

Margit Langemets

## “To think outside the paper”: The case of Estonia

### Abstract

This article intends to give an overview of the lexicographic scene in Estonia from the perspective of modern lexicography, i.e. e-lexicography. We have been developing our in-house Dictionary Writing System EELex for some time already. However, the modern lexicography expects more: the multifunctionality of the database, special needs of the specific user, etc. I will briefly characterise some of the lexical resources of Estonian, tools for lexicographers, as well as some theoretical matters together with the issues of language technology, e.g. the concept-based method vs. word-based method (Tavast 2008) for compiling dictionaries. I will sum up with an overview of typical properties of online Estonian dictionaries.

### 1. The background

Strategic planning of the development of the Estonian language began in 1998 when the Ministry of Education was assigned the task of drawing up a development strategy of the Estonian language. In 2000 the Estonian Committee for the organization of the European Year Languages (later the Estonian Language Council) was set up. In 2004 the first Development Strategy of the Estonian Language 2004-2010 was adopted by the Government. Since 2011 we have been living an age of the new strategic plan, The Development Plan of the Estonian Language 2011-2017 (henceforth, The Plan 2011). All the activities foreseen by The Plan (2011) are expected to be funded from the budgets of the state and local governments, universities and research institutions, national programmes, etc.

This article is written from the perspective of the Institute of the Estonian Language<sup>1</sup> (Tallinn, henceforth the Institute), thus the expression of ‘the case of Estonia’ in its title reflects foremost the lexicographic work practiced at the Institute, not Estonia as a whole. However, it might be justified for at least some reasons.

Firstly, the Institute is one of the many establishments and institutions, which have significant positions from the perspective of use and development of The Plan (2011). It is an important participant in language management, the study of Estonian and language resources, as well as language-technological applications. Secondly, the Institute is the owner and developer of several lexical resources – in the current state of language technology support, characterised as “cautiously optimistic” in the META-NET White Paper Series (Liin et al. 2012, 35). Since 2012 the Institute has been a member of the Center of Estonian Language Resources (CELR),<sup>2</sup> together with the University of Tartu and the Tallinn University of Technology. Thirdly, the Institute was the organiser of the third biennial conference of electronic lexicography in the 21st century eLex 2013,<sup>3</sup> Tallinn,

---

<sup>1</sup> Institute of the Estonian Language (<http://portaal.eki.ee/>, last visited: 28.01.2013).

<sup>2</sup> The Center of Estonian Language Resources ([http://keeleressursid.ee/ressursid\\_eki.html](http://keeleressursid.ee/ressursid_eki.html), last visited: 28.01.2013).

<sup>3</sup> eLex 2013 (<http://eki.ee/elex2013/>, last visited: 28.01.2013).

together with Trojina Institute for Applied Slovene Studies, as well as a participant in previous eLex conferences.<sup>4</sup> And last but not least, the Institute is widely acknowledged in Estonia to compile and edit dictionaries essential for the national culture.

## 2. Tasks for modern lexicography

The overall and very general task for lexicography, also mentioned in The Plan (2011), is to provide society with language resources (dictionaries, manuals, and databases) and guidance (incl. advice and courses). The other very general claim is that resources should be trustworthy and efficient, i.e. the content should be very carefully designed and fulfilled, as well as the mechanisms for reaching it should be fast and comfortable, preferably all in one.

The conferences on electronic lexicography and other forums have discussed the differences between paper and electronic dictionaries, as well as comparisons of their advantages and disadvantages for quite a long time already. However, the focus nowadays has moved to find different ways of how to exploit the rich potential of electronic medium in order to respond (quickly) to the needs of the new types of users, as well as to the needs of modern lexicographers. The slogan of eLex 2013 has been worded as “to think outside the paper”, i.e. one has to forget about conventional approaches and be innovative, to conceptualize the dictionary with an electronic format in mind. This is not at all an easy task for lexicographers to get used to working in new environments, in different dictionary systems, rather thinking systematically following the rules of the advanced database than simply writing down a next dictionary article in line with all information one can remember. For users it is definitely much easier: they just formulate their queries, the simpler the better – in any device, in any place – and expect quick, correct and easily understood answers. The system builder, together with lexicographers, has to perform like a fortune teller to guess all possible needs of a user, when speaking of dictionaries.

Fuertes-Oliveira/Bergenholtz (2011) have proposed broadening lexicographic theory to other reference sources, considering lexicography as an integral part of information science with the aim of ‘solving the problems’ of the user, developing access to different (extralexical) data and different data presentation possibilities (e- and p-lex; pupil, semi-expert, or expert, etc.). They introduce the notion of dynamic SmartDictionaries or ‘true lexicography’, however, they do not focus on the content of the lexicographic system, i.e. gathering, analysing, processing, etc. of the data, expecting somehow that all the necessary content is already there. They even have expressed hostility towards linguistics and the role of linguistic theory in presenting lexicographical data (Fuertes-Oliveira/Bergenholtz 2011, 20-29).

Estonia is still at the stage of e-dictionaries and their pros and cons, not of dynamic SmartDictionaries or ‘true lexicography’, quoting Fuertes-Oliveira/Bergenholtz (2011).

---

<sup>4</sup> eLex 2009 was held in Louvain-la-Neuve, Belgium, and eLex 2011 in Bled, Slovenia.

### 3. Dictionary Writing System EELEX

We have been developing our web-based Dictionary Writing System EELEX<sup>5</sup> since 2005 and by now nearly 50 dictionaries, incl. 7 terminological ones, have been compiled and edited using the system.

One of the earlier public applications was the Estonian-X dictionary database, which was compiled to provide a core for the new bilingual dictionaries to be produced by the system (Langemets et al. 2010). The database contains only source language (Estonian) data: headword, grammatical information, definitions, usage examples, etc. The data of the target language – translation equivalents with the rest of the necessary information – will be supplied by the user compiling the new dictionary. The system presents a preliminary standard form of the entry, open for the lexicographer who can modify, if necessary, both the content and form of the entry.

One of the newest functions of EELEX is generating dictionaries of a particular type by selecting individual elements from the source dictionary database(s) (Kallas/Langemets 2012). It is possible to generate specialized dictionaries by reorganizing the preview (and layout) of the existing dictionary articles, or generating a brand new dictionary database. The latter is still in need of improvement, together with the function of importing data from several databases into one target database. The existing dictionary articles might be reorganized using the article preview generator function or the customization function of the XML query. Both options display the content of the dictionary for a user (i.e. lexicographer) in a new, individualised way (Kallas/Langemets 2012).

There are still many features needed, e.g. a lemmatiser to identify all the base forms of the constituents of (a) the multiword expressions (MWE) (for linking MWE-s to several headwords), or (b) the definitions in the learner's dictionary (for checking the controlled vocabulary). Automatic linking of sense references has yet to be improved. The overall user-friendliness is almost lacking, especially compared to commercial DWS-s,<sup>6</sup> neither do we have very nice tutorials.

### 4. Lexical resources: databases and dictionaries

The META-NET White Paper declares the situation with regard to lexical resources as “reasonably good for Estonian since substantial resources have been built in recent decades” (Liin et al. 2012, 35). However, on a five-point scale the Speech and Text Resources for Estonian were evaluated as last but one (“Fragmentary support”), together with Danish, Finnish, Slovene, etc., a step behind German, Dutch, Hungarian, etc. (“Moderate support”), whereas the best and only one in this category, English, was scored as “Good support” (yet not “Excellent”) (Liin et al. 2012, 64). My personal opinion on this matter is that regarding the dictionaries and lexical databases – not dealt with special care in the White Paper – the situation is much better: “reasonably good” as stated in the above overview. “Fragmentary” would be fairer when evaluating access modes to online dictionaries (see below).

---

<sup>5</sup> EELEX (<http://eelex.eki.ee/>, last visited: 28.01.2013).

<sup>6</sup> See T-Lex (<http://tshwanedje.com/tshwanelex/>), Léacslann (Dublin) (<http://lxln.prettydata.eu/>), iLEX, the integrated XML system ([www.emp.dk/ilexweb/](http://www.emp.dk/ilexweb/)) (all last visited: 28.01.2013).

As mentioned earlier, there are nearly 50 dictionaries of different types (monolingual and bilingual, general and learners' dictionaries, etc.) in our Dictionary Writing System EELEX. The standard XML mark-up and flexible functions should make EELEX a huge multi-purpose lexicographic repository consisting of different databases. In the following, I will mention only some of them.

**(1) The Estonian-X dictionary database** was developed in different sizes: 80,000 head-words for a voluminous dictionary (on the basis of an Estonian-Russian dictionary, 1997-2009), 40,000 for a medium-sized dictionary, 15,000 for a small, and later on, 5,000 for a basic dictionary. Also, according to the type of dictionary required, three standards of morphological description have been developed to facilitate the presentation of Estonian morphology (Langemets et al. 2010). The full morphological description (all basic inflectional forms, the inflectional type number, part of speech) is presented brackets < > (see Figure 1); |.....| marks the place for translation equivalent and |----| the place for the translation of the definition, if necessary.

**kuld** <k'uld kulla k'ulda k'ulda, k'ulda[de k'ulda[sid\_&\_k'uld/i 22 S>  
**1** (*teatud väärismetall* 'a yellow precious metal') |----|  
 ◆ *puhas kuld* |.....|; *kullast ehted* |.....|  
**2** (*värvuselt ja läikelt kulla sarnane* 'colour resembling gold') |----|  
 ◆ *päikesekuld* |.....|; *sügiskuld* |.....|; *kased puistavad juba kulda* |.....|  
**3** (*midagi väärtuslikku ja head* 'sth highly respected') |----|  
 ◆ *tema nõuanded on kulda väärt* |.....|

Figure 1: The preliminary dictionary entry for *kuld* ('gold') (Langemets et al. 2010)

**(2) Dictionary of Word Families** (Vare 2012) serves as a Word Formation database of Estonian. The compounding and derivation are the main productive devices for forming new lexical items. The design of the database is based on the word families method, which consists of the organization of words on the basis of common stem morphemes and word formation relations (Viks et al. 2010). As an agglutinative-fusional language, Estonian is characterised by a rich word formation system in which different word formation kinds, types and means combine in complex ways, and stems are subject to different types of change. It is headed by the simplex word (the head of the family) that represents the common stem. Inside the word family, words (family members) are arranged semasiologically, according to word formation. The words are organized in an integrated hierarchical network on the basis of a stepwise immediate constituent analysis, which permits to visualise the internal structure of complex words by simultaneously showing their base word and their immediate constituents (Figure 2).

**SPORT** 'sport'  
 sport=**lane** ('sport=NOUN SUFFIX, sportsman, athlete')  
 sport|**las=lik** ('sportsman=ADJECTIVAL SUFFIX, sportsmanly')  
 sport|**las|likk=us** ('sportsmanly=NOUN SUFFIX, sportsmanship')

Figure 2: An excerpt of the word family *sport* 'sport' (Viks et al. 2010)

The majority of Estonian vocabulary consists of derivations and compounds: the database contains about 9,000 lexical entries, i.e. word families, with a total of about 120,000 lexical items, and 600 simplex words with no attested derivations or compounds (Vare

2012). The database was compiled manually following a fixed schema of the Dictionary Writing System.

The Word Formation database will be probably most used for research (available on the Web since January, 2014), as well as for language learning and language technology, e.g. to create the independent word formation module as part of the rule-based morphology of Estonian, which already covers the fully regular word formation.

**(3) The Explanatory Dictionary of Estonian** (EKSS 2009, 2<sup>nd</sup> edition, 6 volumes, 150,000 headwords). In the 1960s, the compilation of the first comprehensive dictionary of Estonian was started – roughly 300-200 years later than those of French, English, or German, and 100 years later than in the Nordic countries – and the first fascicle appeared in 1988. Thinking of the Finno-Ugric languages one may say that the publication of a comprehensive explanatory dictionary is highly coincident with statehood, and also with population (ca 1 million for Estonia) (Langemets 2008). Of the eight Finno-Ugric languages that are state languages (Hungarian, Finnish, Estonian, Komi, Udmurt, Mari, Erzyan and Moksha), only three have a comprehensive dictionary: first the Finns and Hungarians (1950s-1960s),<sup>7</sup> and next the Estonians (EKSS 1988-2007; 2<sup>nd</sup>, revised edition EKSS 2009). The EKSS contains about 150,000 words. It is a typical first explanatory dictionary: it abounds in literary examples (ca 50,000) and usage examples devised by lexicographers, the entries are long as many polysemous words (especially verbs) have been subjected to a really thorough meaning differentiation. The dictionary took 50 years to prepare. We were extremely lucky in having the opportunity to produce the dictionary in an electronic form from the very beginning. EKSS has been on the Web since 2010,<sup>8</sup> and on META-SHARE since 2012.

The next higher ‘lexicographic storey’ is under construction, representing an one-volume desktop version of the explanatory Estonian dictionary (ca 100,000 words, to appear in 2018), following the example of several European languages. The focus lies on ‘more theoretical’ lexicosemantic description and usage examples are considerably fewer than in EKSS, while etymological information has been added. Most of the usage examples come from a text corpus, from which they are extracted by the Sketch Engine software (Kilgarriff et al. 2004) and special Estonian Sketch grammar (Kallas 2013). The prospective user is an educated adult native Estonian (Langemets et al. 2010).

Estonian monolingual lexicography of the 20<sup>th</sup> century is represented by two main kinds of dictionary: a descriptive explanatory dictionary (EKSS) and a prescriptive dictionary of Standard Estonian or the language planning dictionary (ÕS).

**(4) The Dictionary of Standard Estonian ÕS 2006** (ÕS 2006). The first language planning dictionary (20,000 headwords), published in revolutionary times, was a contemporary of the Estonian state born in 1918. Now, the new explanatory, one-volume Dictionary of Estonian (to appear in 2018) celebrates the centennium of the Republic of Estonia as well as the line of monolingual lexicography of Estonian.

<sup>7</sup> The Finnish *Nykysuomen sanakirja* (1951-1961), the Hungarian *A magyar nyelv értelmező szótára* (1959-1962).

<sup>8</sup> EKSS 2009: [www.eki.ee/dict/ekss/](http://www.eki.ee/dict/ekss/) (last visited: 28.01.2013).

In 1989 the Law on Language declared Estonian the sole official language of Estonia. According to a Government regulation of 2006, the literary norm should be based on the most recent dictionary of Standard Estonian (then ÕS 2006) issued by the Institute of the Estonian Language – an interesting fact, probably not common to many languages. The new edition of the revised Dictionary of the Standard Estonian (ÕS 2013) is to appear in 2013, thus being the ‘updated’ official norm of the standard language. ÕS 2006 has been on the Web since 2010,<sup>9</sup> and on META-SHARE since 2012, as well as being usable via mobile phone application.

**(5) The Basic Estonian Dictionary** (to appear in 2014). Beside the two above-mentioned mainstreams, the Basic Estonian Dictionary (4,500 headwords) is under compilation (Jürviste et al. 2011). It is a comprehensive print and online dictionary of contemporary Estonian compiled for learners of Estonian as a second or foreign language (A2, B1 proficiency level). In order to support the development of lexical and grammatical competence, the dictionary aims to present explicitly the syntagmatic relations of Estonian content words: substantives, adjectives, adverbs and verbs (Kallas 2013). The Basic Estonian Dictionary is another dictionary compiled using the Sketch Engine software (Estonian Sketch grammar) for extracting relevant syntagmatic collocations. The learners' dictionary will be on the Web from 2015, a year after being published in print. The web version will be prepared in cooperation with Speech Technology, being one of the applications of Speech Synthesis of Estonian.

## 5. Tools for lexicography

**(1) Morphological analysis and synthesis.** Estonian dictionaries have a long tradition of presenting morphological information, as Estonian is characterised by a great number of inflected forms and extensive variation of morphological units. The morphology modules have been successfully used in lexicography for generating the grammar component in dictionary entries (Viks 2000, 2008). The system is semi-automatic, utilising a rule-based module. The module generates the inflectional type, part-of-speech, inflected forms, morphonological marking (degree of quantity, morpheme boundaries, etc.) and morphological references for inflected forms considerably different from headwords (e.g. *peo* SgGEN → *pihk* ‘palm’).

**(2) Syntax and semantics.** Syntactic parsers with a broad coverage of Estonian have been developed only using one rule-based grammatical formalism. Semantic tools and resources are scored low in the White Paper and syntactically and semantically annotated corpora are small in size (Liin et al. 2012, 59). It is true that we lack basic tools in semantics.

We have so far manually assigned semantic class tags (ARTEFACT, HUMAN, etc.) when editing the explanatory one-volume dictionary of Estonian (to appear in 2018, see above) for the purpose of linguistics research (patterns of systematic polysemy, e.g. ACTIVITY-ARTEFACT, MATERIAL/SUBSTANCE-ARTEFACT, etc.) (Langemets 2010). In the future we hope to link a semantic inventory/dictionary to concordance lines (cf. super senses, Word Sketch).

<sup>9</sup> ÕS 2006: [www.eki.ee/dict/qs/](http://www.eki.ee/dict/qs/) (last visited: 28.01.2013).

**(3) Estonian Reference Corpus and Corpus Query tools.** Since 2010 the Sketch Engine for Estonian (Kallas 2013) has been used at the Institute of the Estonian Language to compile two monolingual dictionaries: the explanatory one-volume dictionary of Estonian and the Basic Estonian Dictionary (see above). The Sketch Engine for Estonian uses the Estonian Reference Corpus<sup>10</sup> of 250 million tokens as input. The corpus had previously been annotated morphologically, lemmatised, partially disambiguated, and annotated by clause by the private company FiloSoft LLC. Syntagmatic relations of content words are described in the Grammar for Estonian (Kallas 2013) as lexico-grammatical constructions (ca. 85 rules or grammatical relations) defined by means of morphosyntactic categories (phrase type, part of speech, inflectional categories). Figure 3 shows the Word Sketch for the noun *diskussioon* ‘discussion’. We also use the Good Examples option in the Sketch Engine software, though it is not fully automatic yet. In the future we expect to detect novel senses (by the means of the Sketch Engine), as well as to find good examples of a word sense.

diskussioon ( ) EstonianRC freq = 5934 (23.8 per million)			
<b>Adj_modifier</b> 2206 3.9	<b>subject_of</b> 400 1.6	<b>object_of</b> 216 3.0	<b>omastav_modifies</b> 821 1.2
avalik 240 7.05	käima 57 4.37	alustama 33 4.65	objekt 93 6.2
poliitiline 110 5.83	tekkima 36 4.36	tekitama 23 4.47	teema 71 4.51
elav 87 6.86	toimuma 33 3.77	algatama 18 5.08	tulemus 46 2.91
tõsine 77 6.21	jätkuma 32 5.62	jätkama 14 4.09	küsimus 29 1.36
pikk 72 5.13	algama 26 4.31	pidama 12 0.61	tekitamine 20 5.82
äge 66 8.15	kestma 17 4.37	ärgitama 11 8.45	algatamine 20 5.6
sisuline 49 7.31	järgnema 13 5.29	avama 8 3.11	algus 12 0.73
ühiskondlik 44 6.9	minema 11 1.03	laskuma 6 6.9	arendamine 10 2.8
terav 35 6.63	puhkema 10 5.82	korraldama 6 2.37	tase 10 1.31
käiv 32 7.3	näitama 8 1.6	vallandama 5 5.64	alune 9 0.19
<b>N_PP</b> 117	<b>saav_modifier</b> 172 1.9	<b>alalütlev_modifier</b> 137 1.6	<b>seesütlev_modifier</b> 115 1.2
N_PP_üle 77 25.0	objekt 57 5.51	teema 91 4.88	ajakirjandus 11 2.52
N_PP_ümbär 13 20.7	tekitamine 12 5.17	tulemus 11 0.85	komisjon 8 0.18
N_PP_käigus 8 11.1	alune 10 0.35	taust 4 1.71	ühiskond 7 1.3
<b>predicate_Adj</b> 21 1.4	<b>participle_modifier</b> 292 3.2	<b>ja/või</b> 166 1.0	
elav 3 2.13	toimuma 55 4.51	arutelu 21 4.04	
kahepoolne 2 4.23	puhkema 28 7.33	ettekanne 14 3.95	
mõttetu 2 3.8	kestma 20 4.61	seminar 10 4.4	
loomulik 2 2.6	tekkima 16 3.19	dialoog 7 4.34	
vajalik 2 0.21	avama 13 3.81	loeng 6 3.99	

Figure 3: Word Sketch of the noun *diskussioon* ‘discussion’ in the Estonian Reference Corpus (Kallas 2013, 121)

## 6. What is the best model for compiling bilingual dictionaries?

Arvi Tavast (2008) has been arguing for an onomasiological model for compiling better, i.e. more systematic and consistent dictionaries for some time already. He has called his view *instrumentalist* (opposed to *linguacentric*) to be summarised in the statement that “language is just a tool” (Tavast 2008, 11, cf. Fuertes-Olivera/Bergenholtz 2011). Tavast

<sup>10</sup> Estonian Reference Corpus: [www.cl.ut.ee/korpused/segakorpus/](http://www.cl.ut.ee/korpused/segakorpus/) (last visited: 28.01.2013).

argues that an instrumental view results in better decisions, both in terminography as well as general (bilingual) lexicography. Since 2012 the Institute of the Estonian Language has been participating in the EU project for *Development of the Estonian-Latvian and Latvian-Estonian dictionary* (2012-2015), together with the Latvian Language Agency.

The role of the Institute (as the leading participant) was to adapt the Dictionary Writing System EELEX for producing two dictionaries (Estonian-Latvian and Latvian-Estonian). We decided to compile these dictionaries in a totally new way – in a *technically onomasiological*, i.e. in a *concept-based* database. It has been proved by Tavast (2008, 64) that this is the second best method (after systematic terminology work, unfortunately impossible in many cases) to prevent numerous internal inconsistencies unavoidable when compiling otherwise semasiologically, i.e. in a word-based database. The System is supplied with a mechanism for the conversion the concept-based database into a traditional word-based dictionary. Thus, the two versions of the bilingual dictionaries (Estonian-Latvian, as well as Latvian-Estonian) will be created fully automatically from the same database. This is slightly reminiscent of the Dutch system OMBI using *semantic links* in the 1990-ies for creating bilingual dictionaries from the multifunctional and reusable electronic lexical databases (Martin/Tamm 1996; Martin 2007).

The root data element of the concept-based database is a concept (meaning/sense description), not a term (word). The terms (source words and target words) are associated with the concept (meaning) as well as with definition (in both languages, if needed) and context. Both languages have the same status, i.e. during the conversion into the (paper) dictionary format, both of them may be selected as a source language. Basically everything is related to the concept, e.g. synonyms are together in one concept – or they are not synonyms at all (within the precise definition of the dictionary) (Tavast 2008, 58 and 66-67).

E.g. the concept with the definition ‘eye doctor/ophthalmologist’ may contain the terms *silmaarst*, *okulist*, *oftalmoloog*, *oftalmologs*, *acu ārstis*. The mechanism for converting defines how the data is presented to the user depends on editors' decisions. One might choose between different options: (a) to show all the source words as equal in every article, (b) to show source words as separate headwords, together with the same definition and same target words, or (c) to assign one source word the role of a primary term which is connected to others via references (vt ‘see’) (Figure 4). A database can be converted into the (Estonian-Latvian or vice versa) dictionary in a fully automatic manner.

(a) Estonian-Latvian:

**silmaarst (okulist, oftalmoloog)** – acu ārstis, oftalmologs

**okulist (silmaarst, oftalmoloog)** – acu ārstis, oftalmologs

**oftalmoloog (silmaarst, okulist)** – acu ārstis, oftalmologs

(b) Estonian-Latvian:

**silmaarst** (‘eye doctor/ophthalmologist’) – acu ārstis, oftalmologs

**okulist** (‘eye doctor/ophthalmologist’) – acu ārstis, oftalmologs

**oftalmoloog** (‘eye doctor/ophthalmologist’) – acu ārstis, oftalmologs

(c) Estonian-Latvian:

**silmaarst (okulist, oftalmoloog)** – acu ārstis, oftalmologs

**okulist** – vt silmaarst

**oftalmoloog** – vt silmaarst

Figure 4: Converting of the concept-based database into the word-based dictionary



Tavast (2008, 16) has alleged that, unfortunately, people do not distinguish between words and their meanings, or terms and concepts. One should achieve a small, but crucial change in thinking to get used to working with such a database. Neither non-experts nor lexicographers feel it an easy task.

The onomasiological database for compiling bilingual dictionaries is a free software TERMEKI<sup>11</sup> offered by the Institute.

## **7. To sum up: online dictionaries of Estonian**

Finally, I will give a short overview of typical properties of online dictionaries of Estonian, following criteria and oppositions described in Lew (2011).

The overall policy for the Institute of the Estonian Language, supported by The Plan (2011) and Ministry of Education and Research, has been to offer all our dictionaries free of charge when published. Or, to be precise, one year after publishing in print. There are different types of dictionaries: general and specialised, monolingual and bilingual, descriptive and prescriptive, etc. We have not set up any open, user compiled dictionaries. The feedback from users is sent by mail to the lexicographers via the web page.

Like most traditional dictionaries they exist standalone, however, they are gathered more or less on the single web page.<sup>12</sup> There have been some attempts to create Dictionary portals,<sup>13</sup> but they do not seem very user-friendly so far.

Regarding the access to content the dictionaries are mostly presented like paper dictionaries in an electronic form, fulfilling the goal of searchableness, but nothing more. Usually all the content is displayed: no ‘Show Less’/‘Show More’ options or other customisation by users are available. The step-wise access with content displayed partially is used only for searching an extremely sophisticated database of Word Formation – the Dictionary of Word Families (Vare 2012) – where the user gets the remaining parts of the hierarchical entry by clicking the ‘+’ icons.<sup>14</sup>

Multimedia is yet to come in the future: illustrations and audio recordings will be available when the learners' dictionary – Basic Estonian Dictionary (to appear in 2014) – will be published online.

And last but not least, p-lex or e-lex? The answer is: definitely e-lex, but so far mostly p-dictionaries on the web. We are still largely constrained by the paper model, thus having rather a lot to achieve to come ‘to think outside the paper’.

---

<sup>11</sup> TERMEKI: <http://term.eki.ee/> (last visited: 28.01.2013).

<sup>12</sup> Lexical resources of the Institute of the Estonian Language: <http://portaal.eki.ee/keelekogud.html> (last visited: 28.01.2013).

<sup>13</sup> Keeleveeb: [www.keeleveeb.ee/](http://www.keeleveeb.ee/) (last visited: 28.01.2013).

<sup>14</sup> The Database of Word Families will be publicly available since January 2014.

## 8. References

- EKSS 1988-2007 = Karelson, R./Kullus (Põlma), V./Raiet, E./Tiits, M./Valdre, T./Veski, L. (eds.) (1988-2007): *Eesti kirjakeele seletussõnaraamat 1988-2007* (vihikud 1-26). Eesti Keele Instituut. [The Explanatory Dictionary of Standard Estonian. 26 fascicles.] Tallinn: Eesti Keele Sihtasutus.
- EKSS 2009 = Langemets, M./Tiits, M./Valdre, T./Veski, L./Viks, Ü./Voll, P. (eds.) (2009): *Eesti keele seletav sõnaraamat I-VI*. ("Eesti kirjakeele seletussõnaraamatu" 2., täiendatud ja parandatud trükk). Eesti Keele Instituut. [The Explanatory Dictionary of Estonian, 2nd, improved edition.] Tallinn: Eesti Keele Sihtasutus.
- Fuertes-Olivera, P./Bergenholtz, H. (eds.) (2011): *e-Lexicography: The Internet, digital initiatives and lexicography*. London: Continuum.
- Jürviste, M./Kallas, J./Langemets, M./Tuulik, M./Viks, Ü. (2011): Extending the functions of the EELEX dictionary writing system using the example of the Basic Estonian Dictionary. In: *Electronic Lexicography in the 21st Century: New Applications for New Users: Proceedings of eLex 2011, 10-12 November 2011, Bled, Slovenia*. Ljubljana: Trojina, Institute for Applied Slovenian Studies, 106-112.
- Kallas, J. (2013): *Eesti keele sisusõnade süntagmaatilised suhted korpus- ja õppeleksikograafias*. Tallinna Ülikool, Doktoritöö. [The Syntagmatic Relationships of Estonian Content Words in Corpus and Pedagogical Lexicography.] Tallinn: Tallinna Ülikooli Kirjastus.
- Kallas, J./Langemets, M. (2012): Automatic generation of specialized dictionaries using the Dictionary Writing System EELEX. In: Tavast, A./Muischnek, K./Koit, M. (eds.): *Human Language Technologies – The Baltic perspective*. (= Frontiers in Artificial Intelligence and Applications 247). Amsterdam: IOS Press, 103-110.
- Kilgarriff, A./Rychly, P./Smrž, P./Tugwell, D. (2004): The Sketch Engine. In: Williams, G./Vessier, S. (eds.): *Euralex 2004. Proceedings of the Eleventh Euralex International Congress*. Lorient: Université de Bretagne-Sud, 105-116.
- Langemets, M. (2010): *Nimisõna süstemaatiline poliüseemia eesti keeles ja selle esitus eesti keelevaras*. Eesti Keele Instituut. [Systematic polysemy of nouns and its lexicographic treatment in Estonian.] Tallinn: Eesti Keele Sihtasutus.
- Langemets, M./Loopmann, A./Viks, Ü. (2010): Dictionary management system for bilingual dictionaries. In: Granger, S./Paquot, M. (eds.): *eLexicography in the 21st century: new challenges, new applications. Proceedings of eLex 2009*. (= Cahiers du CENTAL 7). Louvain-la-Neuve: Presses universitaires de Louvain, 425-430.
- Langemets, M./Tiits, M./Valdre, T./Voll, P. (2010): *In spe: ühekõiteline eesti keele sõnaraamat*. In: *Keel ja Kirjandus* 11, 793-810.
- Lew, R. (2011): Online dictionaries of English. In: Fuertes-Olivera, P./Bergenholtz, H. (eds.): *e-Lexicography: The Internet, digital initiatives and lexicography*. London: Continuum, 230-250.
- Liin, K./Muischnek, K./Müürisep, K./Vider, K. (2012): *The Estonian language in the Digital Age/Eesti keel digiajastul*. Edited by Georg Rehm and Hans Uszkoreit. (= White Paper Series). Berlin: Springer.
- Martin, W. (2007): Government policy and the planning and production of bilingual dictionaries: the "Dutch" approach as a case in point. In: *International Journal of Lexicography* 20.3, 221-238.

- Martin, W./Tamm, A. (1996): OMBI: an editor for Constructing Reversible Lexical Databases. In: Gellerstam, M., et al. (eds.): *Euralex '96. Proceedings I-II*. Göteborg: Göteborg University, Department of Swedish, 675-685.
- ÕS 2006 = Ereht, T. (ed.) (2006): *Eesti õigekeelsussõnaraamat ÕS 2006*. Eesti Keele Instituut. [The Dictionary of Standard Estonian ÕS 2006]. Tallinn: Eesti Keele Sihtasutus.
- Tavast, A. (2008): *The translator is human too: A case for instrumentalism in multilingual specialised communication*. Tartu Ülikool. PhD dissertation. Tartu: Tartu University Press.
- The Plan 2011 = *The Development Plan of the Estonian Language 2011-2017*. Tallinn: Eesti Keele Sihtasutus.
- Vare, S. (2012): *Eesti keele sõnapered I-II*. Eesti Keele Instituut. [The Dictionary of Estonian Word Families 1-2.] Tallinn: Eesti Keele Sihtasutus. [Index on the Internet: <http://portaal.eki.ee/dict/spi/> (last visited: 28.01.2013)].
- Viks, Ü. (2000): Tools for the generation of morphological entries in dictionaries. In: *Proceedings of the Second International Conference on Language Resources and Evaluation. Athens, Greece, 31 May-2 June 2000*, 383-388.
- Viks, Ü. (2008): Eesti-X-keele sõnastik ja grammatika. In: *Eesti Rakenduslingvistika Ühingu aastaraamat 4*, 247-261.
- Viks, Ü./Vare, S./Sahkai, H. (2010): The Database of Estonian Word Families: a Language Technology Resource. In: *Human Language Technologies. The Baltic Perspective: Proceedings of the Fourth International Conference, Baltic HLT 2010*. Amsterdam: IOS Press, 169-176.

