Dictionary Management System for Bilingual Dictionaries

Margit Langemets¹, Andres Loopmann², Ülle Viks³
Institute of the Estonian Language (Tallinn, Estonia)

Abstract
Our project focuses on a web-based dictionary management system for bilingual dictionaries. The system is designed for different users (lexicographers, language learners, translators, etc.) who desire to compile their own bilingual (primarily Estonian–other) dictionary via the web. The system enables users to design their own dictionary as they need it, choosing the languages, modifying the entry layout and structure, and using a ready-made description of the source language. The main components of the management system are: (1) EELex system of dictionary administration, (2) an Estonian–X dictionary database, and (3) the management system interface.

Keywords: web-based dictionary management, bilingual dictionaries, automatic morphology.

1. Introduction
The open world of today experiences a growing need for translation dictionaries. Our dictionary management system for bilingual dictionaries is meant to facilitate lexicographer’s work and, at the same time, enhance its quality. We believe that dictionary-makers should be relieved from technical problems, so that they could focus on the most important task of entry compilation. Our aim is a web-based universal dictionary writing system for compilation, editing and publication of a bilingual dictionary, with an interface enabling users to format the dictionary as they like it, choosing the source and target languages, and modifying the entry layout and structure. An additional asset is an Estonian-X dictionary database and morphological synthesis of Estonian wordforms.

2. EELex system of dictionary administration
The EELex (= EELex) system of dictionary administration is a web-based lexicographer’s workbench integrating various language technological tools: linguistic software and language resources (see Langemets et al. 2006). The main features of

¹ Institute of the Estonian Language, margit@eki.ee
² Institute of the Estonian Language, andres.loopmann@eki.ee
³ Institute of the Estonian Language, ylle@eki.ee
EELex are: Unicode support, XML databases and schemas, XSL transformations for generating different views (XML view, Edit view, Layout view), click-to-edit, structural queries and sorting of query results, export to the MS Word layout format, team work option (with different levels of user rights), various tools for entry and dictionary editing, e.g. menu compiler, XML file generator, etc. (cf. Joffe et al. 2008, Mangeot 2006).

The EELex system of dictionary administration has been created at the Institute of the Estonian Language under the project called “Lexicographer’s workbench”. At present, the system involves about twenty dictionaries, of which five have been completed (published or finished), eleven are being edited, and two existing dictionaries (in another format) are being prepared for transferring to an EELex format.

Figure 1 shows the EELex editing window (language option of the user’s interface: English).

![Figure 1. Editing window](image)

### 3. Estonian–X dictionary database

The main EELex application is the Estonian–X dictionary database (= EXDD), which is compiled by means of EELex to provide a core for the new bilingual dictionaries to be produced by the system. The EXDD contains only source language (Estonian) data: entry word, grammatical information, explanations, labels, usage examples, compound words, etc. The data of the target language (X) – translation equivalents with the rest
of the necessary information – will be supplied by the user compiling the new
dictionary.

The EXDD source material comes from a voluminous (c. 80,000 entries) Estonian–
Russian dictionary (Liiv et al. 1997-2009), plus some material from other dictionaries.
The presentation of the material has been tailored for the EXDD, while an essential
part of the entries (c. 40,000) have been subjected to a detailed editing process, using a
standard presentation for sense division and homonyms, as well as for compound
words (headword or example), cross-referencing, labelling (usage information,
domains), etc. According to the type of dictionary required, three standards of
morphological description have been developed to facilitate the presentation of
Estonian morphology for non-native users who might otherwise be taken aback by the
great number of inflected forms and extensive variation of morphological units.

Figure 2 presents an example (the entry kuld ‘gold’) from the EXDD in layout format.
Note that 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.

(1) kuld <k’uld kulla k’ulda, k’ulda[de k’ulda[sid _& k’uldi/i 22 S>
1 (teatud väärismetall ‘a yellow precious metal’) |TE| ♦puhas kuld |TE|; ehtekuld |TE|;
kullast ehted |TE|; kulda pesema |TE|; lõpetasin keskkooli kallaga könek |TE|
2 (värvuselt ja läikelt kulla sarnane ‘colour resembling gold’) |TE| ♦päikesekuld |TE|;
sügiskuld |TE|; kased paistavad juba kuld |TE|
3 (midagi väärtuslikku ja head ‘sth highly respected’) |TE| ♦ tema nõuanded on kuld väärt
|TE|; nendel sõnadel on kulla kaal v hind |TE|

(2) ■Ls: kuld+ (kullast, kullatud ‘made of gold’) ♦ kuldbrokaat tekst |TE|; kuldmedal |TE|;
kudmmünt |TE|; kuldsmõrmus |TE|; kuldvillak miit |TE|
■Ls: kuld+ (kulla värv ‘lustrous yellow’) ♦ kuldblond |TE|; kuldjuukseline |TE|;
kuldkollane |TE|; kuldprõnikas zool (Cetonia aurata) |TE|
■Ls: kulla+ ♦ kullaauk könek, piltil |TE|; kullafond (1) maj |TE|, (2) piltil |TE|; kullakang
|TE|; kullakeetraja |TE|; kullaliiv |TE|; kullalüüge |TE|; kullamõed piltil |TE|; kullotsija |TE|;
kullapalavik piltil |TE|; kullaproov |TE|; kullasoon |TE|; kullastandard maj |TE|; kullatera
|TE|

Figure 2. The dictionary entry kuld (‘gold’)

Section (1) includes the main part of the entry: the headword with a full morphological
description (all basic inflectional forms, the inflectional type number, part of speech).
The numbers differentiate between senses (and explanations), while each sense is
followed by usage examples. Section (2) contains a separate block of compounds
associated with the headword, classified according to the grammatical form and
meaning of the first component. Throughout the entry, |TE| signals a translation
equivalent to be supplied by the user. In addition, the entry provides structured space
for information pertaining to the translation equivalent: grammatical information,
labels, explanations, etc.

Although the Estonian–X dictionary database (EXDD) is still being improved and
updated, the system is used in-house at present for compiling three bilingual
4. Dictionary management system for bilingual dictionaries

The dictionary management system for bilingual dictionaries uses the same software as the EELEX dictionary administration system. The management system is based on a standard XML schema following the structure of a typical bilingual dictionary, and on a standard dictionary layout.

The user interface enables users to create their own dictionary and to adjust the system to their own needs. For every dictionary application appropriate parameters can be selected in the following four domains:

(a) selection of source and target languages accompanied by automatic keyboard switching and a spelling checker option available during the editing process;

(b) layout design: the user can decide upon the style of the elements as well as the markers of the elements or element groups;

(c) the morphological interface enables automatic generation of a morphological description of the Estonian headword: inflected forms, indexes of part of speech and inflectional paradigm; the rule-based morphological system will generate a morphological description for unknown words as well (Viks 2000);

(d) a future option enables modification of entry structure by addition, deletion or rearrangement of its elements.

Figure 3 presents an example of the window for creating a user’s dictionary.

Figure 3. Dictionary creating window

A public version of the Dictionary management system for bilingual dictionaries will be released as freeware (http://exsa.eki.ee).
5. Conclusion

The dictionary management system for bilingual dictionaries enables the user, through the user interface, to design their own dictionary in a web environment, choosing the languages, modifying the entry layout and structure, and using, upon request, a ready-made description of the source language. This will economize on compilation time and improve the output quality as the resulting dictionaries represent universal re-usable language resources in a standard format.

Acknowledgements

This work was supported by the National Programme for Estonian Language Technology (2006–2010) and by the project SF0050023s09 (“Modelling intermodular phenomena in Estonian”).

References


