

Simple and Effective User Interface for the Dictionary Writing System

Kamil Barbierik, Zuzana Děngeová, Martina Holcová Habrová, Vladimír Jarý,
Tomáš Liška, Michaela Lišková, Miroslav Virius
Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i.
kamil.barbierik@fffi.cvut.cz, dengeova@ujc.cas.cz, holcova@ujc.cas.cz,
vladimir.jary@foxcom.cz, tomas.liska@foxcom.cz,
liskova@ujc.cas.cz, miroslav.virius@fffi.cvut.cz

Abstract

A new monolingual dictionary of the contemporary Czech language is being prepared by the Institute of Czech language at the Academy of Sciences of the Czech Republic, v.v.i. A dictionary writing software is being developed as part of a grant supported by the Ministry of Culture of the Czech Republic within the National and Cultural Identity (NAKI) applied research program. We will present the overall architecture of the software and then focus on its user interface and two modules: the referencing system and a new module – the editorial tool (that was promised in (Barbierik 2013)).

Keywords: dictionary writing system; DWS; cross-reference module; editorial tool; lexicography

1 Introduction

Since 2012, the Department of Contemporary Lexicology and Lexicography within the Institute of Czech Language at the Academy of Sciences of the Czech Republic, v. v. i., has been preparing a new monolingual dictionary of contemporary Czech.

Its working title is *Akademický slovník současné češtiny* (The Academic Dictionary of Contemporary Czech). It is a medium-sized dictionary with an expected number of 120,000–150,000 lexical units.

To aid this project, a new Dictionary Writing System (DWS) is developed. More information about the project can be found in (Kochová 2014). A detailed specification of the requirements from the lexicographer's point of view can be found in the article *A New Path to a Modern Monolingual Dictionary of Contemporary Czech: the Structure of Data in the New Dictionary Writing System* (Barbierik 2013).

We introduce basic functionality of our DWS with emphasis on the user interface in this paper. Further, we focus mainly on the editorial tool which will be described in more detail.

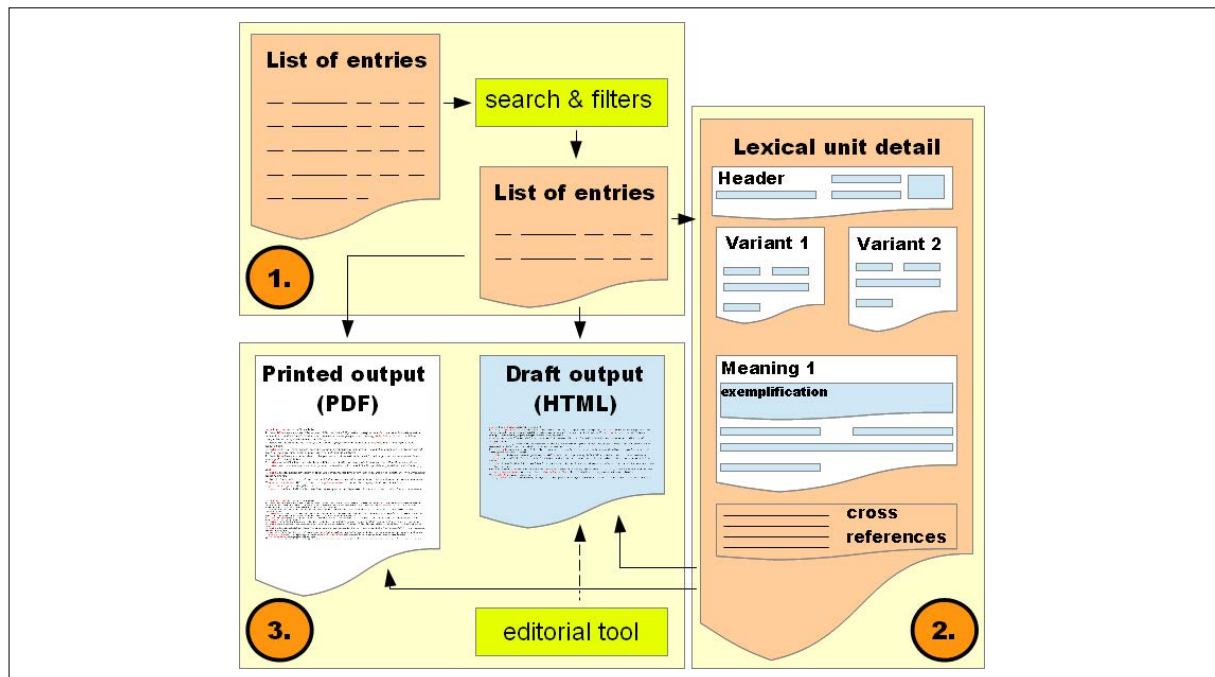


Figure 1: Basic scheme of our Dictionary Writing System.

2 Existing Dictionary Writing Systems

There are several commercial Dictionary Writing Systems (DWS) available (e.g. TshwaneLex (2013), IDM DPS (2013),

iLEX (2013)) as well as open-source systems (e.g. the Matapuna Dictionary Writing System (2013)). The DEB II (2013),

(DEBDict 2013), dictionary editor and browser, is available for the Czech language.

The lexicographic team at the Institute of the Czech Language of the Academy of Sciences of the CR, v. v. i., considered three options: to buy an existing commercial DWS, to use one of the open-source systems or to develop their own system. One significant criterion used for the DWS selection was the amount of necessary adjustments due to specifics of the compilation of the dictionary and the time allocated to this task. Another criterion was the DWS price. After evaluation of the DWSs available, we decided to develop our own DWS that will fully respect the significant specifics of the compilation of the dictionary (Abel 2012: 1-23; Atkins 2008). The greatest advantage of this decision, which the lexicographic team benefits from, is the fact that any request for the user interface, some process modification or a new handy feature implementation etc., can be processed and implemented almost immediately.

3 Basic Functionality of Our DWS

From the common user point of view, the software is divided into three parts as it is shown on Fig. 1: the list of entries, the lexical unit detail and the output. The numbers of these main parts are used in the titles of the following subsections.

3.1 List of Entries (part 1)

After successfully logging into the system, the user inputs the list of entries which mainly represents the macrostructure of the dictionary. It is a list of lexical units with some basic information that is important for linguists at this stage.

Some helpful functions are available for better orientation in the long list of entries. The most important one is probably the quick search engine together with a set of predefined filters. The quick search allows users to search for entries by selecting any field of the microstructure of lexical unit (through which the user intends to search) and entering a search query. The search query can contain wildcards granting users better control over the search results. In addition, the query field is automatically updating its mode according to the type of information the user is searching for to make the search process easier for the user. For instance, if the user is trying to find information in fields where only a few values are available (e.g. type of lemma), the query field updates itself to select box. Thus, the user does not have to guess what values are available within the selected field. To avoid typing errors, an auto-complete function is implemented when searching in fields that contain short texts (like lemma).

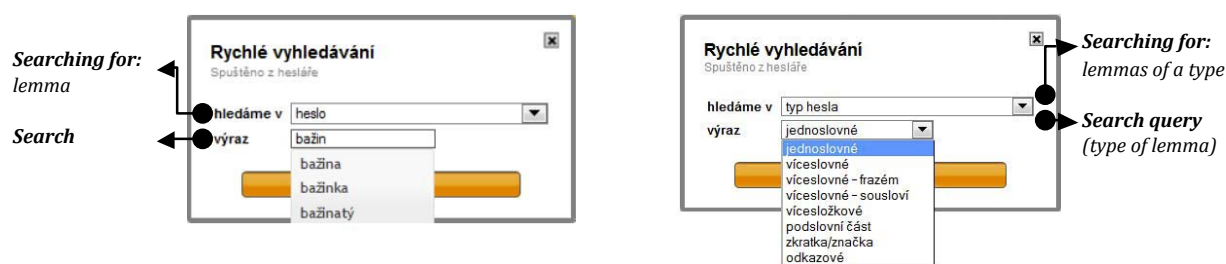


Figure 2: The “Quick search” function with auto-complete (left) when searching for lemma and select box when searching lemmas of certain type (right).

For example, the user is able to filter out, with combination of the quick search with predefined filters, the following entries:

- entries which I (the logged in user) founded and begin with the defined letter,
- entries of a specified word class created in some time interval,
- entries from manual selection containing some phrase in any of its exemplification, etc.

3.2 Detailed View of The Lexical Unit – Editing Module (part 2)

When the lexicographer finds the entry, he or she may continue to the detailed view of it. The detailed view of the entry contains all the information mainly from microstructure of the lexical unit that is available in a well-arranged and sophisticatedly structured way. Additionally, the user can edit any information required in this view. To make the editing process more effective, different input fields are used according to the type of information it needs to gather. The whole editing form is designed so that the linguists editing the lexical unit information do not need to learn any special markup language or have any advanced computer literacy skills.

This editing form is organized into 4 sections (see Fig. 1 – Part: Lexical unit detail):

- (1) Header
- (2) Section of variants
- (3) Meanings
- (4) Cross-references

Header section. General information about the lexical unit can be found in the header section. It contains the entry status indicator which shows the progress of the work on this entry. Also, the output status can be set in this section which indicates in which output (electronic or paper) the entry will be presented. Furthermore, it contains information about the time of creation and about the last editing of the entry. The header section also contains information on the responsible user as well as a field very where the lexicographers may leave a note concerning the entry.

Section of variants. The section of variants may contain one or more variants of the lemma with all required microstructure elements. The variants of variant lemmas are often equivalent in majority of values, thus the function “Add variant as a copy of the last one” was implemented. This function creates a new variant and copies all the values from the last existing variant to it. Thus, only a few values have to be edited in the new variant. Consequently, creating new similar variants is much more efficient.

Section of meanings. The section of meanings consists of one or more panels, where the meaning of the word is described together with other related information. The section is organized as a set of panels. Each panel contains a large form, where the information about the meaning can be edited. The user can change the order of the panels; this will affect the order in the dictionary printed or electronic output. Meanings are numbered and when reordered, the numbering of meanings is automatically updated. The panel containing the form, with information related to the meaning, can be minimized or maximized according to which panel the user intends to work with. It helps for better orientation, whereas some lemmas may have quite a lot of meanings. The quick navigation is also helpful when a word has a lot of meanings.

This allows navigating directly to the meaning with the certain number, without scrolling the page.

Implementation of the editing form. As we mentioned at the beginning of this section, the whole detailed view is basically a well arranged set of fields of different types. These field types were chosen according to the types of information contained in lexical unit microstructure. For the gathering of short textual information (mostly comments, but also pronunciation for instance, synonyms, etc.) we have used simple one row text input field. Multiline longer texts are collected using textboxes. Often it is necessary to format the input text in some way. Special textboxes with rich text functions were implemented for this purpose. Such fields are used to store the exemplifications or the meaning explanations. Probably the most complex input fields are administrated select boxes which provide lexicographers with finite number of options prepared by the administrative user. This prevents editors from committing typing errors and unifies the values in certain places in microstructure through the whole dictionary. But it does not limit them thanks to the option of adding their own entry if it is necessary. The statistics of these entries are collected, and if some value is used too frequently, the administrative user may integrate it to the select box and “standardize it” very easily. Due to limited range we cannot provide an adequately descriptive picture of the editing form. For more information about the editing form, please refer to our poster “Simple and effective user interface of our new DWS”.

3.3 The Output Module (part 3)

It is possible to evoke the output view of one or more lexical units from the detail of the lexical unit as well as from the list of entries. The output module takes the information collected using the editing form described above and utilizes some complex and very strict formatting rules on them to form the output. Thus, the user has a great possibility to preview the entry (or more entries at once) in its printed form and to see how it will exactly look like in the printed dictionary.

Two outputs are available in our DWS system: printed and electronic output. The printed output is not editable and it is presented in PDF format ready to be printed on the paper. The electronic or draft output is presented in HTML form. Even this output is not editable, but it is possible to implement additional interactive functions for it. Thanks to HTML format the user can interact with it using a web browser. One of the interesting functions we designed and implemented in this output is an editorial tool. It allows the lexicographers to fine tune the output or to correct mistakes or inconsistencies in cooperation with other lexicographers or editors.

4 Recently Implemented Features

4.1 Cross-References Module

A very necessary feature of the system, especially from the linguistic point of view, is the ability to define relations between entries. Relations are defined between two dictionary entries; one of the entries is considered to be the main or “master” entry, the second is the “slave”. The system always allows creating the connection from both sides. This means that the user may define the relation if he is editing the slave as well as the master entry.

There are different types of relations from the linguistic point of view: run-on entries, references between one-word and multi-word lexical units and linked entries. From the user point of view, each type of relation needs a slightly different approach, but from the system perspective it is always just a relation between two entries supplemented with some information that is important from the linguistic point of view. When the user is at the detailed view of the lexical unit, he can always define all available types of cross-references to another entry. The window of the referencing module is evoked by clicking on the buttons at certain sections of the editing form. These buttons are placed according to the element of the microstructure from which the user references the other entry. For instance, run-on entry may be referenced from the whole entry (the button is at the end of the form) or from any of its meanings or exemplifications (buttons are under the corresponding input fields). The referenced units are then displayed at correct places according to this information when the output is compiled.

Other type of referenced entry is the linked entry. It can be referenced from the whole lexical unit or from the particular meaning of the entry to other foreign entry or its meaning. Thus, the button for bringing up the referencing tool popup is always at the end of the editing form.

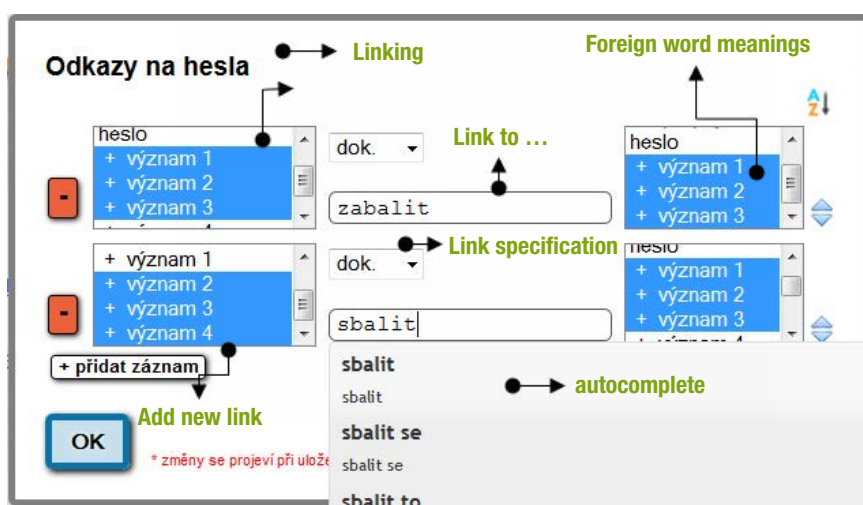


Figure 3: The cross-reference dialog box for linking words.

The interface for referencing is very simple and contains some clever functions to help the user to reference and to manage entries effectively.

Fig. 3 above is a snapshot of popup of referencing module evoked from lemma “balit” and it allows creating references (of type linked entry) to foreign entries and their meanings. As can be seen, references to multiple entries may be defined at once. The reference is created by putting the desired referenced lemma in the middle text input field. These inputs are provided by auto-complete functionality to make the process easier for the user. After writing in the foreign lemma, all its numbered meanings are loaded to the right select-box. Thus, the user knows how many meanings the foreign entry contains and he can comfortably choose the desired ones. Additional information to each reference can be added using a select box. When more references are defined (two in our case), it is reasonable to have an ability to sort the entries. It is possible to do it manually using little arrows next to each referenced entry, or to sort it alphabetically by the program using the AZ button in the top right corner. Links, if any, are according to the formatting rules for creating the monolingual dictionary attached to the end of each entry and our example will produce the output shown in Fig. 4 at the end of the word “balit” definition in the printed output.

► dok. k 1 - 3 → **zabalit**1 - 3, ► dok. k 2 - 4 → **sbalit**1 - 3

Figure 4: The printed output of linked words to the word “balit”.

Editorial Tool

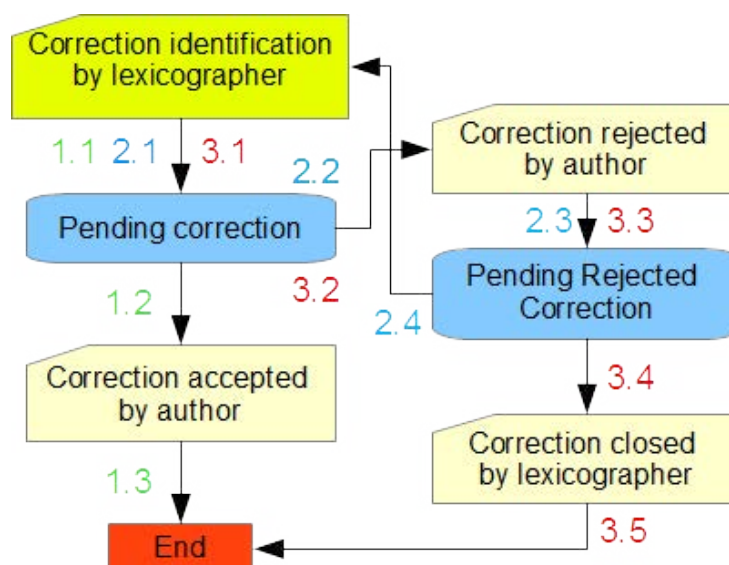


Figure 5: The editorial process.

In order to produce and maintain a high quality dictionary, our DWS system implements an editorial tool, which is, as mentioned above, connected to the electronic HTML output. This editorial tool has been projected to replace the standard editorial process, where some portion of the submitted entries is printed on the paper, sent to the other lexicographers or editors, and received back reviewed. By implementing this feature directly in the DWS system, we save a vast amount of paper, as well as the time and money for the transaction agenda. The whole dictionary, each and every entry of it, is always ready to be reviewed without printing or posting anything.

The editorial process on one entry is captured schematically on Fig. 5. When the author (lexicographer) of the entry submits it to the system, the reviewer is able to see it and to review it using the draft (electronic) output. This draft output is very similar to the final printed output, so he or she is revising the entry almost as it was printed on the paper.



Figure 6: The correction founding and sending it to editors.

By clicking on the information in the draft output, that the lexicographer wants to correct, he or she gets a popup window – see Fig. 6, where he or she can input his suggestion for correction, make a note about the correction for the author and send it to the author with a single click.

This is how the correction identification happens. There is a pending correction from this moment. This is indicated to the author of the entry (and not only him or her, but to other signed in users too) by highlighting the field that contains the corrected information – see Fig. 7.

By the click on the yellow “correction icon” nearby the highlighted field the popup will appear with the suggested correction from the lexicographer.

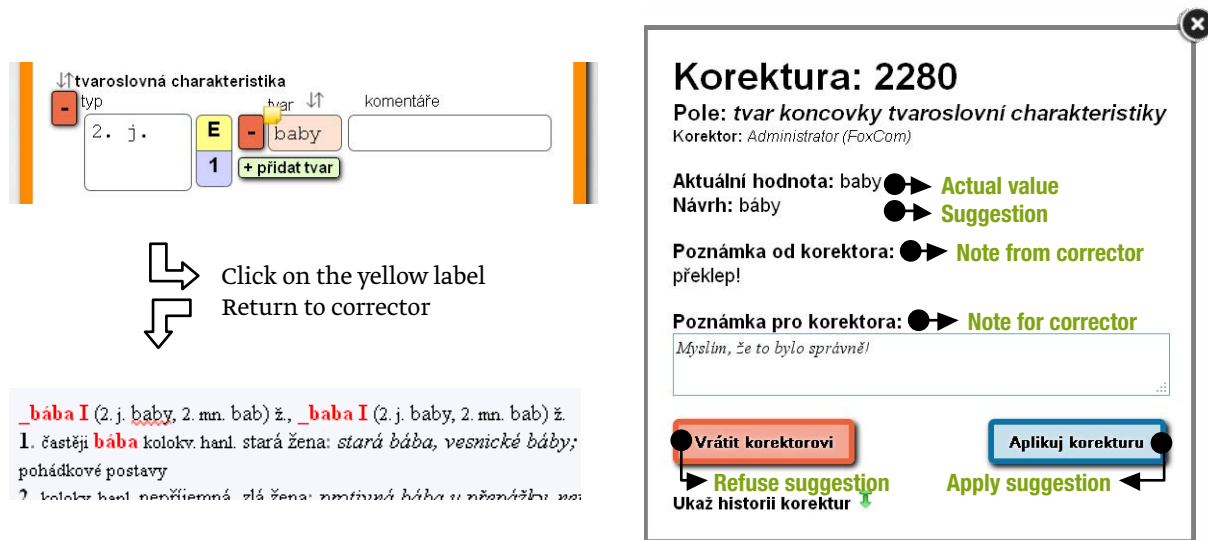


Figure 7: The correction from the author's point of view.

The author may now decide whether he accepts the correction (take the green path number 1 in scheme on Fig. 5), or reject it (the blue path number 2 or the red path number 3 on Fig. 5). In the case he accepts the pending suggested correction, the system automatically updates the entry and the process successfully ends.

If author rejects the pending correction suggested by the (other) lexicographer – the corrector, it will be indicated in the draft (electronic) output – see Fig. 8. By clicking on it, the (other) lexicographer may view it together with the author's note on why it was rejected. At this stage the lexicographer who suggested the correction has two options: to close the correction (following the red path number 3 on Fig. 5) or to suggest a new one (following the blue path number 2 on Fig. 5). If the correction is closed, no changes in the entry are made and the process ends. If a new correction is designed, the process is started again.



Figure 8: The correction refused by editor.

There is one more option for the lexicographer when he or she defines the correction that is not indicated in Fig. 5. He or she may immediately accept it – see Fig. 6 (the “Apply suggestion” button) - instead of creating a pending correction. Doing so, he or she directly updates the entry. This feature speeds up the process, when obvious typing errors are detected, because the pending correction does not have to wait for the author’s approval.

Every correction made using this editorial tool is recorded together with the old and new value of the time stamp, every action is signed by lexicographer, who changed the value or status, and their comments are also recorded. Thus, all the information that was ever corrected has an editorial history. It never gets lost and is always available in the editorial module pop-up. Thus, when the author is deciding whether to accept or reject some suggestion from another lexicographer, he can check the history of corrections made, find out who and when the suggestions were made. With the inclusion of a notes feature, he may even know why and under what circumstances they were made.

datum vzniku (korektor)	date and time corrector	heslo (pole) the lemma the field	zpracovatel poznámka zpracovatel editor note to editor
2014-04-14 10:28:28 (*ja*)	(druhá) pražská defenestrace (vznam-vyklad_vznamu)	*ja*	""
	vyhození dvou královských místodržících z oken Pražského hradu roku 1818		
	vyhození dvou královských místodržících z oken Pražského hradu roku 1818		
2014-04-14 10:01:12 (*ja*)	barzoj (vznam-vyklad_vznamu)	Pernicová	
	plemeno dlouhosrstých chrtů pocházejících z Ruska		"Nezačínat výklad slovem odrůda, druh, typ, forma, plemeno... - https://docs.google.com/document/d/1MZxKskjoCyIS2B1A_NNHSUIPsfxYGAjXAFixKoP/edit "
	plemeno dlouhosrstých chrtů pocházejících z Ruska		
2014-04-14 09:59:59 (*ja*)	step aerobik (vznam-vyklad_vznamu)	Pernicová	

Figure 9: The history of corrections.

5 Conclusion

Our DWS has been released and the lexicographic team uses it in their everyday work. Nevertheless, we are preparing additional modules for our DWS. This article was devoted to the editorial tool that has been deployed recently and we present it here for the first time.

We have strictly emphasized the quality of the user interface of our DWS. It must be designed according to the needs of the lexicographers that use it for the processing of large amount of lemmas.

Lexicographers are now processing lemmas using the described tools and preparing them to be published. Meanwhile we are preparing, except printed output, several applications, where published lemmas will be available for public. Using these applications like web pages, mobile applications for iOS or Android operating systems, users will be able to search and browse the dictionary on different devices.

Currently, we are preparing a very strong relation based search tool called xFilter which we will present sometime in the future.

6 References

- Barbierik, K. et al. A New Path to a Modern Monolingual Dictionary of Contemporary Czech: the Structure of Data in the New Dictionary Writing System. In Proceedings of the 7th international conference Slovo, 13-15 November 2013. Slovenská akadémia vied, Jazykovedný ústav Ľudovíta Štúra, pp. 9-26.
DEB II. Accessed at: <http://deb.fi.muni.cz/index-cs.php> [11/10/2013]
DEBDict. Accessed at: <http://deb.fi.muni.cz/debdict/index-cs.php> [11/10/2013]
IDM DPS. Accessed at: <http://www.idm.fr/products/dictionary-writing-system-dps/27/> [11/10/2013]
iLEX. Accessed at: <http://www.emp.dk/ilexweb/index.jsp> [11/10/2013]
Kochová, P., Opavská, Z., Holcová Habrová, M. (2014). At the Beginning of a Compilation of a New Monolingual Dictionary of Czech (A Report on a New Lexicographic Project). *Poster presented on this conference. Matapuna*. Accessed at: <http://sourceforge.net/projects/matapuna/> [11/10/2013]
TshwaneLex. Accessed at: <http://tshwanedje.com/tshwanelex/> [11/10/2013]
Abel, A. and A. Klosa 2012. 'The lexicographic working environment in theory and practice.' In R. V. Fjeld and J. M. Torjusen (eds.), *Proceedings of the 15th EURALEX International Congress*. Oslo: University of Oslo, 1-23.
Atkins, B. T. Sue and M. Rundell 2008. *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Acknowledgement.

This work has been supported by the grant project of the National and Cultural Identity (NAKI) applied research and development program A New Path to a Modern Monolingual Dictionary of Contemporary Czech (DF13P01OVV011).

