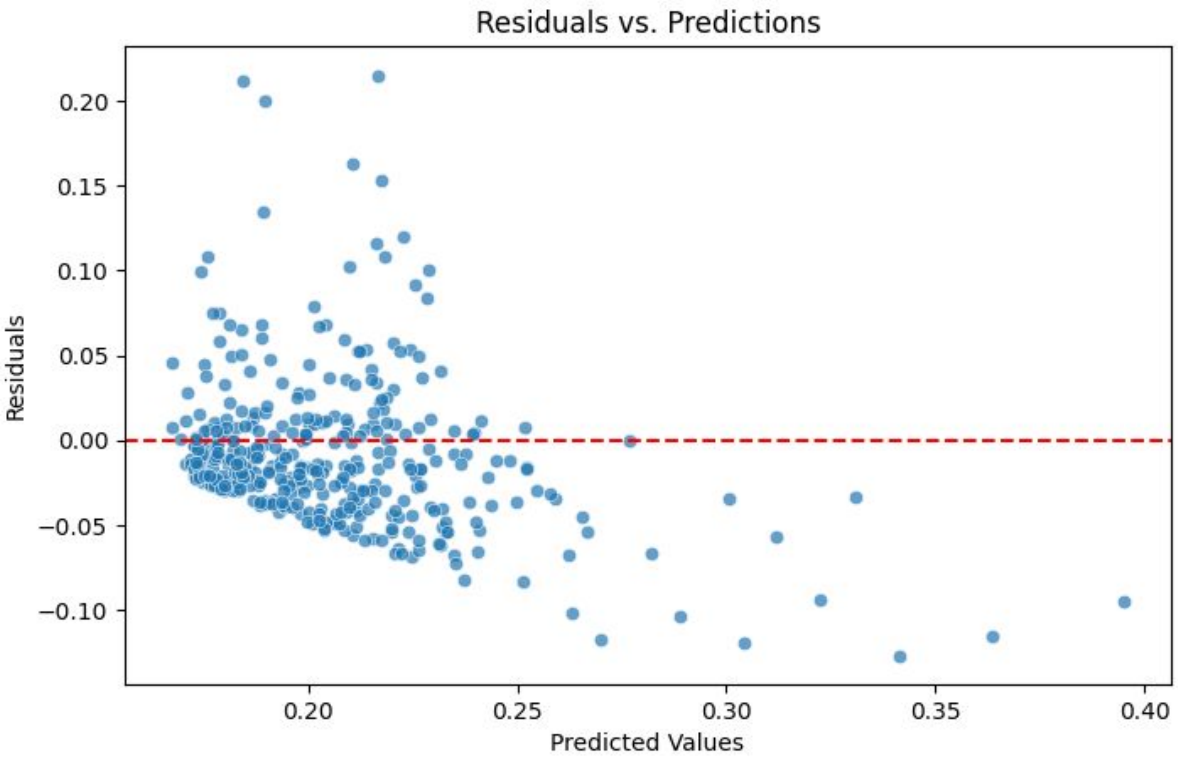




Random Forest

Prediction Performance

Error plot:



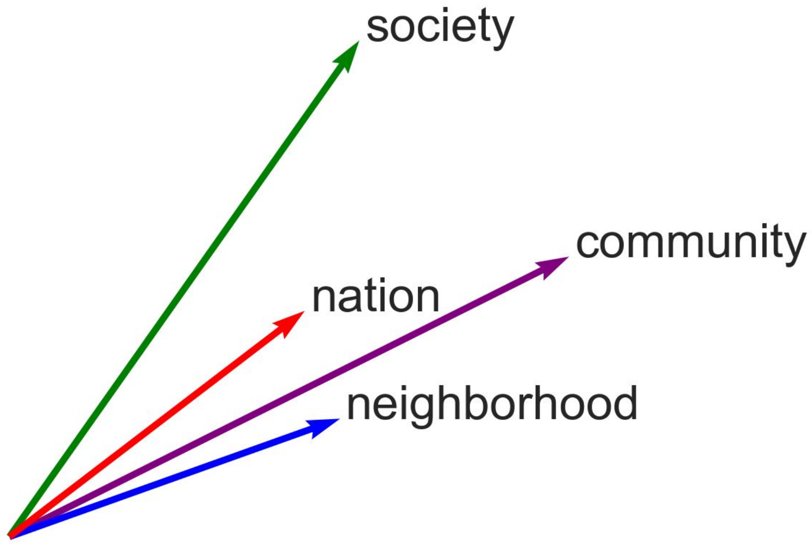
Correlation with gold data:

ρ=0.35; p-value=1.05e-12

Feature Importance

From multiple SHAP plots we observe that good features are:

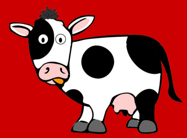
- Noun similarity



Feature Importance

From multiple SHAP plots we observe that good features are:

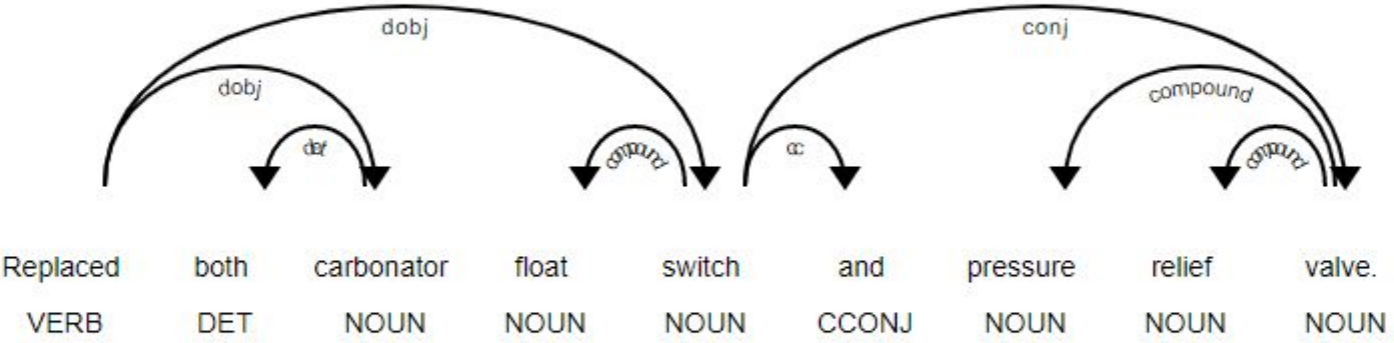
- Noun similarity
- Hapaxes (lexical density)



Feature Importance

From multiple SHAP plots we observe that good features are:

- Noun similarity
- Hapaxes (lexical density)
- Verb edges



Feature Importance

From multiple SHAP plots we observe that good features are:

- Noun similarity
- Hapaxes (lexical density)
- Verb edges
- Distribution of monosyllabic words

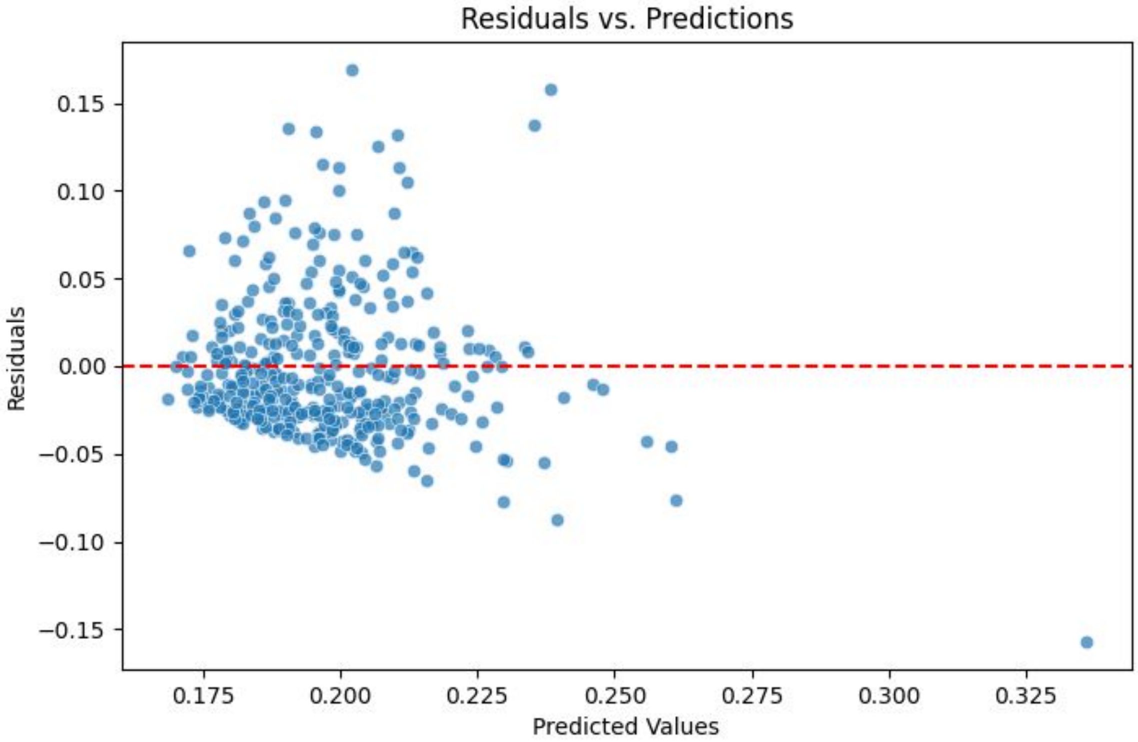
university of groningen / faculty of arts

Baseline Performance

Random forest with n-grams:

Correlation with gold data:

ρ=0.3; p-value=3.3e-09



Baseline Importance

The baseline's most important n-grams were by far:

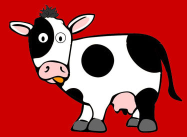
Snein **Sneon**

Baseline Importance

The baseline's most important n-grams were by far:

Snein **Sneon**

The Frisian words for Sunday and Saturday, respectively



Baseline Importance

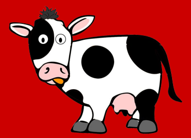
The baseline's most important n-grams were by far:

Snein **Sneon**

The Frisian words for Sunday and Saturday, respectively

What if we make a distinction between weekend and weekdays?

Baseline Importance



The baseline's most important n-grams were by far:

Snein **Sneon**

The Frisian words for Sunday and Saturday, respectively

What if we make a distinction between weekend and weekdays?

- Weekdays: **ρ=0.22**; p-value=9.1e-05
- Weekend: **ρ=0.36**; p-value=0.005

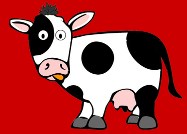
→ New angle of disentangling text complexity from reader interest

Generative AI models

Model Evaluation

GPT-4o and Llama-3-8b-instruct:

- Provide valid explanations
- Look at: structure, tone, information layers, and comprehension
- Corr. with gold reading time: ρ=0.87; p-value=0.001



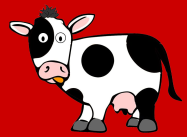
Model Evaluation

GPT-4o and Llama-3-8b-instruct:

- Provide valid explanations
- Look at: structure, tone, information layers, and comprehension
- Corr. with gold reading time: ρ=0.87; p-value=0.001

Fietje-2-chat:

- Confuses input with the provided example
- Can only take two or three examples



What we will do next

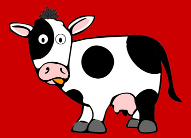
In the upcoming months we will:

- Finalise the systematic assessment
- Disentangle text complexity from reader interest
- Train LLMs for text simplification
- Field test the simplification with a target group

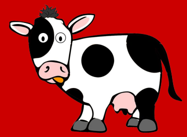
What do we need?

- Parallel data with original and simplified pairs (by humans)
- Human judgements of the simplified text to validate performance

Takeaways

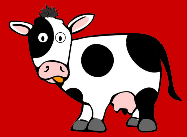


- Focus on readers first
- Good features emerge at different linguistic levels (lexical, semantic, syntax)
- LLMs can look into more subtle features: style, tone, and comprehension
- Weekday news reading is different from weekend news reading



Feel free to ask questions!

Contact Information



Name: Sijbren van Vaals

Email: s.j.van.vaals@rug.nl

LinkedIn: https://linkedin.com/in/sijbren-vv

GitHub: https://github.com/sijbrenvv

Project website: paginaproject.nl