

Towards Building the LEM Readability Platform for Children's Literature in the Romanian Language



Madalina Chitez¹, Mihai Dascalu², Aura Cristina Udrea², Cosmin Strilețchi³, Karla Csürös¹, Roxana Rogobete¹, Alexandru Oravițan¹

¹West University of Timisoara, Romania

²National University of Science and Technology POLITEHNICA Bucharest, Romania

³Technical University of Cluj-Napoca, Romania



Introducing...



an open-source readability platform and a dynamic digital repository of

children's literature texts in Romanian.

<u>Lectură</u> = Reading pentru <u>mine</u> = for me

http://www.lemi.ro/



Why LEMI?

- ★ Readability: Key characteristic of texts as linguistic constructs to be processed by the human mind, referring to the set of features that influence the reader's understanding and reading efficacy (Collins-Thompson, 2014)
- ★ School text readability: Focus of numerous studies (e.g., Shanahan et al., 2016) with impact on educational interventions and policies
- ★ Studies on readability formulas for Romanian: Relatively scarce (e.g. Garais, 2011; Garais & Enaceanu, 2011); text complexity features (e.g. Toma et al., 2021)
- ★ As of today, there is no platform, tool, or app that integrates a research-based readability formula for the Romanian language

Related work



Readability: Beyond Education

- ★ Readability Evolution: Interest spans over a century, stemming from educational needs during the Great Depression.
- ★ Early Discoveries: Leavy & Grey's 1935 study identified style, especially sentence length and word complexity, as crucial readability factors.
- ★ Formula Development: Emerged in the 1950s and remain stable, validated by over 200 formulas and numerous studies by the 1980s.
- ★ Core Metrics: Word and sentence length, lexical choices.
- ★ Diverse Applications: Education, healthcare, law, business, public administration, research, enhancing accessibility for specific audiences.



Readability Formulas

- ★ Formula Evolution: primarily developed by U.S. researchers, analyze English text characteristics.
 - 1940s: Dale-Chall, Flesch Reading Ease
 - 1950s: Gunning Fog Index
 - 1970s: Fry Readability Graph, Simple Measure of Gobbledygook, Flesch-Kincaid Grade Level Test
 - Other indices: Coleman-Liau, Automated Readability Index
- ★ International Development: in the 1960s, Sweden's Läsbarhetsindex (LIX),
- ★ Global Formulas: French's Kandel and Moles Index, Italian's Gulpease Index, and Spanish's INFLESZ scale.
- ★ Romanian Readability: L1 and L2 formulas, automated writing assessment tools (ReaderBench)



Tool Integration

- ★ Current State: No platform, tool, or app currently integrates a research-based readability formula for the Romanian language. This also applies to digital tools that provide access to children's literature texts based on readability levels or offer automatic readability assessment.
- ★ English-Language Tools: Text Inspector (U.K., Bax, 2012) and ARTE (U.S., Choi & Crossley, 2021), not applicable to Romanian.
- ★ Unique Features: Our tool provides access to a self-compiled digital repository of children's texts, with filtering options (age, grade, and readability level)

Method



Building the LEMI Romanian children's literature corpus

- ★ Analysis of Romanian language and literature textbooks (grades I-IV)
 - > Database of all reading texts that appear in primary textbooks
- ★ Surveys to three target groups (grades I-IV): teachers, parents, children
 - 1. Evaluate their **satisfaction** with their Romanian **textbooks**;
 - 2. Give examples of texts that children at their level enjoy;
 - 3. Give level-appropriate recommendations of authors and/or texts.



The **LEMI** Romanian children's literature corpus

\star Reading texts digital repository:

- Original or adapted to level;
- Primarily fiction or poetry;
- Romanian or foreign writers;
- Classic or modern (publication date);
- Assigned age group and level.

★ This study: 80 texts (33,154 words)

- 45 classic Romanian texts;
- 9 modern Romanian texts;
- > 9 classic foreign texts;
- 17 modern foreign texts.

www.lemi.ro | clasa 0 | nivelul 3

Țup. Imposibil e doar un cuvânt (Fragment adaptat; text pentru audiere prin citire de către adult)

Alex Donovici

A fost odată ca niciodată o pădure mare și frumoasă, din care oamenii încă nu tăiaseră copacii și unde animalele și păsările trăiau fericite. La marginea unui luminiş își avea cuibul o ciocârile. Da, o ciocârile, acea pasăre micuță care cântă atăt de frumos cum puține păsări o pot face. În fiecare dimineață, ciocârila zbura sus, sus de tot, cât mai aproape de soare, și cânta. Soarele era fericit să o audă de fiecare dată când răsărea și, drept mulţumire, lumina cu bunătate pădurea și îi învăluia vitețuitoarele cu razele lui calde.

Ciocàrlia cànta cu atât mai frumos cu cât era mai ferricită. În cuib avea șase ouă de care avea grijă ca de ochi din cap și abia aștepta ca din ele să iasă puii. Minunea s-a întâmplat într-o dimineață, chiar după ce pasărea îi cântase, ca de obicei, soarelui. Puii au început să iasă unul câte unul din ouă, să ciripească vioi și să întindă cioculețele flămânde după mâncare. Dar nu toți. Doar cinci. Cel de-al șaselea ou se mișca prin cuib, se rostogolea, ca și cum puiul din el încerca din răsputeri să-i spargă coaja, dar nu aveă forța s-o facă. Ciocârlia îi privea îngrijorată, cu sufletul la gură. Mă iertați, la

www.lemi.ro | clasa 0 | nivelul 3

1

cioc. S-a hotărât să ajute puiul să iasă la lumină și a lovit cu grijă coaja oului, ca s-o spargă. O bucată de coajă s-a desprins și ciocărila a văzut fericită cum ultimul său puiuț, o fetiță-ciocârlie, scoate căpșorul la lumină. Dar tot nu putea să iasă, așa că mama ei a luat una câte una bucățile de coajă din jurul ei, iar puiuța s-a eliberat în sfârșit. În acel moment, bucuria s-a transformat în durere. Deși ciripea veselă, micuța ciocârile nu vea aripioare, iar piciorușele îi erau scurte, mult prea scurte. Frățiorii ei săreau veseli prin cuib, dar ea abia se mișca.

Ca o mamă bună, ciocârlia a încercat să aibă grijă de toți puii. Chiar dacă cei cinci cu aripioare fugeau care încotro și cea mică stătea mai mult în cuib. În fiecare dimineață, mama le aducea semințe și își hrănea toți puii, rând pe rând, ca să nu rămână flămând niciunul dintre ei. Seara le cânta tuturor înainte de culcare.

Micuța era prima care îi ieșea mamei în întâmpinare de fiecare dată când se întorcea la cuib. Țopăia veselă pe piciorușele ei scurte, așa că ciocârlia i-a dat numele Țup.



★ Visitors:

• Access LEMI information.



Despre LEMI

De ce nu citesc cu plăcere copiii din ziua de azi?

Unul din motive este faptui că textele pe care le citesc copiii, în primii ani de contact cu activități de lecturare, sunt prea complexe din punct de vedere lingvistic.

Există studii care recomandă "sporirea motivației elevilor pentru lectură prin creșterea încrederii în el însți ca citorif îvezi Reading literacy în irelardi în PISA 2018: Performance, policy and practice, pagina 150; document complet adu). In frincederea crește atunci când texte de au un grad de complexitate și lizibilitate (readability în literatura de specialitate) că mai adaptat vărstei sa invelului çolar.

LEMI este primul instrument digital din Romània care integrează metode de analiză a complexității lingvistice a textelor pentru lectură școlară. Pentru crearea instrumentului digital LEMI, <u>echipa</u>



★ Visitors:

- Access LEMI information.
- Register for free.



				Despre	Documentație	Ştiri	Echipa	Conta
+ Creare c	ont							
* Prenume:								
* Nume:								
* E-mail:								
* Ești un:	nespecif	icat 👻						
Organizația din care faci								
parte (dacă e cazul):								
* Nume de utilizator:								
• Parolă:								
(intre 8 și 15 caractere)								
* Confirmă parola:								
* Sunt de acord cu	0							
Termenii & condițiile:	0							
* Textul din imagine:								
	-							
		✓ Salveaz	ă					



Authenticated users:

 Browse & filter text collection. Results displayed with metadata + excerpt. Texts readable online or downloadable in PDF format.



www.lemi.ro | clasa 2 | nivelul 2

Marco Polo (Fragment adaptat din "Micii mei eroi")

Javier Alonso Lopez

Nu-i așa că e amuzant să explorezi? Să vizitezi castele imense, să vezi amile săbatice și să întânești persoane cu tradiții total diferite de ale tale. Și e aproape la fel de amuzant să le povestești prieternilor despre tot ce ai întâlnit pe parcursul aventurilor tale și să-i vezi cum fac ochii mari când descrii minunățile pe care le-ai văzut.

Sunt Marco Polo. În vremurile când m-am născut eu, majoritatea oamenilor nu-și părăseau niciodată satul sau orașul. Era foarte complicat și periculos să călătorești.

Mie însă nu-mi era frică, eram curios! Am avut ocazia să călătoresc în locuri îndepărtate și n-am lăsat-o să-mi scape.

Datorită poveștilor despre călătoriile mele, și alți oameni au putut afla despre minunățiile care se găsesc pe pământuri îndepărtate. Secole la rând, poveștile mele i-au fascinat pe toți, stârnind și altora aceeași sete de aventură! Vrei să știi cum au fost călătoriile mele?

lată povestea mea. Mă numesc Marco Polo și m-am născut la

Veneția, cu aproape 800 de ani în urmă, în 1254.

www.lemi.ro | clasa 2 | nivetul 2

1

Cum sigur știi, Veneției e un oraș ceva mai deosebit: în loc de străzi avem canalel Așa că pentru noi, venețienii, e ceva firesc să ne deplasăm pe apă. Să nu uităm că am fost mari negustori: în fiecare zi plecau din oraș și soseau sute de corăbii cu produse din toate coliturile lumii.

Să te deplasezi pe uscat era mult mai complicat. Dacă nu voiai să mergi pe jos, alternativa era să folosești animale: cai, măgari sau câmile. Asta îngreuna călătoriile și le făcea să dureze foarte mult timp. Așa păteau tată și unchul meu, care făceau comerț cu Asia. Fiecare câlătorie de-a lor dura ani de zile. Da, ai citti brie: ani de zile!

Când am venit pe lume, tatăl meu era departe de Veneția, într-una din lungile lui călătorii. Se dusese în China pe Drumul Mătăsii. Eu îmi

Se dusese in China pe Drumul Matasii. Eu imi petreceam zilele imaginăndu-mi ce făcea el în celălalt capăt de lume. Chiar dacă azi știu că poveștile istorisite de negustorii mai bătrâni nu erau intru totul adevărate, ele aveau darul de a-mi stârni imaginația.

 De ce se numeşte Drumul Mătăsii?
 Pentru că pe acest traseu transportăm mătasea din Asia.

 E adevărat că în Asia există dragoni mari care scuipă foc?

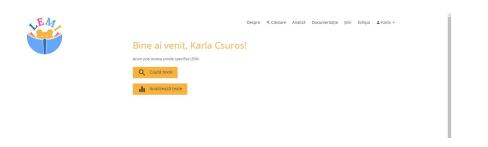
 — E adevărat! Sunt şi războinici cu două capete care îi atacă pe negustori!
 Abia asteptam să mă fac mare, ca să pot

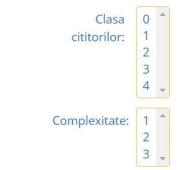
vedea cu ochii mei toate aceste minunății!



Authenticated users:

- Browse & filter text collection. Results displayed with metadata + excerpt. Texts readable online or downloadable in PDF format.
- Analyze complexity of own texts. Results: recommended age and school grade, linguistic complexity levels, overall text intricacy.







★ Administrators:

- Secured CMS to manage database with CRUD facilities.
- View registered users and their uploaded texts, if consent is provided.

NEM -	ŀ	ADMINISTRARE	
exte & metadate	Texte & meta	odate ⊙ Adaugă texte €Încărcare metadate ⊙ Vizualizare metadate	
3 leșire		scărcare texte	
	Identificatori Identificatori RD1,R0,C,001 ThrL: La scladat (Pagment adaptat din "Amintiri din cogliane") Agentification (Pagment Accord); en claime Teme: coplane; cel asic Najonalitatea autorului: autor romain Verhime: clasic Domeniu: alte Variat a citorolior: 10, 11 Clasa: citorolior: 4	Continut filtrat Intro zi, aproape de Sânt-lile, mama avea multe trebi de fâcut. Parea obosită și grâbită, Se trezise deverme și mă trezise și pe mine. — Nică, mis a spus cu băndețe, azi avem mult de lucru. Tatăl tâu a pletar să se ocupe de o treabă la câmp, iar eu am adr. de multe trebi că aba ma îm pasul cu ele. Te rog, stal langă mămuca și ai grîță de sont-cu. Și dacă o să mă ajuți. Când merg la Fâti.	Acţiune conţinut filtrat conţinut original şterge

LEMI and Digital Humanities

- ★ Access to first digital repository of children's literature texts in Romanian;
- ★ Literary heritage enhanced by inclusion of most popular pieces of children's literature and literary samples of texts not available to the general public in user-friendly format;
- ★ Computational interface reflects latest developments in Digital Humanities research, where linguistics and computational linguistics intersect (Luhmann & Burghard, 2022)

Results and Discussion



Corpus Analysis with ReaderBench

★ Textual complexity indices: surface-based, syntax, semantics and discourse

Feature	Description	Importanc e
Max(Dep_xcomp / Par)	Maximum number of 'xcomp' (open clausal complements) dependencies per paragraph	0.266
SD(ParseDepth / Sent)	The standard deviation of the depth of the parsing tree per sentence	0.233
Max(NgramEntr_2 / Word)	Maximum entropy of bigrams per word	0.122
M(WdEntr / Par)	Mean of word entropy per paragraph	0.067
Max(UnqPOS_noun / Sent)	Maximum unique nouns per sentence rBench complexity	0.019

Table 2: ReaderBench complexity indices representative for the corpus

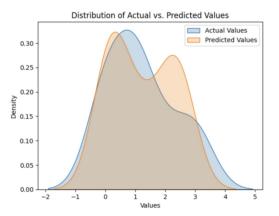


Figure 2: Random forest regression model

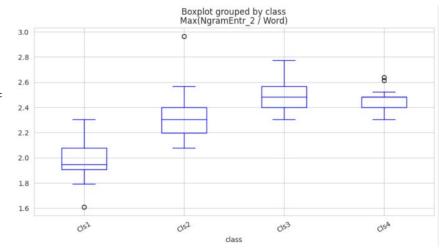
Mean Squared Error = 0.6R² = 0.4





Calibration of a readability formula for children's literature texts in Romanian

- ★ Initial Classification: reading complexity system per grade
- ★ Alignment with ReaderBench Indices: initial assessment and text classification per grade align with the results of ReaderBench textual indices
- ★ Textual Complexity Indices: ReaderBench's maximum number of bigram entropy per word effectively discriminated between grades, higher values => more complex texts
- ★ Validation Process: 'user calibration' conducted in schools for grades 3 and 4 resulting in a 96.48% acceptance rate of text-per-level among 256 participants
- ★ Next Step: validation process replicated for a larger set of texts across all school grades (0-4)





To sum up...

- ★ Creating educational language apps: crucial and complex process that requires linguistic data collection, appropriate analysis, didactic validation.
- \star Detailed major steps in the creation of:
 - **Repository of children's texts** distributed on readability levels (grade 1 to 4);
 - **Readability platform** for children's literature in Romanian, launch Nov 2023.
- ★ Corpus analysis with ReaderBench text complexity framework: satisfactory match between corpus texts and relevant metrics (e.g., bigrams).
 + Classroom validation = Readability formula calibration before automatic text evaluation of user texts.
 - \rightarrow Impact on education, corpus linguistics, Digital Humanities in Romania.



Notes on ethical issues

- ★ Research based on the voluntary participation of research subjects schoolchildren, their parents/legal tutors and their instructors.
- ★ All data available exclusively to research project members and were processed for statistical and scientific research purposes only.
- ★ All participants were informed about the key elements of the research study and signed informed-consent forms.
- ★ Protection of all personal data was assured following GDPR regulations.
- ★ All children's literature texts comply with the Romanian and international legislation in point of copyright.
- ★ Short fragments (2-3 pages) of larger literary work, which complies with Article 35, paragraph (d) from the Copyright Law No.8 (14 March 1996).

References



- Barrio-Cantalejo IM, Simón-Lorda P, Melguizo M, et al. [Validation of the INFLESZ scale to evaluate readability of texts aimed at the patient]. Anales del Sistema Sanitario de Navarra. 2008 May-Aug; 31(2): 135-152. DOI: 10.4321/s1137-66272008000300004. PMID: 18953362.
- Bax, S. (2012). Text inspector. Online text analysis tool. Available at: https://textinspector.com/. Begeny, J. C., & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials?. Psychology in the Schools, 51(2): 198-215.

- Begeny, J. C., & Greene, D. J. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials?. *Psychology in the Schools*, *51*(2): 198-215.
 Björnsson, C. H. (1968). *Läsbarhet*. Liber, Stockholm.
 Björnsson, C. H. (1983). Readability of newspapers in 11 languages. *Reading Research Quarterly*: 480-497.
 Chicco, D., Warrens, M.J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, *7*: e623. doi: 10.7717/peerj-cs.623. PMID: 34307865; PMCID: PMC8279135.
 Choi, J.S. and Crossley, S. (2021). *Readability Assessment Tool for English Texts* (No. 6190). EasyChair.
 Coleman, M. and Liau, T. L. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, *60*(2): 283–284. https://doi.org/10.1037/h0076540.
 Collins-Thompson, K. (2014). Computational assessment Tool fer *English Texts* (No. 6190). EasyChair.
 Dale, E. and Chall, J.S. (1948) A Formula for Predicting Readability. *Educational Research Bulletin 27*(1): 11-28.
 Dascalu, M., Dessus, P., Trausan-Matu, S., Bianco, M., and Nardy, A. (2013). Readerbench, an environment for analyzing text complexity and reading strategies. In *Artificial Intelligence in Education:* 16th International Conference, *AIED* 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16, Springer, pp. 379-388.
 Dascalu, M., Gutu, G., Ruseti, S., Paraschiv, I. C., Dessus, P., McNamara, D. S., Crossley, S.A., & Trausan-Matu, S. (2017). *ReaderBench:* A Multi-lingual Framework for Analyzing Text Complexity. In E. Lavoué., H. Drachsler, K. Verbert, J. Broisin, & M. Pérez-Sanagustin (Eds), *Data Driven Approaches in Digital Education*. EC-TEL 2017. Lecture Notes in Computer Science, vol 10474. Springer, Cham. https://doi.org/10.1007/978-3-319-66610-5-48
 DuBay, W. H. (2004). *The principles of readability.*

- DuBay, W. H. (2004). *The principle's of readability. Online Submission.* Available at: https://files.eric.ed.gov/fulltext/ED490073.pdf.
 Flesch, R. (1948). A new readability vardstick, *Journal of Applied Psychology*, 32(3): 221-233. doi: 10.1037/h0057532
 Fry, E. (1968). A readability formula that saves time. *Journal of reading*, 11(7): 513-578.
 Garais, G. E., & Enaceanu, A. S. (2011). Determining quality levels for improving maintenance processes. *Annals of DAAAM & Proceedings*.
 Gunning, R. (1952). *The technique of clear writing*. McGraw-Hill.
 Kandel, L. and Moles, A. (1958). Application de l'Indice de Flesch à la langue française. *Cahiers d'Etudes de Radio-Television 19*: 253-274.
 Kincaid, J. P., Fishburne Jr, R. P., Rogers, R. L., & Chissom, B. S. (1975). Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. *Research Branch Report*.
 Leary, B. E., & Gray, W. S. (1935). *What Makes a Book Readable: With Special Reference to Adults of Limited Reading Ability*... University of Chicago Press.
 Lee, B. W., & Lee, J. H. J. (2023). Traditional readability formulas compared for English. *arXiv preprint arXiv:2301.02975*.
 Lucisano, P., & Piemontese, M. E. (1988). Gulpease: una formula per la predizione della leggibilita di testi in lingua italiana. *Scuola e città* 3: 110-124.
 Luhmann, J., & Burghardt, M. (2022). Digital humanities A discipline in its own right? An analysis of the role and position of digital humanities in the academic landscape. *Journal of Reading*, 12(8): 639-646.
 Moley, P. F., Bandré, P. E., & George, J. E. (2011). Moving beyond readability: Considering choice, motivation and learner engagement. *Theory into Practice*, *50*(3): 247-253.
 Shanahan, T., Fisher, D., & Frey, N. (2016). The challenge of challenging text. In: Scherer, M. (Ed.). (2016). On developing readers: readings from education

- Shahanan, I., Fisher, D., & Frey, N. (2016). The challenge of challenging text. In: Scherer, M. (Ed.). (2016). On developing readers: readings from educational leadership (EL Essentials) (pp. 100-109). Alexandria, USA: ASCD.
 Sirbu, M. D., Botarleanu, R. M., Dascalu, M., Crossley, S. A., & Trausan-Matu, S. (2018). ReadME–Enhancing Automated Writing Evaluation. In Artificial Intelligence: Methodology, Systems, and Applications: 18th International Conference, AIMSA 2018, Varna, Bulgaria, September 12–14, 2018, Proceedings 18 (pp. 281-285). Springer International Publishing.
 Smith, E. A., & Senter, R. J. (1967). Automated readability index (Vol. 66, No. 220). Aerospace Medical Research Laboratories, Aerospace Medical Division, Air Force Systems Command. Wright-Patterson Air Force Base, Ohio.
 Spache, G. (1953). A new readability of primary-grade reading materials. The Elementary School Journal, 53(7): 410-413.
 Stellner, B. (2013). Readability of Quarterly Reports: Do Companies Mislead Investors?. Anchor Academic Publishing (ap_verlag).
 Toma, I., Marica, A.-M., Dascalu, M., & Trausan-Matu, S. (2021). ReaderBench Automated Feedback Generation for Essays in Romanian. Scientific Bulletin, University Politehnica of Bucharest Series C 83: 21-34.

- Bucharest, Series C, 83: 21-34.
- Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes. Cambridge, MA: Harvard University Press.