

Estonian as a Second Language Teacher’s Tools

Tiiu Üksik, Jelena Kallas, Kristina Koppel, Katrin Tsepelina, Raili Pool
The Institute of the Estonian Language / Roosikrantsi 6, 10119 Tallinn, Estonia
(tiiu.yksik, jelena.kallas, kristina.koppel)@eki.ee,
katrin.tsepelina@eki.ee, raili.pool@ut.ee

Abstract

The paper describes “Teacher’s Tools” (et *Õpetaja tööriistad*) developed by the Institute of the Estonian Language for teachers and specialists of Estonian as a Second Language. The toolbox includes four modules: vocabulary, grammar, communicative language activities and text evaluation. The vocabulary module provides graded word lists for young (CEFR levels pre-A1–C1) and adult (CEFR levels A1–C1) learners. The grammar module provides descriptions of young learners’ grammar competence. The communicative language activities module gives teachers an overview of the typical situations that learners should be able to cope with. The text evaluation module marks lemmas in texts according to their CEFR assignment in the vocabulary module. So far the grammar and the communicative language activities modules have been developed only for young learners (CEFR levels pre-A1–B2). The toolbox is aimed at providing support for the development of Estonian as an L2 courses, educational materials, exercises and tests within a CEFR-based framework. The project started in 2017 and is ongoing.

1 Introduction

“Teacher’s Tools” is a compatible toolkit developed for teachers and specialists of Estonian as a Second Language, providing an extensive overview of the development of lexical and grammatical competence in Estonian among L2 learners. The toolkit is accessible as a sub-page of the language portal [Sõnaveeb](https://sonaveeb.ee)¹. The framework of the project is based on the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR, 2001), its Companion Volume with New Descriptors (CEFR/CV, 2018) and the Collated Representative Samples of Descriptors of Language Competences Developed for Young

¹<https://sonaveeb.ee/teacher-tools>

Learners for Ages 7–10 (Szabo, 2018a) and 11–15 (Szabo, 2018b). The toolkit distinguishes between adult (CEFR levels A1–C1) and young learners (CEFR levels pre-A1–C1). The methodology of the project is adapted from similar projects for other languages (e.g. Capel, 2010, 2012; O’Keeffe and Geraldine 2017; Alfter et al., 2019).

2 Resources of Estonian as an L2

There are quite a few corpora available for the research of the Estonian Language (the biggest Estonian corpus is the “Estonian National Corpus 2019” (1,5 billions words))². However, at the beginning of the project in 2017 there was a clear lack of resources available for the research of Estonian as an L2. In order to fill this gap, the Institute of Estonian Language compiled two types of corpora: 1) coursebook corpora and 2) learner corpora. As Volodina and Kokkinakis (2013) point out, the first type of corpora provides information about what and when education professionals found important for students to learn (passive linguistic competence), while learner corpora provide an indication of active linguistic competence. Both need to be studied since the second is directly influenced by the first. The coursebook corpora were compiled in several stages and resulted in two groups of corpora: “Estonian as a Second Language Coursebook Content Corpus 2017”³ (contains eight coursebooks for adult learners at levels A1–C1; 500 000 tokens) and “Estonian as a Second Language School Coursebook Content Corpus 2021”⁴ (contains 27 school coursebooks for grades 1–12; 1 300 000 tokens). The learner corpus was compiled in 2019–2021 and contains 6700 Estonian as an L2 young learner’s texts. These are texts written mostly as state standard-determining tests (grades 3 and 6), basic school final examinations

²DOI: [10.15155/3-00-0000-0000-0000-08489L](https://doi.org/10.15155/3-00-0000-0000-0000-08489L)

³DOI: [10.15155/3-00-0000-0000-0000-06ADEL](https://doi.org/10.15155/3-00-0000-0000-0000-06ADEL)

⁴DOI: [10.15155/3-00-0000-0000-0000-0888BL](https://doi.org/10.15155/3-00-0000-0000-0000-0888BL)

Keeleõppija: noor keeleõppija täiskasvanu

Otsi sõna

Keeletase:
 eelA1
 A1
 A2
 B1
 B2
 C1

Sõnaliik:
 nimisõna
 omadussõna
 asesõna
 tegusõna
 määrsõna
 kaassõna
 sidesõna
 hüüdsõna
 arvsõna

Temaatilised loendid:
 Vali teema

Tühista filtrid

Otsingu tulemused (6274) TXT XLSX Näita 20 ▾

Sõna	Sõnaliik	Tase
aabits	nimisõna	A1
aadress	nimisõna	eelA1
aare	nimisõna	C1
aasta	nimisõna	eelA1
aastaaeg	nimisõna	eelA1
aastane	omadussõna	A2
aastapäev	nimisõna	A2
aastatuhat	nimisõna	B2
aastavahetus	nimisõna	B1

Figure 1: Vocabulary module

(grade 9) and state exams (grade 12) for 2017–2021. The data is available for analysis through the interface of the [Estonian Learner Language Corpus EMMA](#)⁵. The texts were not corrected or altered. Both types of corpora were used for the development of “Teacher’s Tools”. The coursebook corpora were used primarily to create passive vocabulary word lists and for the analysis of explicit and implicit grammar teaching. The learner corpus was used for the creation and analysis of active vocabulary word lists and the dynamics of grammar acquisition.

3 Modules of “Teacher’s Tools”

“Teacher’s Tools” consists of four modules: vocabulary, grammar, communicative language activities and text evaluation.

3.1 Vocabulary

The vocabulary module includes young and adult learners’ word lists compiled on the basis of coursebook and learner corpus word lists in comparison with frequency distribution in the “[Estonian National Corpus 2017](#)”⁶. Currently, the vocabulary list for adult learners includes about 12 500 words for levels A1–C1. The young learners’ word list includes approx. 9000 words for levels pre-A1–B2. All words are presented as lemmas. Users can generate word lists based on learner type (young or adult) and POS. In addition, 24 topics (*clothes, furniture, birds* etc.) are compiled. The search results can be sorted alphabetically, by level or by POS (see Figure 1).

⁵DOI: 10.15155/1-00-0000-0000-0000-001A5L

⁶DOI: 10.15155/3-00-0000-0000-0000-071e71

3.2 Grammar

The grammar module is the first attempt to create a systematic overview of Estonian L2 learner’s grammar competence development. The general methodology was adopted from the [English Grammar Profile](#)⁷ project (O’Keeffe and Geraldine, 2017). Currently, the grammar module provides descriptions of grammar competence on the morphology, derivation, phrase and sentence levels, from level pre-A1 up to level B2 (see Figure 2). The same grammatical category (e.g. the use of the genitive case for nouns or the use of imperative mood for verbs) may be described at all levels, but the functions and usage (how often the learner makes mistakes and what kinds of mistakes) are level-specific. Search options offer users choices of four main categories: morphology, derivation, phrase formation and sentence formation. The main categories contain 19 subcategories. For example, subcategories of *morphology* are *verb, noun, adjective, numeral, pronoun, adverb*. Every subcategory has grammatical markers (e.g. *number* and *case for noun*). The grammatical markers have values, for example *number* has values *singular* and *plural*. All descriptions are equipped with example sentences compiled either by experts or taken from coursebook and learner’s corpora.

3.3 Communicative language activities

The communicative language activities module is based on CEFR/CV descriptor scales, and examples are taken from Szabo (2018a,b), which include reception processes and strategies of written

⁷<https://www.englishprofile.org/english-grammar-profile/egp-online> (13.01.2021)

Grammatika

Grammatika osa sisaldab noore keelearnija grammatikapädevuse kirjeldust. Noore keelearnija tasemekirjelduste vastavused keelepädevusega on järgmised: eelA1 = koolieelikud (6–7-aastased), A1 = I kooliaste (1.–3. klass), A2 = II kooliaste (4.–6. klass), B1 = III kooliaste (7.–9. klass), B2 = gümnaasium (10.–12. klass). Täiskasvanute grammatikapädevuse kirjeldus on väljatöötamisel.

TAGASISIDE

Keelearnija: noor keelearnija

Keeletase: eelA1 A1 A2 B1 B2

Peakategooria

Alamkategoriad

VORMIMOODUSTUS

TEGUSÖNA

arv

mitmus

+

OTSI

+ Lisa peakategooria

Otsingu tulemused (37)

Näita 20

Peakategooria	Alamkategoriad	Tase	Kirjeldus
VORMIMOODUSTUS	TEGUSÖNA	eelA1	<p>kindla kõneviisi olevik jaatavas kõnes (mitmuse 1., 2. ja 3. pööre)</p> <p>Oskab sagedasi eakohaseid tegusõnu kasutada kindla kõneviisi oleviku mitmuse esimese pöörde vormis (nt <i>teeme</i>). Kasutab neid vorme eakohasel viisil moodustatud lausetes, et väljendada enda ja kaaslase koos toimuvat tegevust (nt <i>sööme</i>) või ergutada kaaslast endaga koos mingit tegevust alustama (nt <i>lähme</i>). Mõistab ka mitmuse 2. ja 3. pöördes tegusõnade kasutust (nt <i>sööte</i>, <i>söövad</i>), kuid ise kasutab neid väga harva. Omandab verbivormid valdavalt terviküksustena. Eksib pöördelõppude kasutamisel.</p> <p>Näited: <i>Meie sööme. Lähme õue! Teie mängite. Poisid jooksevad.</i></p> <p>Tunnused: kõneviis <i>kindel</i> tegumood <i>isikuline</i> aeg <i>olevik</i> pööre <i>1 2 3</i> arv <i>mitmus</i> kõneliik <i>jaatav</i></p>

Figure 2: Grammar module

Keelearnija: noor keelearnija

Keeletase: A1 A2 B1 B2

Suhtlustoiming

Näidiskaala

TEKSTI VASTUVÖTT

üldine kuulamisoskus

suhtlusvorm

suuline

+

OTSI

+ Lisa suhtlustoiming

Otsingu tulemused (16)

Näita 20

Suhtlustoiming	Näidiskaala	Tase	Kirjeldus
TEKSTI VASTUVÖTT	üldine kuulamisoskus	A1	<p>Kasutusolukord: kirjeldusest arusaamine</p> <p>Tingimused/piirangud: väga lihtne, lühikesed laused</p> <p>Kasutusolukorra näide: <i>Saan aru minu magamistoa või klassiruumi kirjeldusest. Saan aru mõne asja suuruse, värvi või asukoha kirjeldusest. Saan aru, kui keegi räägib ilmast, nt täna on külm. Saan aru, kui keegi kirjeldab oma kodus tubade asukohta, nt köök on elutoa kõrval.</i></p> <p>Tunnused: suhtlusvorm <i>suuline</i> tekstiliigi näited <i>monoloog</i> teemade näited <i>asjad ilm klass kodu kujundid, vastandid, värvid ruumid</i> teksti alaliigi näited <i>kirjeldus</i></p>

Figure 3: Communicative language activities module

Keelearnija:

noor keelearnija täiskasvanu

Sisesta tekst:

667 / 2000

Eesti keeles on 14 käänet, mille üle eestlased on iseäranis uhked. Selletõttu aga ei ole eesti keele lauses sõnade järjekord kindlate reeglitega määratud, nagu on seda paljudes indo-euroopa keeltes. Käänete rohkus ei garanteeri veel, et lauseliikmete süntaktilised funktsioonid oleks nende käänete vahel võrdselt ära jaotatud. Ei, nii lihtne eesti keel

Keeleoskustase: A1 A2 B1 B2 C1

Eesti keeles on 14 käänet, mille üle eestlased on iseäranis uhked. Selletõttu aga ei ole eesti keele lauses sõnade järjekord kindlate reeglitega määratud, nagu on seda paljudes indo-euroopa keeltes. Käänete rohkus ei garanteeri veel, et lauseliikmete süntaktilised funktsioonid oleks nende käänete vahel võrdselt ära jaotatud. Ei, nii lihtne eesti keel siiski ei ole! Enamasti sõltub sõna süntaktiline funktsioon ka veel lause tähendusest, see aga on arvutile seni veel ületamatu raskus. Näiteks kui vaadata kahte lauset « Lapsed sõid need kommid ära » ja « Need kommid sõid lapsed ära », siis nendes lausetes saab alust ja sihitist määrata ainult lause tähendust teades.

Sõnu kokku: 98

Tase	Sõnu	%
A1	55	56.12%
A2	18	18.37%
B1	11	11.22%
B2	4	4.08%
C1	5	5.10%
määramata	5	5.10%

HINDA

Figure 4: Text evaluation module

and spoken texts (listening and reading), written and oral text production and production strategies, and written and spoken interaction and interaction strategies. It covers young learners at levels A1–B2. The architecture of the module is similar to that of the grammar module, containing main categories (reception, production, interaction and mediation) and subcategories, which are followed by detailed descriptions, including limitations: when and how well the language user can manage certain communicative activities (see Figure 3). For example, by choosing the main category *reception* and the subcategory *overall listening comprehension*, a teacher can see that at the A1 level the learner should be able to understand a simple description with short sentences, for example about the weather or how big a certain object is.

3.4 Text evaluation

The text evaluation tool helps to assess Estonian texts for their degrees of complexity according to the CERF (see Figure 4). Currently, the tool takes into account only lexical information and defines the CEFR level of each particular lemma based on CEFR vocabulary lists (see Chapter 3.1). Similar tools have also been developed for many other languages (see e.g. Alfter, 2021). The program runs on EstNLTK v 1.6, which offers functionality to lemmatise and perform morphological analysis. The tool needs to be developed further. First, there is a need to implement methods for the improvement of the analysis of homonyms (for example *tamm* can mean either an *oak* or a *dam*, which is assigned different levels in word lists) and multi-word expressions. So far, the analysis is based only on single word lists, which is not sufficient.

4 Summary

“Teacher’s Tools” is the first attempt to provide a systematic overview of the development of lexical and grammatical competence in Estonian as a Second Language. The project is a work in progress and further development of the toolkit is foreseen. We plan to add descriptions of the development of grammar competence and communicative language activities for adult learners. The enrichment of the text evaluation module with the possibility of measuring grammatical difficulty and readability (by for example adding Lix-index value) is under development. “Teacher’s Tools” as a resource can be used for different CALL tasks, including auto-

matic CEFR-related vocabulary and grammar exercise generation or lexical simplification tasks. The use of NLP for the development of such computer-assisted tools has enormous potential for enhancing the teaching and learning of Estonian as an L2.

References

- Estonian Learner Language Corpus EMMA = Eesti keele õppija korpus EMMA. University of Tartu.
- Language portal Sõnaveeb. The Institute of the Estonian Language.
- CEFR = the common european framework of reference for languages: Learning, teaching, assessment. 2001. Council of Europe, Strasbourg.
- Estonian as a Second Language Coursebook Content Corpus 2017. 2017. The Institute of the Estonian Language.
- Estonian National Corpus 2017. 2017. The Institute of the Estonian Language.
- CEFR/CV = Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion Volume with New Descriptors. 2018. Council of Europe, Strasbourg.
- Estonian National Corpus 2019 (.vrt format). 2019. The Institute of the Estonian Language.
- Estonian as a Second Language School Coursebook Content Corpus 2021. 2021. The Institute of the Estonian Language.
- David Alfter. 2021. *Exploring natural language processing for single-word and multi-word lexical complexity from a second language learner perspective*. Ph.D. thesis, University of Gothenburg, Faculty of Humanities.
- David Alfter, Lars Borin, Ildiko Pilán, Therese Lindström Tiedemann, and Elena Volodina. 2019. *Lärka: From Language Learning Platform to Infrastructure for Research on Language Learning*, Linköping Electronic Conference Proceedings. Linköping University Press. CLARIN Annual Conference ; Conference date: 08-10-2018 Through 10-10-2018.
- Annette Capel. 2010. A1-B2 vocabulary: Insights and issues arising from the English Profile Wordlists project. *English Profile Journal*, 1(1):1–11.
- Annette Capel. 2012. Completing the English Vocabulary Profile: C1 and C2 vocabulary. *English Profile Journal*, 1(2):1–14.
- Anne O’Keeffe and Mark Geraldine. 2017. The english grammar profile of learner competence: Methodology and key findings. *International Journal of Corpus Linguistics*, 4(22):457–489.

Tunde Szabo. Collated Representative Samples of Descriptors of Language Competences Developed for Young – Volume 2: Ages 11-15. 2018a. Council of Europe.

Tunde Szabo. Collated Representative Samples of Descriptors of Language Competences Developed for Young – Volume 2: Ages 11-15. 2018b. Council of Europe.

Elena Volodina and Sofie Johansson Kokkinakis. 2013. *Compiling a corpus of CEFR-related texts*. Antwerpen, Belgium. May 27–29.