From Legibility to Text Simplification: Measuring and Improving Text Comprehensibility

Irina Temnikova
GATE Institute, Bulgaria
irina.temnikova@gmail.com

1 Brief description of the workshop topic and content

This is a half-day/one-day-long (depending on the number of submissions) interdisciplinary multilingual workshop covering several fields, such as Linguistics, Translation and Translation Technologies, Psycholinguistics, Education, Computational Social Science, and Natural Language Processing. The workshop provides an opportunity for scientific researchers to present their completed or ongoing work, as well as serves as a discussion place for the research, developers and practitioners (such as teachers, representatives of Ministries and other public institutions, writers and publishing houses) communities. The workshop covers all areas and issues at least partially linked to analysing, evaluating and improving text and speech comprehensibility, specifically (among others):

- Analysing and evaluating the readability of texts, including fonts legibility; readability indices

- Readability indices

- Psycholinguistic studies of text and speech comprehensibility

- Eye-tracking and finger-tracking tool approaches to text readability and accessibility

- Linguistic resources and automatic tools for readability, simplification, and comprehensibility

- Assistive reading tools

- Comprehensibility in translation and interpreting; cross-linguistic studies

- Manual rules and automatic text simplification

- Lexical, syntactic, discourse, and document-level text complexity and simplification

- Comprehensibility of texts and conversations in medical, legal, disaster management, and other domains

- Measuring school children's reading skills and related education

- Public institutions' procedures regarding the comprehensibility of public documents

- Multimodal adaptation for different audiences

- Generation of accessible and simplified texts

- Evaluation of the comprehensibility of content generated by Language Models (LLMs) and LLM use in related fields

The workshop will invite:

- Long research articles (maximum 8 pages, excluding references) on completed research

- Short research articles (maximum 4 pages, excluding references) on ongoing research

- Student articles (maximum 4 pages, excluding references)

- Articles from industry (e.g., publishing houses, translation agencies, and IT companies), practitioners (e.g., teachers, translators), and representatives of public institutions (e.g., Ministries of Education), presenting institutional procedures and issues arising in their work (maximum 4 pages)

2 A list of invited speakers, if applicable, with an indication of which ones have already agreed

We are planning to have either one invited talk or a hybrid online/onsite panel of invited speakers from research and industry. The list below contains speakers who have not been contacted yet:

- Ben D. Sawyer – `https://scholar.google.com/citations?user=mynHwzkAAAAJ&hl=en`

- Shaun Wallace – `https://shaunwallace.org/`

- Zoya Bylinskii – `http://zoyathinks.com/`

- Sofie Beier (Legibility research)

- Vera Demberg – `https://scholar.google.com/citations?user=l2CFSAMAAAAJ&hl=de`

- David M. Howcroft – `https://scholar.google.com/citations?user=gfyCkKcAAAAJ&hl=en`

- Bulgarian publishing house representative (Prosveta or another publishing house)

- Vito Pirelli

3 An estimate of the number of attendees

20

4 A discussion of measures planned to ensure the workshop is successful and productive for both in-person and virtual attendees

The measures include a broad coverage of areas, an opportunity for discussions between different specialists, including scientific researchers, practitioners and companies, to allow sharing current issues and finding solutions.

We also plan to collect participants' expectations and feedback for future editions of the workshop.

5 Description of special requirements and technical needs

A hall, electricity plugs, a projector, an Internet connection strong enough to maintain online talks and a hybrid panel discussion.

6 The names, affiliations, and email addresses of the organisers, with a brief statement on their research interests, expertise, and organisational experience

- **Irina Temnikova**, Institute GATE, Bulgaria (`irina.temnikova@gmail.com`)
  10 years co-organiser of the Recent Advances in Natural Language Processing (RANLP) conference (PC communication, START/SoftConf management, reviewer assignment, programme writing, local organisation). Co-organiser of three RANLP Student Workshops; co-organiser of the Post-Truth Workshop (with Preslav Nakov); co-organiser of two editions of the Human-Informed Translation and Interpreting Technology (HiT-IT) workshop; main organiser of the Winter Seminar on Recognising Disinformation and Untrue Information. Research interests: text readability and simplification, MT evaluation, disaster management, emotion and deception detection.

- **Vito Pirelli**, Istituto di Linguistica Computazionale "Antonio Zampolli", Italy
  Research Director at the CNR Institute for Computational Linguistics; co-editor of *The Mental Lexicon* and *Lingue e Linguaggio*; head of the Physiology of Communication Lab; elected member of Academia Europaea (since 2021). Research interests: language architecture and physiology at the intersection of cognitive linguistics, psycholinguistics, neuroscience, and information science. Vito Pirelli has extensive experience with analysing readability using eye-tracking and finger tracking, as well as in organising scientific events.

- **Ruslana Margova**, Institute GATE, Bulgaria
  20 years of experience in journalism and editing; currently working on Information Integrity at GATE. Bridges linguistics, media, and technology, with focus on media and digital literacy. She was assisting with the organisation of the Winter Seminar TRACES, and the Information Integrity Forums 2025 and 2026 (https://sofia-info-integrity.eu/).

- **Nevena Grigorova**, Institute GATE, Bulgaria
  BA in Philology; MA in Computational Linguistics and Advertising. Experience as editor

of educational literature, teacher of Bulgarian language and literature, and organiser of cultural and social events. Researches linguistic markers of propaganda in political and advertising language.

- **Andrea Nadalini**, CNR Institute for Computational Linguistics, Italy
Psycholinguist and research fellow. Research interests: processing and representation of written words in the lexical-semantic system; combines experimental techniques (semantic categorisation, lexical decision times, ERPs) with computational models. Involved in the Readlet project on children's reading behaviour.

- **Dimitar Kazakov**, University of York, United Kingdom
Extensive experience in organising research events (AISB 2011; AAMAS 2002, 2003; CA2HL summer school; Evolang 2026). Research interests include text readability and simplification.

- **Maria Todorova**, Bulgarian Academy of Sciences, Bulgaria
Extensive experience in co-organising CLIB events.

7 A list of Program Committee members, with an indication of which members have already agreed

We are starting to invite PC members, and this is our preliminary list:

- Vito Pirelli (Institute for Computational Linguistics "Antonio Zampolli", Italy) [not yet contacted]

- Matthew Shardlow (Manchester Metropolitan University, United Kingdom) [not yet contacted]

- Dimitar Kazakov (University of York, United Kingdom) [not yet contacted]

- Victoria Yaneva (National Board of Medical Examiners, United States) [not yet contacted]

- Horacio Saggion (Universitat Pompeu Fabra, Spain) [not yet contacted]

- Thomas François (Université catholique de Louvain, Belgium) [not yet contacted]

- Rémi Cardon (Université catholique de Louvain, Belgium) [not yet contacted]

- Elena Volodina (University of Gothenburg, Sweden) [not yet contacted]

- Marcello Ferro (Institute for Computational Linguistics "Antonio Zampolli", Italy) [not yet contacted]

- Cengiz Acarturk (Jagiellonian University, Poland) [not yet contacted]

- Burcu Can (University of Stirling, United Kingdom) [not yet contacted]

- Le An Ha (Ho Chi Minh City University of Foreign Languages and Information Technology, Vietnam) [not yet contacted]

- Ivelina Stoyanova (Bulgarian Academy of Sciences, Bulgaria) [agreed]

8 Answers to the Questions of the Reviewer

We will create our own website for the workshop and will share its link with the CLIB organisers as soon as it is created (in the next 2 weeks).

We will rely on the CLIB organisers to assist us with setting an EasyChair account, as our knowledge of the system is outdated. We will run our own Call for Papers campaign, reviewing, and the acceptance/rejection of the workshop papers.

We will need CLIB's organisers assistance in assembling the workshop's proceedings, especially, as they will be composed of a mixture of LaTeX and Microsoft Word/OpenOffice-created documents.