Development of the User Interface Tools for Creation of National Language Modules

Tigran Grigoryan, Vahan Avetisyan

Institute for Informatics and Automation Problems
P. Sevak Str. 1,
375014 Yerevan, Armenia
{va@ipia.sci.am; tigrangr@ipia.sci.am }
http://ipia.sci.am; http://www.unl.am

Abstract. The paper describes the UNL Toolbox, software for development of national language modules of UNL, designed at the Institute for Informatics and Automation Problems of the National Academy of Sciences of the Republic of Armenia. The software provides tools for creating dictionaries, enconversion and deconversion rules. There are also enconversion and deconversion modules which output the converted text and the list of occurred errors and their descriptions (if any). This software also can be used as an educational tool to learn creating UNL dictionaries and conversion rules.

1 System Overview

The UNL Toolbox is an integrated environment for UNL development. It contains tools for performing the most common tasks arising during UNL development such as dictionary creation and conversion rules creation. The Toolbox makes routine tasks like compilation of dictionary and conversion rules transparent to the end user. It allows setting options for an individual tool as well as for a whole system (for example the common output directory).

The main window of the Toolbox is divided into two parts (Fig. 1). On the left side of the window the toolbar is located. Pressing the buttons on the toolbar brings up appropriate tool in the right side of the window. Currently four tools are available – Dictionary Editor, Enconversion Rules Editor, Deconversion Rules Editor and Converter.

2 Dictionary Editor

Dictionary Editor provides a user friendly interface for creating UNL dictionaries and editing the existing ones. It uses XML to store the dictionary. When needed it is possible to export the dictionary in a standard plain text UNL dictionary format.

The dictionary in the dictionary editor has a tree-like structure. Each word is represented as a node of a tree with its stems (if any) represented as child nodes (Fig. 2).

© J. Cardeñosa, A. Gelbukh, E. Tovar (Eds.)
Universal Network Language: Advances in Theory and Applications.
Research on Computing Science 12, 2005, pp. 210–214.

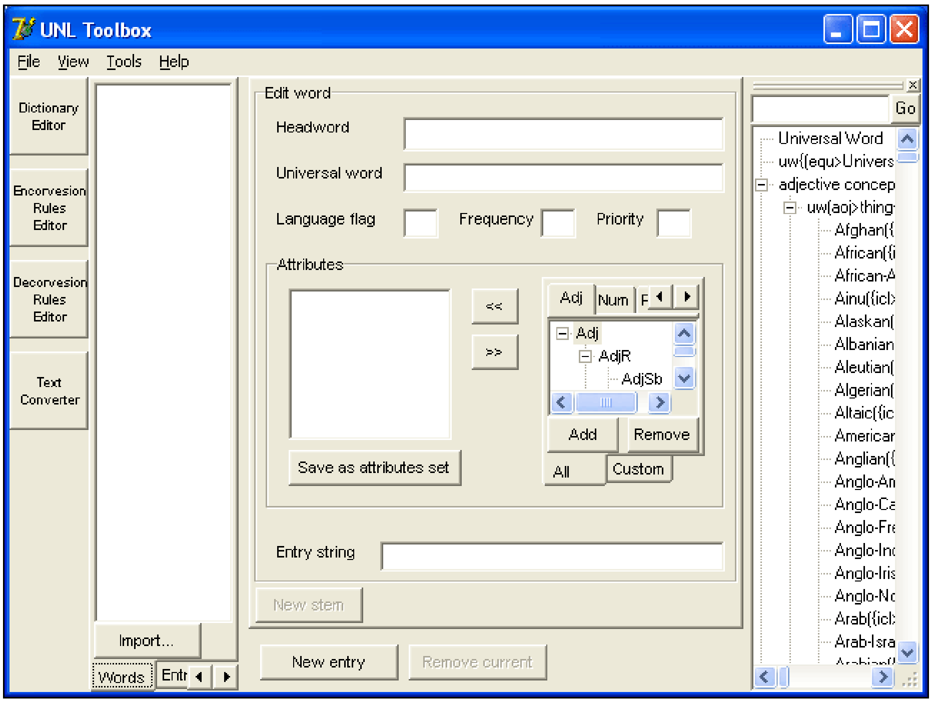


Fig. 1. Main window of the Toolbox.

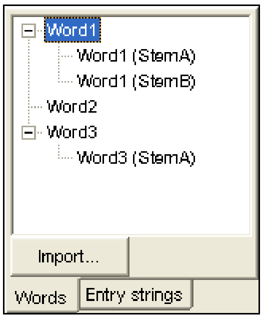


Fig. 2. Representation of a word.

In the UNL dictionary words and their stems are considered as separate dictionary entries, and when exporting the dictionary into plain text format the tree-like structure of the dictionary is lost. The tree-like structure of the dictionary is maintained only for ease of use for the end user. Before saving the dictionary the words are ordered alphabetically, preserving the tree structure.

The dictionary editor window is divided into three areas – dictionary words (on the left), attributes and properties of the word (in the center), and the UNL knowledge base tree (on the right), as shown in Fig. 3. The tree-like structure of the dictionary is shown on the left side of the dictionary editor. It has two views, which are switched using the tabs bellow the tree view. One of the views is called “Words” and it shows only the native language words, another view is called “Entry strings” and it shows entries of the dictionary in the UNL dictionary format (word, UW, attributes, etc.). The ‘Import’ button makes it possible to paste a set of words into a dictionary. When clicking on a word in the dictionary tree, the word becomes active and all its properties are shown in the central part of the window and can be edited. The attributes from the editable language-specific attribute list are assigned to the word and removed from the word using ‘>>’ and ‘<<’ buttons. It is also possible to define custom sets of attributes that are frequently used together (by clicking on the ‘Save as attributes set’ button) and assign an attribute set to the word. Attribute sets appear on the “Custom” tab of attributes tree.

The knowledge base tree (KB) on the right is the UNL knowledge base downloaded from the undl.org. On the top of the KB tree there is a text field for typing in a keyword to be searched in the KB. When the keyword is typed and the ‘Go’ button is clicked the search is performed over the KB tree and the matching (or partially matching) entry is highlighted. Click on the universal word (UW) in the KB will assign that UW to the current word in the dictionary.

If an adequate UW is not found in the KB, a new (made up) UW corresponding to the word can be typed in as a UW property of the word (in the UW textbox in the center). These new UWs are *not* added to the KB.

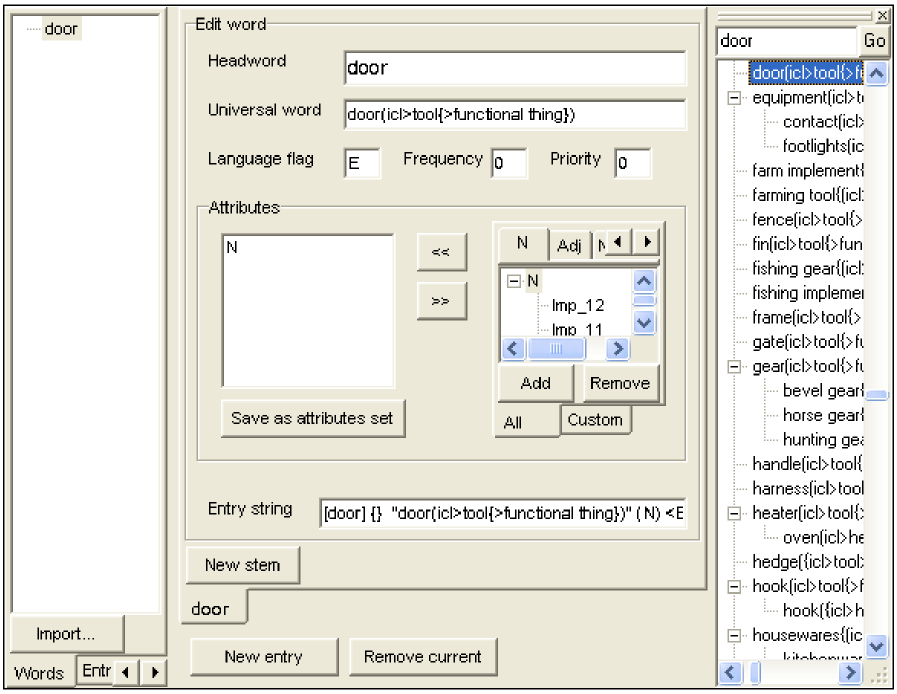
There are two approaches in adding new words to the dictionary: adding a single word and adding more than one word. A single word is added by clicking on the ‘New entry’ button after which a new entry with headword “New word” is added to the dictionary and becomes active so its properties can be set. Multiple words are added by clicking on the ‘Import’ button and pasting their headwords in the appeared popup window. After importing a set of entries each of them must be activated and the properties must be set.

3 Enconversion and Deconversion Rules Editors

Important tools in the UNL Toolbox are enconversion and deconversion rules editors. Although they are separate tools but their functionality is similar. Below the enconversion rules editor will be described (deconversion rules editor is mostly the same). As the structure of enconversion/deconversion rules is pretty complicated for novice users this tool will be very helpful for them. The tool provides an interface to the user for setting conditions and actions on the *condition windows* and *analysis windows*. It is also enables the user to assign labels (names) to the rules making it easier to manage them.

On the left side of the enconversion rules editor there is a list of the defined rules, represented by names (Fig. 4). Clicking on the name of the rule in the list will show the content of the rule in the right side of the window making it available also for editing.

To add a new rule the ‘Add’ button is used. When clicking on it the new rule will be created and will become active for editing.


Edit word

Headword: door

Universal word: door(icl>tool{>functional thing})

Language flag: E Frequency: 0 Priority: 0

Attributes: N

Entry string: [door] {} "door(icl>tool{>functional thing})" (N) <E

Fig. 3. Dictionary editor.

An existing rule can be removed by clicking on it in the list view and then on the ‘Remove’ button.

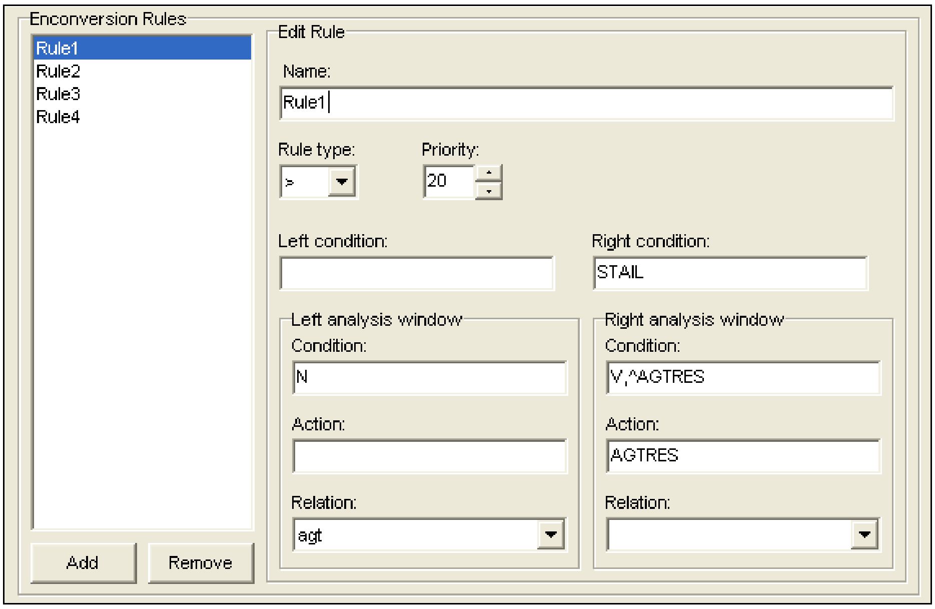
The enconversion and deconversion rules can be exported in the plain text UNL format files which can be used with EnCo and DeCo.

4 Converter

The last tool in the UNL Toolbox is the Converter. The Converter uses external enconversion and deconversion tools, such as EnCo and DeCo (from UNL Development Set) to do native language – UNL and UNL – native language translations and returns the result of the translation. An important feature of the Converter is that it also performs an analysis of the errors occurring during the translation process and represents them in a user friendly style.

5 Use and Future Development of the System

Currently the UNL Toolbox is used for creating the UNL Armenian module. Its user friendly interface and automation of routine tasks greatly increases the speed of dic-



Name: Rule1

Rule type: >

Priority: 20

Left condition:

Right condition: STAIL

Left analysis window Condition: N

Left analysis window Action:

Left analysis window Relation: agt

Right analysis window Condition: V,^AGTRES

Right analysis window Action: AGTRES

Right analysis window Relation:

Fig. 4. Ruses Editor.

tionary and conversion rules creation. The users of this system also benefit from its error handling, which makes the process of finding errors and typos easier and faster.

The system can also be used as an educational tool for those who are new to UNL. It is easy to understand and to use for novice users. And again the error handling functionality has a great educational role because the user can see where he did the mistake and after correcting it try it again.

Having integrated system that contains dictionary editing, conversion rules editing and converting tools is also beneficiary as it allows performing centralized management of the dictionary and the rules that are used with that particular dictionary. When developing a dictionary or conversion rules it is possible to check the current state of the dictionary and the set of rules using the converter to see if the change we just made brings up any anomalies or not.

The UNL Toolbox is a fully functional system but it is still being improved. The additions to the system will include a support for plug-in modules so that the other tools may be designed for the system and plugged to it as plug-in modules.

References

1. Uchida H., Zhu M., The Universal Networking Language (UNL) specifications version 3.0,*1998*. Technical Report, United Nations University, Tokyo, 1998.
2. EnConverter Specification Version 2.1, UNU/IAS/UNL Centre, Tokyo 150-8304, Japan; http://www.unl.ias.unu.edu/unlsys/enco/index.html, 2000.