

Estonian L2 vocabulary test LexEst

Kaidi Lõo, Katrin Leppik, Agu Bleve, Anton Malmi

Eesti ja üldkeeleteaduse instituut, Tartu ülikool

Eesti Rakenduslingvistika Ühingu Kevadkonverents

27.04.2023

Motivation

- Language proficiency is often measured with questionnaires when assessing the language skills of an L2 speaker
- Self-evaluation of vocabulary is subjective and difficult
- Quick and objective tests are needed to evaluate L2 language skills

LexTale vocabulary tests

- Vocabulary size could be measured with a simple 5-minute lexical decision task
- The task is to decide whether a stimulus is a word (e.g., **mug**) or a nonword (e.g., **muk**).
- A percentage of correct vs. incorrect answers reflect L2 speakers' vocabulary knowledge
- Also found to be a good indication of L2 general proficiency

LexTale vocabulary tests

- LexTale (*Lexical Test for Advanced Learners of English*) originally developed to test advanced L2 speakers (Lemhöfer & Broersma, 2012)
- Later on shown suitable for testing language knowledge beyond advanced language levels (Amenta et al. 2021, Salmela et al. 2021)
- Possibly also suitable for testing L1 speakers language knowledge: e.g., children and L1 speakers of different language background (Salmela 2021)
- Widely used in psycholinguistic experiment settings

- LexTale - Lexical Test for Advanced Learners of English (Lemhöfer & Broersma, 2012)
- LexIta - Assessment Tool for Italian L2 Receptive Vocabulary Size (Amenta et al., 2021)
- Lexize: A test to quickly assess vocabulary knowledge in Finnish (Salmela et al. 2021)
- LexTALE_FR (Brysbaert, 2013)
- Lextale-Esp (Izura, Cuetos & Brysbaert, 2014)
- LEXTALE_CH (Chan & Chang, 2018)
- Estonian and Swedish versions currently in development

Estonian

- Estonian language skills, incl. vocabulary measured via CEFR level testing in schools and by Education and Youth Board
- EKI tools enable classifying vocabulary in Estonian according to CEFR levels (Kallas et al. 2021)
- No standardized quick Estonian L2 test available to test language skills, vocabulary knowledge or otherwise
- Lexical decision task has been used before on Estonian (Lõo et al. 2018)

LexEst - Estonian L2 vocabulary test

- Almost ready to start collecting data for validating the test
- LexEst modelled after the Finnish version in collaboration with researchers from the University of Turku
- Hosted on ViLLE - a collaborative learning platform
- To be used for all levels of Estonian L2 speakers



Thank you for taking an interest in our experiment. We invite you to participate in an experiment that studies language learners' vocabulary. The results of this experiment will help us to create a vocabulary test that quickly estimates the language proficiency of a language learner.

If you choose to participate, you are asked to fill out a questionnaire about your background. After that, you are directed to the experiment page, where you are presented with letter sequences of which you have to decide whether it is a word in Estonian (e.g., aeg) or not (e.g., trui). By completing the test, you will substantially contribute to the research in language learning and acquisition. Your data is stored anonymously.

Find a quiet place where you can concentrate without being disturbed for 15 minutes. Your participation is voluntary. You can stop or withdraw from the study at any point in time without any consequences. This experiment has been approved by the Research Ethics Committee of the University of Tartu.

I have read and understood the information letter about this vocabulary study. I understand that I may withdraw from the study at any time without consequences. I agree that the data collected within the context of this study may be publicly shared as long as no person-identifying information will be made public. By clicking "submit" I consent to the above statements. I certify that I am at least 18 years old and willing to participate in the experiment.

The experiment has been created by the Institute of Estonian and General Linguistics, University of Tartu. For more information contact us at lexest@lists.ut.ee.

☐ I am at least 18 years old and willing to participate in the experiment.



Background questionnaire



What is your email address? (Note that this information is collected for connecting your background information with the results of the test. We will not send you spam or distribute your email to third parties.)



What is your gender?

- ☐ male
- ☐ female
- ☐ non-binary



How old are you?

word



yes

S

no

L

Next you will see Estonian words or nonwords. If the item is a word, select yes (S on keyboard). If the item is a nonword, select no (L on keyboard). If you are using a tablet or a phone, press the buttons on the screen.

Let's practice first with a few items. During practice, you will be informed if you answer incorrectly. In such case, please choose the correct alternative.

Start

SOOV

yes

S

no

L

umuk

yes

S

no

L

Stimulus creation

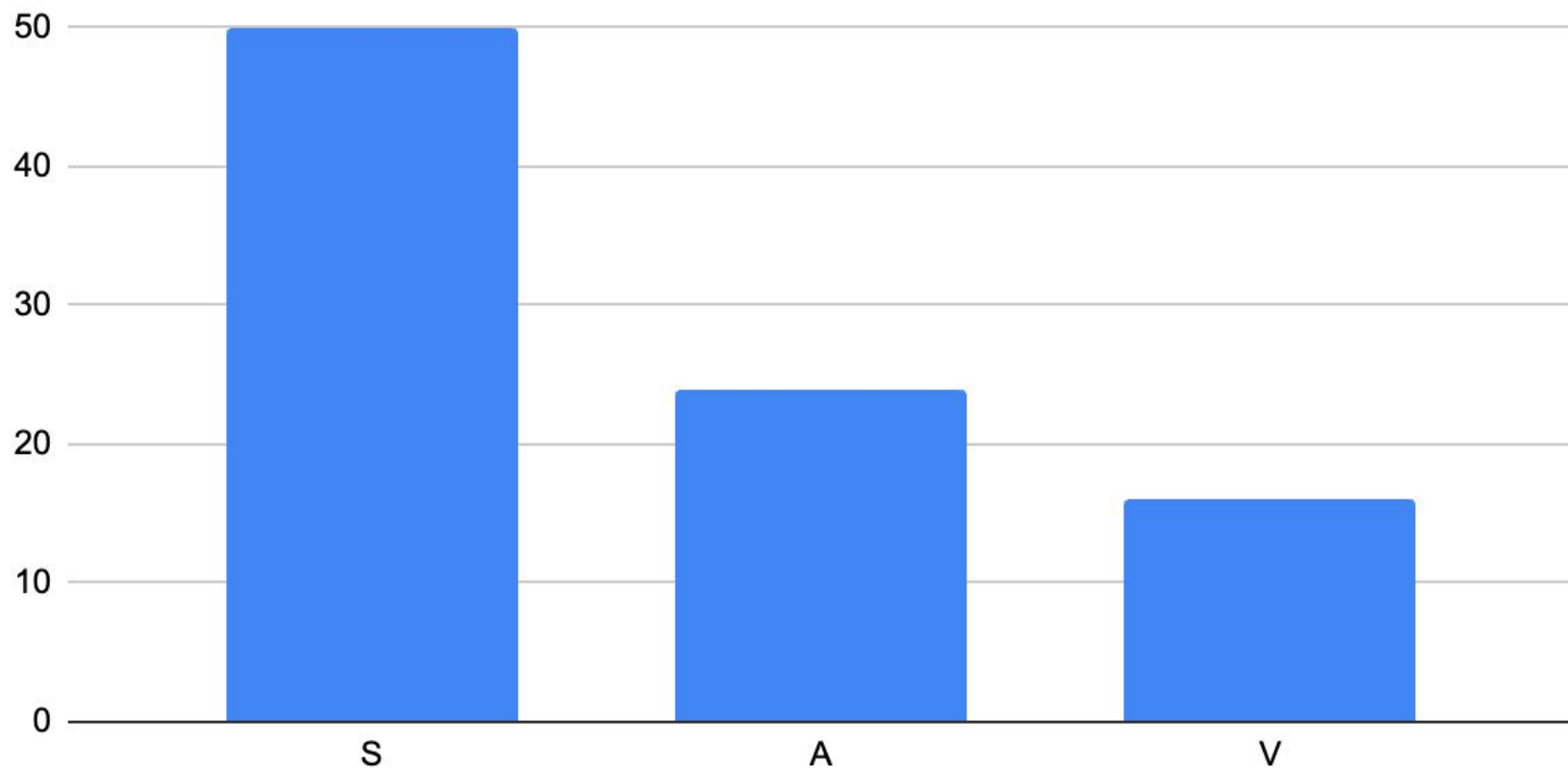
- Balanced Corpus of Estonian (15 mil tokens) divided into six frequency bands
- Random sample of 30 words from each frequency band
- Excluding words with
 - foreign/cognate origin (e.g., **šõu** - show, **klaver** - piano (das Klavier))
 - possibly offensive words
 - words with lexical ambiguity (e.g., **luud** - bones/broom)
 - derivation (e.g., **janune** - thirsty).

Word list

- **90 items**

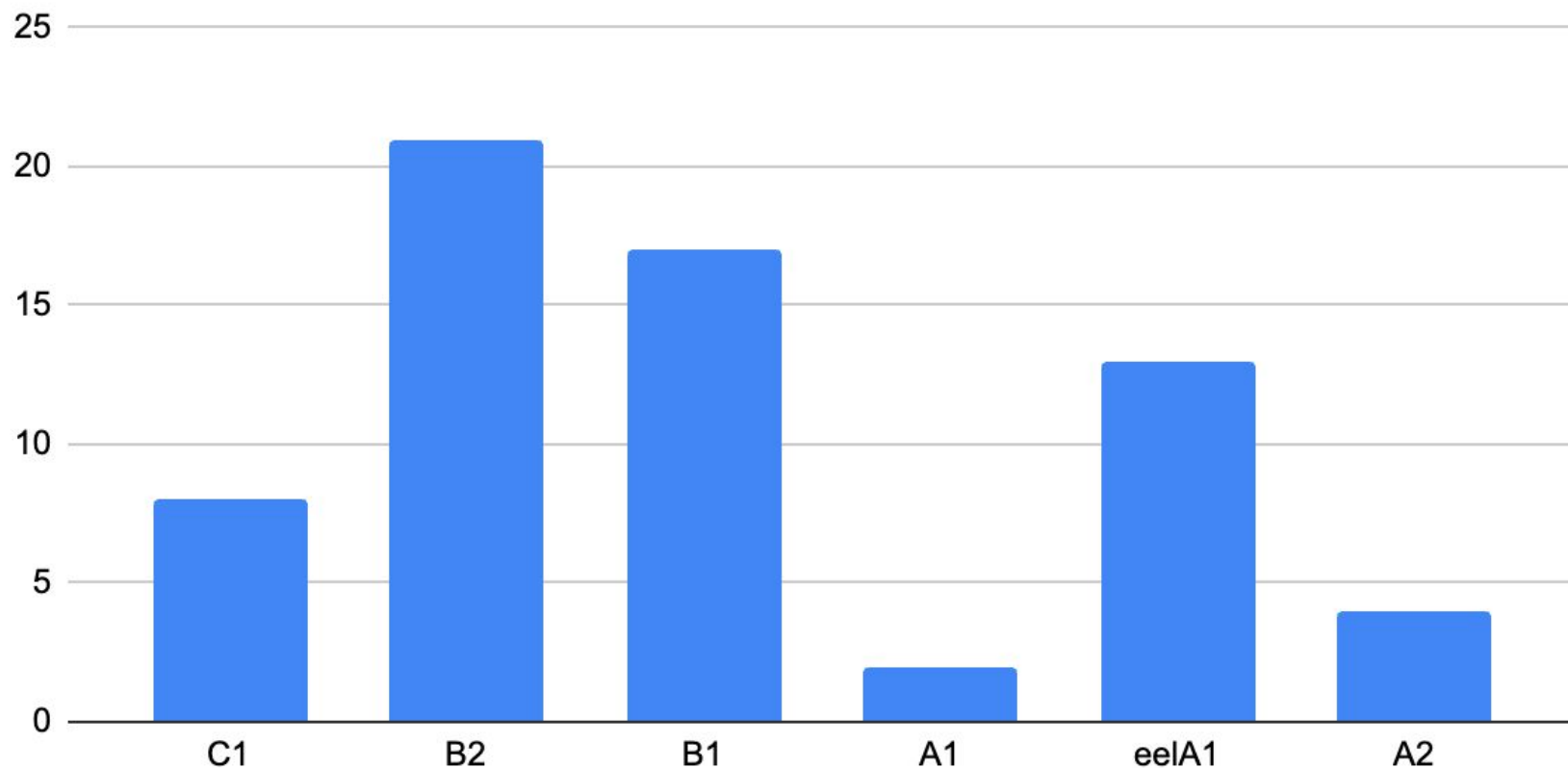
- 10 words with frequency <1 per million (pm);
- 24 words with frequency 1–5 pm;
- 22 words with frequency 5–10 pm;
- 18 words with frequency 10–20 pm;
- 10 words with 20–100 pm;
- 6 words with frequency more than 100 pm

Count of POS



Count of POS

Count of Proficiency EKI



Count of Proficiency EKI

Pseudoword list

- **45 items**

- Another set of Estonian words from a corpus mirroring the final word list in frequency and length
- Changing 1-3 letters in real Estonian words to make them pseudowords (e.g., m**a**ja > k**l**a**j**a)
- Checking that pseudowords were not words in other major foreign languages in Estonian (e.g., Finnish - **talo**, English - **tear**) or words in Estonian dialects (e.g. **loda**)

Problems with selection of stimuli

- How to define derivation (e.g. is it important to exclude unproductive derivation or etymologically derived words whose derivation is not transparent)?
- How to define words of foreign origin, what is the time depth to perceive a word as of not foreign origin?
- How to create good pseudowords?

Expected participants

- First phase
 - 200 adult nonnative speakers of Estonian from different language levels
 - 100 adult native speakers of Estonian of different background (not only university students and university employees, gender and age balance)
- Future
 - Estonian children, Estonian bilinguals in Finland

Conclusions

- There is a practical need for a quick test to assess Estonian L2 speakers' knowledge of Estonian (especially in experimental settings)
- LexTale and similar tests have shown to be suitable for this purpose.
- LexEst will be quick vocabulary size assessment tool freely available for everyone to use

Participate in our study!

<http://sisu.ut.ee/lexest>

Acknowledgements

This study is supported by the Kadri, Nikolai and Gerda Rõuk Memorial Fund.

References

- Amenta, S., Badan, L., & Brysbaert, M. (2021). LexITA: A quick and reliable assessment tool for Italian L2 receptive vocabulary size. *Applied Linguistics*, 42(2), 292-314.
- Brysbaert, M. (2013). Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica*, 53(1), 23-37.
- Chan, I. L., & Chang, C. B. (2018, May). LEXTALE_CH: A quick, character-based proficiency test for Mandarin Chinese. In *Proceedings of the 42nd Annual Boston University Conference on Language Development* (
- Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, 35(1), 49-66.
- Kallas, J., Koppel, K., Pool, R., Tsepelina, K., Üksik, T., Alp, P., & Epner, A. (2021). Eesti keele kui teise keele õpetaja tööriistad Eesti Keele Instituudi keeleportalis Sõnaveeb. *Eesti Rakenduslingvistika Ühingu aastaraamat*, 17, 61-80.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44, 325-343.
- Lõo, K., Järvikivi, J., & Baayen, R. H. (2018). Whole-word frequency and inflectional paradigm size facilitate Estonian case-inflected noun processing. *Cognition*, 175, 20-25.
- Salmela, R., Lehtonen, M., Garusi, S., & Bertram, R. (2021). Lexize: A test to quickly assess vocabulary knowledge in Finnish. *Scandinavian journal of psychology*, 62(6), 806-819.