



Designing LEMI: the Romanian language Tool that Makes Kids Love Reading

Mădălina Chitez, Roxana Rogobete, Alexandru Oravițan

West University of Timisoara, Romania

Abstract

In Romania, the attraction of young schoolchildren to literature is doubtlessly deficient in many ways: the temptation of the digital environment as a recreational refuge and the lack of reading practice, primarily due to inappropriate selection criteria of materials contribute to increased rates of functional illiteracy and school dropout, which automatically translate into poor results recorded in the PISA tests. Despite the wide range of school textbooks and compulsory supplementary reading activities, no unified school reading recommendations based on linguistic research have measured the complexity of the recommended texts specific to each age or competence level category. A team of linguists from Centre for Corpus Related Digital Approaches to Humanities at the West University of Timișoara, Romania, has recently (January 2023) initiated a research project in which such problems are addressed and integrated into a literacy support tool: LEMI. This tool includes several functionalities: a children's literature repository searchable according to selected criteria, the automatic assessment of short reading text complexity and a children's reading recommender system. This paper presents the tool's design, showcases its basic functionalities, and explains its expected impact, particularly by highlighting the didactic applicability of this new Romanian language instrument.

Keywords: LEMI, children's literature repository, text complexity automatic assessment for Romanian, literacy support tool

1. Introduction

The International Association for the Evaluation of Educational Achievement (IEA) defines reading ability as "the ability to understand and use those forms of written language required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts" [1]. In this context, reading is an essential competence that impacts children's lives in many aspects. Interest in reading activities is genuine and visible: school textbooks and auxiliary materials include a wealth of texts to be read and analysed; schools and non-profit organizations are constantly undertaking projects to promote reading, particularly in vulnerable groups; publishers are prolific in offering the most attractive reading anthologies possible. This interest is motivated by school requirements and by recognizing the significant impact of reading on school results and especially on the child's holistic development.

In Romania, despite numerous reading resources and reading support projects, the attraction of young schoolchildren to literature is deficient in many ways. One of the reasons seems to be the temptation of the digital environment as a recreational refuge. The second reason is the lack of reading practice, primarily due to inappropriate selection criteria of materials and their adaptation to the specifics of the educable. This results in increased rates of functional illiteracy and school dropout, which automatically translate into sub-mediocre results recorded in the PISA tests [2]. The texts included in Romanian textbooks and supplementary reading volumes have not been linguistically assessed. Because of this, no measures of linguistic complexity have been applied before recommending texts for a particular grade, age or reading competence level.

In this context, one effective method of reversing this trend is to choose texts appropriate to the child's level of comprehension, which is already happening in countries with good reading test scores (such as the multiple series of graded readers available in the UK). Thus, through the recurrent use of easy-to-use tools to respond to the needs of the target individual or group, either as an integrated part of preparing didactic activities associated with reading in primary school or as a widely available children's app, it is expected that the obstacles in reading comprehension will diminish, and the degree of reading motivation will increase. A team of linguists from the Centre for Corpus Related Digital Approaches to Humanities ([CODHUS](#)) at the West University of Timișoara, Romania, has



recently (January 2023) initiated a research project in which such problems are addressed and integrated into a literacy support tool – LEMI – that can be easily used by teachers, parents, teaching materials creators, educational NGOs, and producers of children’s literature to cultivate a passion for reading and thus promote self-confidence and autonomy.

2. Children’s literature reading apps

To our knowledge, only one recently launched (October 2022) application targets book recommendations for children in Romania: NABU. Developed by NABU, UiPath Foundation and OvidiuRO, the app has been marketed as “the reading app for children”. It includes 100 bilingual books (Romanian and English), which can be downloaded and read on mobile devices. It is addressed to 2 to 10-year-olds. NABU states that books are organized “by reading fluency levels,” but, to our knowledge and judging by the public descriptions of the app, no linguistic research supports text classification. Individual teachers have simplified texts based purely on observation, without quantitative text complexity evaluation or readability assessment.

Linguistic research-based apps that can be applied to children’s literature are rare; most of them have been designed for English and focus on linguistic complexity assessment. Two valuable examples, which can serve as models for linguistics-supported educational apps like LEMI are ARTE (Automatic Readability Tool for English), which “offers batch processing of texts through readability formulas” while its particularity is “to produce specific readability scores as outcome instead of particular individual linguistic features” [3], and Text Inspector [4], advertised as a tool which can be used to “analyse the difficulty level of English texts”, which includes several text assessment feature sets: Statistics and Readability, Lexical Diversity, Tagger, Lexis (EVP, KVL, NBC, COCA, AWL), PHRASE List, Metadiscourse Markers and a Scorecard.

3. LEMI – the first Romanian tool that uses computational linguistics methods to assess school children's literature complexity and readability

3.1 Purpose and design

LEMI aims to stimulate children’s interest in individual and collective reading when participating in their first reading activities (ages 7-11). The tool uses linguistic research methods to automatically evaluate and recommend texts suitable for reading at various stages of the primary school cycle (grade or performance level in reading texts). If the text is not adapted to the child’s level of comprehension, the learner will tend to associate reading with insurmountable challenges or cognitive stress. If, on the other hand, the contact with the text is adapted to the child’s level of knowledge through linguistic selection, reading becomes a motivating and recurrent act. In this way, by reading more, the child participates with pleasure in the act of knowledge.

The design of the LEMI tool includes the following main functionalities (F):

F1: a digital repository of children’s literature comprising 250 school readings (texts representative of Romanian and universal cultural heritage).

F2: an interface of the automatic assessment of language complexity and reading comprehension of texts for young school-age readers (ages 7-11).

F3: a book recommender system based on user profile (complexity level, topic preferences).

The testing and validation of the LEMI features will be conducted with the involvement of researchers, teachers, children, and other categories of users. The impact (see section 3.4) of LEMI will be amplified through the involvement of the local community, representatives of organizations and those with decision-making functions in the educational and cultural environment.

3.2 Preliminary steps

Until now, the creation of the LEMI children’s literature assessment app has gone through several preparation and development stages:



(1) School reading survey

A first questionnaire of needs and observations on school reading has been administered. This questionnaire has included response opportunities for all groups that have given relevant opinions on schoolchildren's literature texts: teachers, parents, and linguistic experts.

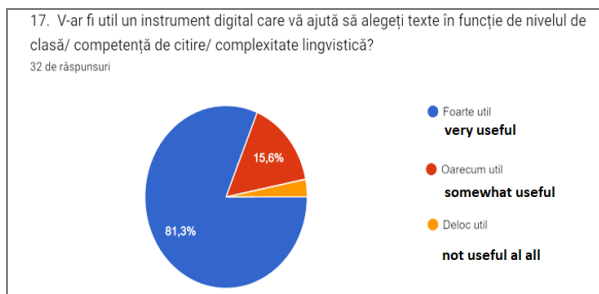


Fig. 1. How teachers answered the question: "How useful is a digital instrument which helps you select texts adapted to grade/reading competence/linguistic complexity level?"

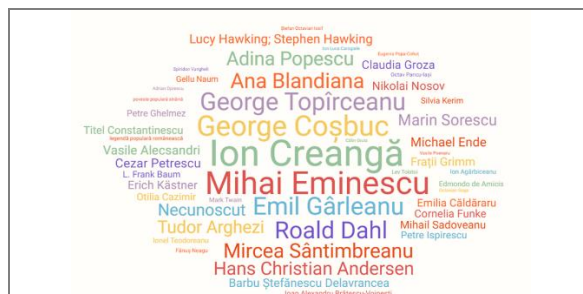


Fig. 2. Children's literature authors encountered in the school textbooks approved by the Romanian Ministry of Education: frequency-based cloud

(2) LEMI team training

This activity aimed at clarifying the objectives related to the methods of technological transfer from the scientific field into the educational-cultural field. The team received training and instructions for developing collaboration protocols with schools for collecting and validating data that will contribute to digital tools to facilitate school reading, developing the data collection protocol for the digital repository, and making copyright or digital text consent forms.

(3) LEMI visual identity

Before creating and launching the LEMI app, we created two elements that would help shape the visual identity and community visibility of the final product: the LEMI logo and the LEMI project webpage. The logo has already been added to the pages describing the project ([project webpage](#), social media pages) and will soon be integrated into marketing products.



Fig. 3. LEMI logo

(4) Digital reading repository

The development of the digital repository involves activities in which the expert linguists have extensive expertise:

- Collection of all 250 texts in primary formats.
- Processing the texts into digital corpus format.
- Management of the database that will form the basis of the corpus.
- Testing the use of the corpus using digital analysis tools.
- Testing the distribution of texts according to complexity level by using linguistic methods.



- Validation of the level of complexity through an in-class questionnaire; this questionnaire will correlate the level of the texts' perceived complexity with the one suggested by the linguistic analysis.
- Discussions with teachers to establish the linguistic parameters that would correspond to a given reading level (e.g., short sentences correlate with level 1)
- Elimination of texts that do not correspond to a target level (grades 0-4)
- Supplementing the repository with the remaining texts and validating them.

The repository is planned to contain 250 short texts, extracts (maximum three pages each) from canonical or modern texts in Romanian literature and universal literature written by authors such as Ion Creangă, George Coșbuc, Ioan Slavici, Ion Luca Caragiale, Emil Gârleanu, Otilia Cazimir, Alexandru Mitru, Cezar Petrescu, Alex Donovici, Mark Twain, Jules Verne, Roald Dahl, Franz Hodjak, Bartis Ferenc etc.

The data collection sessions will be organized either at the schools' premises (for texts delivered by teachers) or through specific working sessions: digitization of textbook extracts, online content processing, teacher-driven validation. Copyright specifications forms will accompany the collected texts.

3.4 Expected impact

Educational impact: Kids read more!

Since LEMI's functionalities involve digitization at multiple levels (as a digital repository of texts, NLP analysis of literary text, and book recommender), LEMI is widely and effectively accessible, thus ensuring a high educational impact.

Social impact: rapid implementation in disadvantaged areas

By creating the LEMI digital repository of readings and didactic recommendations for using texts suitable to children's cognitive level, the premises are created for the sustainable promotion of reading as part of the written culture. In this way, children and teachers from all backgrounds, especially those from disadvantaged areas, will have free resources, easy to use and organized by levels of reading complexity.

Cultural impact: age-adapted "traditional" literature is more engaging

The texts in the digital repository will have a heterogeneous profile: traditional Romanian literature, modern Romanian literature, minority Romanian literature, and universal literature. All genres and styles will be represented in the repository.

4. Discussion and conclusion

During these initial stages, we have encountered multiple challenges to be addressed before taking the next linguistic analysis steps. First, few solutions exist for the computational analysis of text complexity in Romanian. For example, a simple measure such as syllable segmentation and counting requires testing and validation of programs which are not freely available. However, collaborations have been initiated with other Romanian researcher teams who can assist us in implementing their automatic text-processing approaches for LEMI.

Second, there have been numerous dilemmas concerning the texts that should be included in the digital repository. The decision was made to aggregate information extracted from the school reading survey, partner schoolteachers and the research team's expertise to create a collection of texts with both reading motivation and cultural immersion potential.

Finally, it has been demonstrated that such complex projects can be effectively conducted when all actors involved in the creation of the tool collaborate. In the case of LEMI, linguists must join forces with IT specialists while also receiving ongoing support from all school partners and potential



stakeholders (e.g., school literature publishing houses). For this reason, the creation of a functional network that will address different challenges and contribute to the final design, development, testing and distribution of the tool is required. We expect to launch the first version of the LEMI app in November 2023.

Acknowledgement



<http://lemi.projects.uvt.ro/>

This work was supported by a grant of the Administration of the National Cultural Fund (AFCN) of the Romanian Ministry of Culture, in the framework of the programme *Promotion of Written Culture*, session I/2023, for the project LEMI (*Lectură pentru mine. Știința în slujba copiilor – LEMI I* Reading for Me. Science for Children; January – November 2023), contract no. P0299/10.02.2023. The project was awarded to the West University of Timisoara for the proposal submitted by the project coordinator, CS II Dr. Habil. Madalina Chitez.

References

- [1] Mullis, I.V.S., Kennedy, A.M., Martin, M.O., and Sainsbury, M. (2006), PIRLS 2006: Assessment Framework and Specifications (2nd ed.), Chestnut Hill, MA: TIMSS and PIRLS International Study Center, Lynch School of Education, Boston College, <https://timss.bc.edu/PDF/P06Framework.pdf>.
- [2] OECD (2019), PISA 2018 Results (Volume I): What Students Know and Can Do, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/5f07c754-en>.
- [3] Choi, J.S. and Crossley, S. (2021). *Readability Assessment Tool for English Texts* (No. 6190). EasyChair.
- [4] Bax, S. (2012). Text inspector. *Online text analysis tool*. Available at: <https://textinspector.com/>.