Wordnet Development Using a Multifunctional Tool

Ivan Obradović, Ranka Stanković

Wordnet Development Using a Multifunctional Tool

Ivan Obradović
University of Belgrade
Faculty of Mining and Geology
Đušina 7, 11000 Belgrade, Serbia
ivano@rgf.bg.ac.yu

Ranka Stanković
University of Belgrade
Faculty of Mining and Geology
Đušina 7, 11000 Belgrade, Serbia
ranka@rgf.bg.ac.yu

Abstract

In this paper we present a multifunctional tool for manipulating heterogeneous language resources. The tool handles electronic dictionaries, wordnets and aligned texts, and provides for their synchronous use in various tasks. We focus here on the description of the possibilities this tool offers in the development of wordnets. Besides the wordnet module which enables parallel handling of two wordnets, other modules, such as the module for morphological dictionaries and the module for aligned texts, as well as available finite state transducers, can also be used to aid the user in developing and refining the wordnet.

Keywords

Wordnet development, language resource integration, HLT tools

1. Introduction

The first wordnet, namely the Princeton WordNet (PWN), or simply WordNet, was conceived in 1985 by George Miller, a renowned professor of psychology at Princeton University and his associates from the Cognitive Science Laboratory. They started to develop PWN as a linguistic database that maps the way the mind stores and uses language, namely as some sort of a mental lexicon to be used in the scope of psycholinguistic research projects [6]. PWN was formalized as a semantic network of concepts, abstract ideas or mental symbols that denote objects in a given category or class of entities, interactions, phenomena, or relationships between them. In PWN, concepts are lexicalized by one or more synonymous English words (simple or compound) and represented by a synset, a set of synonymous English word-sense pairs accompanied by a definition of the concept. Concepts are interconnected by various semantic relations, such as hypernym/hyponym (kind of, e.g. animal/dog) or holonym/meronym (part of, e.g. hand/finger). As of 2006, this database contains about 150,000 words organized in over 115,000 synsets for a total of 207,000 word-sense pairs.

Wordnets for other languages followed, developed by individual teams or through multilingual projects, such as EuroWordNet, when wordnets for English, Dutch, Italian, Spanish, French, German, Czech, and Estonian were developed, based on the Princeton wordnet, and aligned by interconnecting synsets representing the same concept in different languages via an Inter-Lingual-Index (ILI) [14]. This index also gives access to a shared top-ontology that provides a common semantic framework for all the languages, while language specific properties are maintained in the individual wordnets. BalkaNet, a project aimed at developing wordnets for Bulgarian, Greek, Romanian, Serbian and Turkish and expanding the Czech wordnet, followed an approach similar to EuroWordNet [13]. Namely, BalkaNet wordnets were also developed on basis of PWN and the top-ontology accepted in EuroWordNet, and aligned by using ILI.

From a lexicographer's point of view, the development of a wordnet, perceived as a specific form of dictionary and hierarchical thesaurus for a particular language, opens two critical issues. The first pertains to the organization of the conceptual network. Simply put, the issue is how to define the concepts for a particular language and how to establish links among them? In other words, how to place a concept in the right position within the wordnet? Once this issue is resolved, the second issue needs to be tackled. Namely, how should the concept be lexicalized, namely, how two select the set of word-sense pairs for the synset that represents the concept?

Many wordnets approached the first problem by relying on the conceptual network of PWN as the basis for development. This approach appeared especially convenient in cases of aligned multilingual wordnets, such as EuroWordNet and BalkaNet, since a common conceptual network substantially alleviated the alignment. However, within the BalkaNet project the following questions have often been raised: are concepts linguistically independent or not, are the lexicalization patterns for concepts universal, is the structure of PWN valid for other languages as well, is the set of semantic relations built in PWN sufficient for all languages [15]. Although the work on the development of specific wordnets for Balkan languages often pointed to a negative answer to these questions, this approach has essentially not been abandoned. However, language specific concepts were also developed for each particular wordnet, as well as a set of concepts common to BalkaNet languages and unknown to PWN [10].

Once a concept has been accepted and placed within the conceptual framework of a particular language, the lexicographer is confronted with the problem of its lexicalization. Besides selecting the appropriate synonyms, he/she also needs to provide a gloss, and preferably usage examples. As synset elements appear as a word-sense pairs the lexicographer has to assign senses to all chosen words. It goes without saying that other linguistic resources, such as electronic dictionaries, bilingual word lists and corpora can be of invaluable help to the lexicographer in accomplishing this task.

In this paper we present a multifunctional tool which, among its other functionalities, serves as an aid in developing wordnets that offers more possibilities for resolving the aforementioned problems than other wordnet development tools.

In the next Section we will give a brief overview of the best known wordnet development tools. Section 3 outlines some of the functionalities our tool offers, which are developed for resources other than wordnets, but which can be very useful as an aid to the lexicographer in wordnet development. Section 4 describes the wordnet module of our tool, and how it operates in conjunction with other modules in the wordnet development task. A conclusion follows in Section 5.

2. Wordnet development tools

A number of software tools for wordnets have been developed in the past decades. As it could have been expected, the first wordnet browser was developed for PWN. Its latest version is freely distributed with the version 2.1 for Windows of the Princeton wordnet,[1] while a web application for PWN browsing is also available.[2] Other wordnet tools have been initialized within larger projects, such as EuroWordNet and BalkaNet, and we will describe their basic features briefly. However, there are also many other tools, developed for individual languages, such as Russian [1].

Polaris, a tool for creating, editing and exporting wordnets [11], and Periscope, a graphical database viewer for viewing and exporting wordnets [5] were the two main tools implemented and used within the EuroWordNet project. Polaris allows the user to import wordnets from properly formatted ASCII files, to edit and add relations in the wordnets and to formulate queries. This tool also makes it possible to visualize the semantic relations as a tree-structure that can directly be edited. Trees and sub-trees can be expanded, shrunk, and stored as distinct sets of synsets, which can then be separately manipulated, saved or loaded. In Polaris, it is also possible to switch between the wordnets via the ILI, and to match sets of synsets across wordnets. Periscope is a public viewer that can be used to look at wordnets created by the Polaris tool. It has some of the functionalities of Polaris, but it cannot be used for importing or changing wordnets. Although Polaris, a property of Lernout and Hauspie, can still be licensed either directly from Lernout and Hauspie or from ELRA, whereas Periscope is freely distributed,[3] the development of both tools ceased with the termination of EuroWordNet. Other tools, such as WEI (Web EuroWordNet Interface), have been developed within the EuroWordNet project [2], but their further development has also terminated.

Another browser and editor, VisDic, was developed within the framework of the BalkaNet project and used as the main tool for building all BalkaNet wordnets [8]. Although VisDic, available for both Linux and Windows platforms, has been primarily aimed at browsing and editing wordnets, it has been developed as a more general tool, namely, as a graphical application for viewing and editing various types of dictionary databases stored in XML format. This tool can be configured to handle simultaneously up to 10 dictionaries, which can be monolingual or translational dictionaries, but also thesauri or plain corpora. Thus, VisDic went a step further as a tool which can do more than just editing and browsing wordnets. In addition to that, and contrary to the EuroWordNet tools, the development of VisDic did not discontinue with the termination of the BalkaNet project. Although the development of VisDic itself has finished, a completely new client-server version of this tool, DEBVisDic, is now being developed [9], and can be obtained free of charge, subject to registration.[4]

WS4LR (WorkStation for Lexical Resources), the tool that we present in this paper builds on the features developed by previous tools, especially VisDic, when wordnets are concerned. However, it differs substantially from other wordnet tools by the simple fact that it has not been conceived primarily as a wordnet tool, and that handling wordnets is only one of its functionalities. Namely, WS4LR is a software tool aimed at manipulating heterogeneous lexical resources, of which wordnets are just one type. The tool enables integrated handling of electronic dictionaries, wordnets, aligned texts and transducers equally, and has already proved very useful for various tasks. Although the tool has a module especially developed for manipulating wordnets, the fact that all other resources are also at hand, and that they can be exploited simultaneously with wordnet development, means that the lexicographer developing the wordnet can get more support in the tasks he/she is confronted with, than the majority of wordnet tools can offer.

[1] http://wordnet.princeton.edu/obtain

[2] http://wordnet.princeton.edu/perl/webwn

[3] http://www.illc.uva.nl/EuroWordNet/sample.html

[4] http://nlp.fi.muni.cz/projekty/visdic/

3. A Multifunctional Language Resource Tool

3.1 Motivation

The Human Language Technology group at the University of Belgrade has been developing various lexical resources over quite a long period, reaching a considerable volume to date. Given the fact that these resources have been developed for many years, they have naturally been conceived within different projects and frameworks, both from the conceptual and the technological point of view.

Although the HLT group made every reasonable effort to keep the ever growing pool of resources as coherent and standardized as possible, a certain level of heterogeneity was inevitable. Hence, due to the growth of the volume of resources as well as their heterogeneity, there was a rising need for developing a tool that would facilitate the maintenance, exploitation and integration of available resources as well as their further development. Embarking on this task, the HLT group produced an integrated and easily adjustable tool, the workstation for language resources, labeled WS4LR, which greatly enhances the potentials of manipulating each particular resource as well as several resources simultaneously. Exploiting the synergy of various resources, this tool proved very useful in many HLT tasks, including wordnet development.

3.2 Structure

WS4LR is composed of several modules which perform the following main functions (Figure 1):

- development and refinement of wordnets, where both work with a monolingual wordnet and simultaneous usage of two wordnets for different languages is supported

- management of a system of electronic dictionaries which consist of morphological dictionaries of lemmas for simple and compound words, but also of bilingual and multilingual dictionaries

- manipulation of parallel aligned texts, allowing for various forms of their presentation and usage

- conversions from different formats such as one character encoding set to another, or one resource format to another

The tool is developed in C# and operates on the .NET platform. An important feature of WS4LR is its flexibility expressed both by the possibility of setting environment parameters and by the possibility of invoking command-line routines and using external Perl, Awk, and XSLT scripts. WS4LR functions and their usage are explained in a printed manual that accompanies the software, as well as in a concise on-line context sensitive help.

In this section we will briefly describe the WS4LR functionalities which are not directly related to wordnet manipulation, but which can be very useful in performing this task.

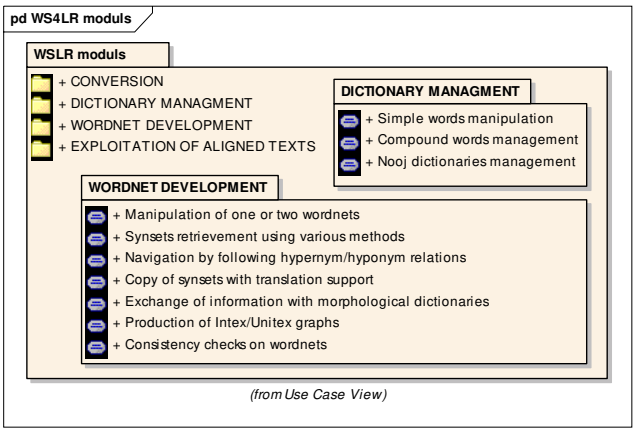

Figure 1. The UML diagram showing WS4LR modules

3.3 Dictionary management

The electronic dictionaries that WS4LR manipulates are monolingual morphological dictionaries, but also a bilingual word list and a multilingual dictionary of proper names. However, the main task of this module is to enable the manipulation of the system of morphological dictionaries of canonical forms, or lemmas, for both simple and compound words. Morphological dictionaries are of great importance for highly inflective languages, such as the group of Slavic languages. The absence of morphological information in wordnets has turned out to be a serious flaw in many applications. Thus the possibility, offered by WS4LR to simultaneously exploit both resources proved to be a great advantage in wordnet development. Given the importance of morphological, but also bilingual and multilingual dictionaries in wordnet development, we will now briefly describe the basic features of the dictionary management module.

The lemma in a morphological dictionary of simple words has the following format:

*lemma.Knnn [+SinSem]**

where *lemma* is the word form used in traditional dictionaries, *K* represents the part of speech (noun, verb, adjective, etc.), and *nnn* the inflectional class code of the lemma, whose characteristics are described by a corresponding transductor labeled *Knnn.* A set of optional tags *+SinSem* follows, which describe the syntactic, semantic, derivational and other properties of the lemma. The format of the lemmas for compound words is more complex, but it basically relies on the same principles.

The format used in the system of morphological dictionaries is known as the LADL format [4]. The first system developed for processing of texts using dictionaries in LADL format was a system called Intex [12]. Intex uses dictionaries in combination with regular expressions and inflectional and morphological finite state transducers (FSTs) to locate morphological, lexical and syntactic

patterns, remove ambiguities, and tag simple and compound words in texts. The text parsing possibilities offered by regular expressions and FSTs proved also useful in wordnet development, and we will give some more details on that in the following section.

Although Intex has been developed for many years and used by over 80 HLT laboratories, it had a serious shortcoming. Namely, Intex did not support the processing of texts in Unicode, and as the usage of this encoding became more and more frequent, the development of a new tool that could handle text in Unicode became inevitable. Building on the functionalities of Intex, but allowing the processing of texts in Unicode, such a new tool has been developed under the name of NooJ[5]. Another system, Unitex, based on LADL format and supporting resources in Unicode has been developed in parallel, and is also available[6].

Since all three systems (Intex, Unitex, and NooJ) provide for processing of texts on basis of dictionaries, in combination with regular expressions and FSTs, and each of them has some useful specific features, the dictionary management module allows the user to activate the functions of the Intex, Unitex and/or NooJ system, and select a list of dictionaries he/she wants to use. As none of the three systems offers possibilities for managing the content of dictionaries themselves, the WS4LR dictionary management module has been developed to enable the entry, editing and review of lemmas of simple and compound words, supporting the specific features of all three solutions. Dictionaries are organized in a modular fashion, in several sub-dictionaries as separate files. This is not only important from the practical point of view, since smaller files are easier to manipulate, but also because of the fact that in text recognition by Intex/Unitex the usage of all dictionaries is not always necessary, or even recommended.

Without going into details of dictionary management, we will just point out that the dictionary management module enables the user to modify or delete all the information attached to a lemma, or the lemma itself, as well as to add new entries. A new entry can be generated from scratch or by copying an existing lemma, which in some cases facilitates the work. The regular expression or a FST graph describing the inflectional properties of the selected lemma can be inspected and corrected if found inadequate.

An important feature of this module is the ability of retrieving efficiently a subset of lemmas by matching the lemmas, their part of speech, inflectional class code, syntactic and semantic markers or their Boolean combination. For instance, one can look for all the dictionary entries starting or ending with a search string. (Figure 2). The latter is particularly useful when the inflectional class code of a new lemma is being established, since this code depends on the lemma ending.

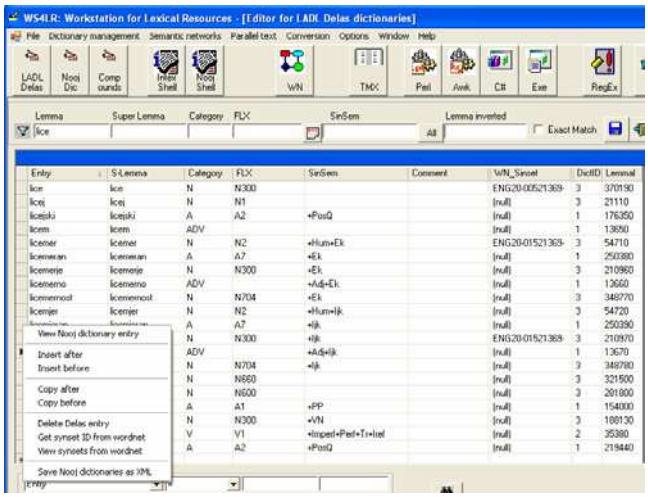

Figure 2. Retrieval of all words starting with "lice"

Given the fact that compound words pose a specific problem for lexicographers, and that in the wordnet development task the dictionaries of compound words can be a valuable resource, we will now very briefly sketch the handling of dictionaries of compound lemmas.

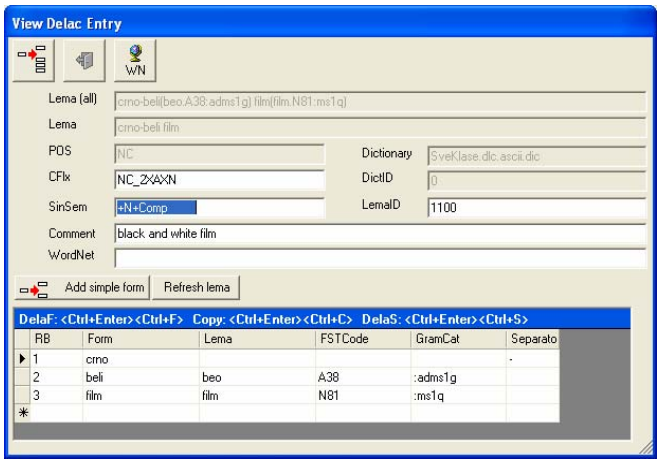

Figure 3. Form for compound entries

The form for new entries in these dictionaries is more complex since more information need to be supplied. In the upper part of the form the information pertaining to the entry as a whole is displayed or typed, while in the lower part the information associated to the compound lemma constituents is entered (Figure 3). For inflected compound constituents additional information is needed: the lemma, its inflectional class code, as well as the list of grammatical categories of the form that appears in the compound lemma. For example, in the compound *crno-beli film* (black and white movie), the lemma for the constituent form *beli*

[5] http://www.nooj4nlp.net

[6] http://igm.univ-mlv.fr/~unitex/

is *beo*. The form of this adjective in the compound lemma is inflected in order to agree in gender with the noun *film*.

As we have already mentioned, WS4LR also handles bilingual word lists, but also multilingual dictionaries, such as Prolex, the multilingual dictionary of proper names based on an ontology built around the conceptual proper name and its relations [7]. This adds additional functionality to the integration of lexical resources offered by WS4LR in various tasks, including wordnet development.

3.4 Management of aligned parallel texts

Parallel texts, which usually originate from a text in one language and its translation in another, are often aligned at a certain level (paragraph, sentence, etc) by matching the corresponding segments of the original and its translation. Aligned parallel texts are a valuable lexical resource which can be used for many HLT tasks, but we also found them very useful in wordnet development, as we shall illustrate in the next section.

The WS4LR module for management of aligned parallel texts uses texts which have previously been aligned using Xalign as an alignment tool [3]. The module converts these texts to the Translation Memory eXchange (TMX) format, which is becoming the standard format for aligned texts. Figure 4 depicts the form with parameters for TMX file generation and a part of a TMX document. Needless to say, the module can use texts that are already in that format.

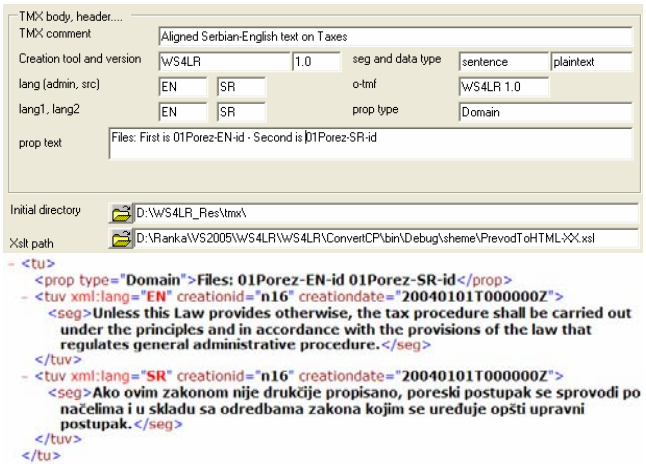

Figure 4. TMX generations parameters and part of the TMX document

Aligned texts can be visualized in various ways by choosing the appropriate XSLT stylesheet. Namely, the user can obtain the aligned texts in HTML format, but also in textual, XML, tabular or TMX format.

3.5 Conversion

The conversion module is important since it adds to the flexibility of resources exploitation. Conversion from one character encoding set to another is extremely important for languages such as Serbian, where two alphabets, Cyrillic and Latin are equally used. WS4LR enables the exploitation of language resources both in Cyrillic and Latin alphabet, as well as in a special encoding, that uses the ASCII character set and that can be unambiguously transformed into Serbian Latin or Serbian Cyrillic alphabet. WS4LR offers to the user the option to apply the transformation only to a part of the file, such as an XML file where only the text should be converted while the XML tags shouldn't be altered. Similarly, when a dictionary type file is transformed, only lemmas and word forms are converted, not the part of speech and grammatical codes. The user can choose a conversion Perl or awk script suitable for the specific file type, or produce his/her own script easily. The module also makes switching between Intex and Unitex easy. This would otherwise be a problem since Intex does not support Unicode and Unitex works only with Unicode.

4. Wordnet management

The wordnet management module supports search of wordnets, their visualization, as well as their development and refinement. When this module is activated, the main form opens with two wordnet windows, thus offering to the user the possibility to work with one or two wordnets. In the current version of WS4LR these two wordnets are the Serbian and English wordnet, but the tool can be easily adapted for any two wordnets. If the user decides to work with both wordnets in parallel, he/she can always synchronize them via the ILI. The main form for wordnet management also opens a window with a bilingual word list (Figure 5).

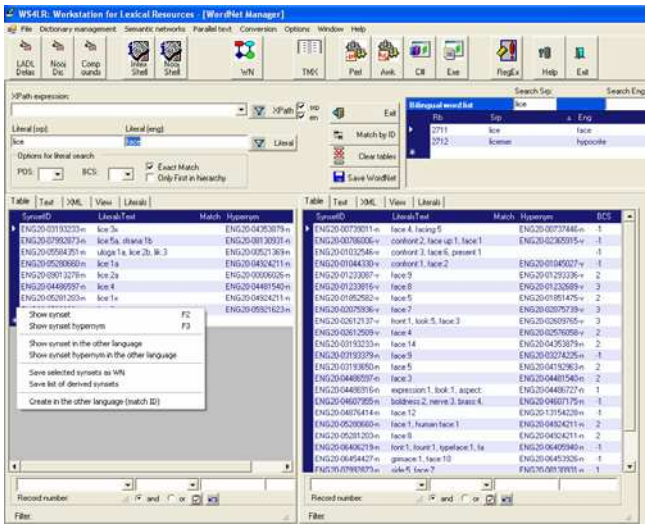

Figure 5. Main form for wordnet management

4.1 Search, visualization and modification

The search of wordnets can be performed in several ways, and the user can always choose whether he/she wants to

search just one wordnet or both of them. Namely, synsets can be retrieved from wordnets into the two available wordnet windows using various methods, from simple string matching to complex Xpath expressions. The user can, for example, specify one or two strings, depending on whether he/she wants to search one or both wordnets for synsets containing words that match the string(s). The user can also specify whether he/she wants an exact match or not, and in the latter case the system will retrieve not only all synsets with words matching the search string(s), but also those that contain words which contain the specified string(s) as their part. On the other hand, the user can use an Xpath expression to retrieve synsets on basis of various other criteria, such as the domain synsets belong to. Thus, for instance, by means of the Xpath expression:

"//SYNSET[DOMAIN='geology']"

the user can retrieve all synsets from the wordnet that belong to the domain of geology, or more precisely, that contain the XML tag <DOMAIN> with the content "geology". WS4LR offers predefined Xpath expressions, but the user can also define these expressions him/herself.

Once the user has retrieved the synsets of interest from the wordnet, he/she can now proceed to their modification or generation of new synsets. Every retrieved synset can be visualized in various forms: as text, XML or hypernym/hyponym tree (Figure 6). In the hypernym/hyponym view, the user can easily navigate through its hypernym/hyponym tree and proceed to further modifications of synsets. There is also an edit view for the synset which allows the user to modify the synset contents: words-sense pairs, definition, usage, but also other properties, such as semantic relations to other synsets (Figure 7).

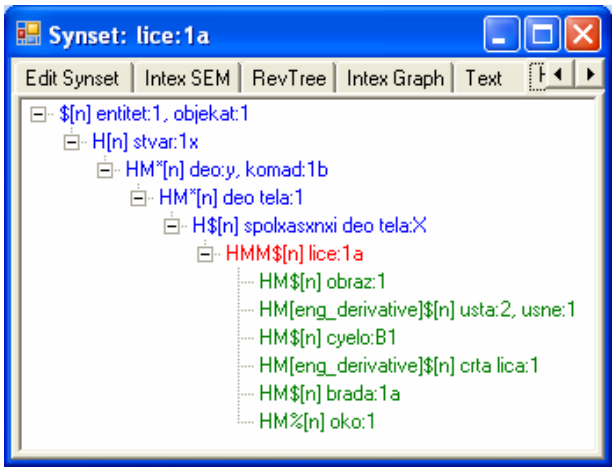

Figure 6. The hypernym/hyponym view of a synset

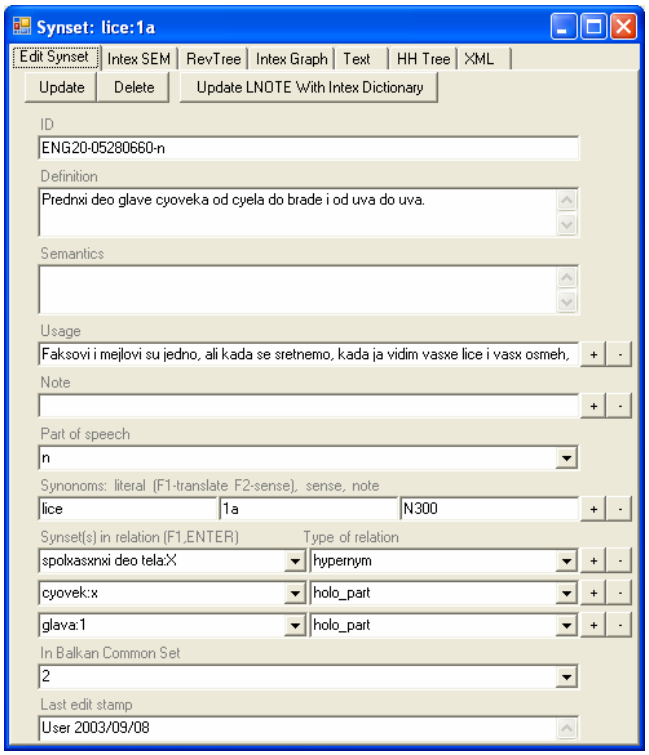

Figure 7. The edit view of a synset

4.2 Adding new synsets

WS4LR allows for adding of new synsets to wordnets using predefined forms. As we have noted previously the two main problems is how to place the synset in the conceptual network, and how to select the appropriate word-sense pairs to represent the concept of the synset.

With a particular concept in mind, one approach of the lexicographer to the solution of the first problem would be an inspection of the part of the wordnet where he/she believes a hypernym of the new synset might be found, and if an appropriate hypernym is found, placing the new synset as its hyponym. In order to find such a hyponym the search possibilities offered by WS4LR could be exploited, as well as the possibility of navigation through the hypernym/hyponym tree.

Another, more frequently used option is to exploit the possibility of working simultaneously with two wordnets. Namely, as we have already pointed the majority of wordnets tend to be aligned with PWN. Thus, if the wordnet that is being developed is used in parallel with PWN, then the lexicographer might first attempt to identify the English synset that corresponds to the concept he/she wants to add to the wordnet. The available bilingual word list can help the user locate the candidate PWN synsets by typing the words in Serbian that denote the concept he/she has in mind, and retrieving all synsets containing the corresponding English words from the bilingual list. This procedure is fully automated in WS4LR.

If the corresponding synset in PWN is found, the new Serbian synset can easily be inserted in the appropriate place in the wordnet using the WS4LR option "Create synset in the other language". This option creates a copy of the PWN synset in the Serbian wordnet, and if necessary, it also creates copies of all its missing hypernyms, to prevent the new synset of becoming a "dangling" synset. Once a copy of a synset is created the user has to make the necessary modifications: its definition, usage examples, and above all, the synonymous word-sense pairs denoting the concept this synset relates to. As we noted before, the two wordnets are presently the Serbian and English wordnet, but they can be any two wordnets, and the new synset can be created by aligning the wordnet that is being developed with some other wordnet, which is not necessarily PWN. This could be the case, for example, when new synsets are defined for concepts not recognized in PWN but which exist in other wordnets, such as Balkan specific concepts.

WS4LR also offers substantial aid in solving the second task, namely, the selection of synonymous words for the synset, and the assignment of meanings to these words. Although it is reasonable to assume that the wordnet developer has a pretty good idea of the candidate words for the synset of the concept he/she wants to add to the wordnet, it is also possible that he/she might neglect some of them. As the simplest and most straightforward aid the bilingual wordlist can be used. Words from the source (English) synset can be matched with words in the target language as probable candidates. The multilingual dictionary Prolex could be used in a similar manner.

Figure 8. Aligned texts with highlighted words

Another, more complex option is to use aligned texts. If PWN is used for the source synset, then the language of one of the parallel texts must be English. Namely, WS4LR allows the user to search aligned texts using words from both parallel texts. All of the words found in both texts will be highlighted (in blue color) (Figure 8). A lexicographer can use this option to extract possible candidate words for a synset by searching aligned texts with words from the original PWN synset and words he/she has already selected for the target synset. Then, if a highlighted word found in the text in English does not have a highlighted match in the text in the target language, the lexicographer should inspect the sentence in the target language for a possible match, which would then be a new candidate for the synset.

Once the user has rounded all the candidate words for the synset he/she might be in doubt whether one or more words properly fit into the synset. In that case the user might want to observe these words within a context, which can be done by searching a corpus for these words and obtaining concordances. By getting the occurrences of the words within the context, the user will be able to make a better assessment whether they are really appropriate or not. In WS4LR this can be realized by creating a regular expression or FST graph from one or more words, and using it to search a text in the target language (Figure 9).

In the case when the user is working on a particular set of synsets, e.g. on adding concepts to a certain domain, he/she might find it useful to retrieve all PWN synsets from that domain by means of an Xpath expression and then using the wordnet module option "Match by ID" which matches synsets in the source and target wordnet via the ILI. This options indicates which PWN synsets have a match in the target language, regardless of the fact whether these target language synsets have previously been retrieved from the wordnet by the user or not, and which PWN synsets do not have a match. The latter are obviously candidates for new synsets in the target language.

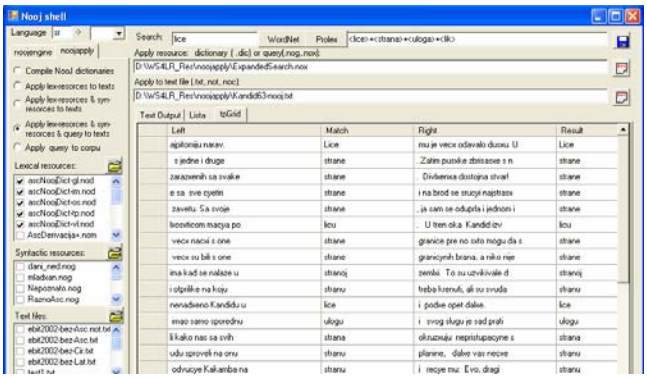

Figure 9. Concordances obtained by searching a text with a regular expression

The WS4LR wordnet module also performs various consistency checks on wordnets. For example, when word senses are in question, WS4LR provides information of the senses that have already been used for a word, so the user can assign a sense tag that has not previously been assigned, thus preventing duplicate word-sense pairs. The wordnet module can also detect dangling relations, and the use of the same word in a hypernym/hyponym pair, which is not allowed.

Morphological dictionaries extend the search possibilities within WS4LR by enabling searches with all inflected forms of the words, as can be observed on

wordnet related searches depicted on Figures 8 and 9. This is of great importance in the case of highly inflective languages, such as Serbian. The possibility of performing the searches in aligned and monolingual texts using all inflective forms, greatly improves the recall and thus the usability of their results.

Finally, we should note that WS4LR enables the enrichment of synsets with morphosyntactic information from morphological dictionaries. The tool can search for all synset words in morphological dictionaries of simple or compound lemmas, retrieve their inflectional class codes, and assign them to synset words using the <LNOTE> XML tag. If more lemmas of the same form exist, they are all offered to the user to choose the appropriate one. The missing morphosyntactic information can thus be added to wordnets.

5. Conclusion

Wordnet refinement and development described in this paper relies on WS4LR, a multifunctional tool that integrates diverse language resources and is thus more powerful than the majority of other wordnet tools. The desktop version of WS4LR is fully operational and is already being used as the main tool for developing resources in Serbian, including the Serbian wordnet, but its commercial applications have not yet been considered. Although a systematic evaluation of WS4LR has not been performed, there have already been several enhancements of the tool on basis of user feedback. Such enhancements include specific search mechanisms and mechanisms for extraction of groups of synsets on basis of various criteria.

The HLT group is now working on porting the most important functions of WS4LR on the web. The ambition of the HLT group is to make a full-scale web version of this tool which would, among other things, enable its usage in wordnet development by several lexicographers concurrently, with all the possibilities the desktop version now offers. Presently, some of the WS4LR functions are available on the web for searches based on morphological (using dictionaries) semantic (using wordnets) and multilingual (using aligned multilingual wordnets) expansions of the initial query.

6. References

[1] V. Balkova, A. Sukhonogov and S. Yablonsky. Russian WordNet, Proceedings of the Second International WordNet Conference – GWC 2004, Brno, Czech Republic, pp. 31-38, 2004.

[2] L. Benítez, S. Cervell, G. Escudero, M. López, G. Rigau, and M. Taulé. Methods and tools for building the Catalan WordNet. Workshop on Language Resources for European Minority Languages on LREC 1998, Granada, Spain, 1998.

[3] P. Bonhomme, T.M.H. Nguyen and S. O'Rourke. XAlign: l'aligneur de Langue & Dialogue, 2001, http://www.loria.fr/equipes/led/outils/ALIGN/align.html

[4] B. Courtois and M. Silberztein (eds.) Dictionnaires électroniques du français. Langue française 87. Paris: Larousse, 1990.

[5] I. Cuypers and G. Adriaens. Periscope: the EWN Viewer, EuroWordNet Project LE4003, Deliverable D008-D012. University of Amsterdam, Amsterdam. 1997.

[6] C. Fellbaum, (ed.). WordNet: An Electronic Lexical Database, The MIT Press, Cambridge, Massachusetts, London, England, 1998.

[7] T. Grass, D. Maurel and O. Piton. Description of a Multilingual Database of Proper Names. Lecture Notes in Computer Science, Advances in Natural Language Processing, Third International Conference, PorTAL, June 2002, Faro, Portugal, 23-26, Springer, Berlin, Vol. 2389, pp.31-36, 2002.

[8] A. Horák and P. Smrž. VisDic – wordnet browsing and editing tool, Proceedings of the Second International WordNet Conference – GWC 2004, Brno, Czech Republic, pp. 136–141, 2004.

[9] A. Horák, K. Pala, A. Rambousek, M. Povolný. DEBVisDic - First Version of New Client-Server Wordnet Browsing and Editing Tool. Proceedings of the Third International WordNet Conference - GWC 2006, Seogwipo, Jeju Island, Korea, pp. 325-328, 2006.

[10] C. Krstev, I. Obradović, D. Vitas. Developing Balkan specific concepts within BalkaNet - a multilingual database of semantic networks, Proceedings of the Fifth International Conference Formal Approaches to South Slavic and Balkan Languages, Sofia, Bulgaria, October 2006, S. Koeva, M. Dimitrova-Vulchanova (eds.), pp. 94-98.

[11] M. Louw. Polaris User's Guide: The EuroWordNet Database Editor, EuroWordNet (LE-4003), Deliverable D023D024, Lernout & Hauspie - Antwerp, Belgium, 1998, available at http://www.illc.uva.nl/EuroWordNet/docs.html

[12] M. Silberztein. Le dictionnaire électronique et analyse automatique de textes: Le systeme INTEX, Paris: Masson, 1993.

[13] D. Tufiş, (ed.). Special Issue on BalkaNet Project, Romanian Journal on Information Science and Technology, Bucureşti, Publishing house of the Romanian academy, Vol. 7, No.1-2, 2004.

[14] P. Vossen, (ed.). EuroWordNet: A Multilingual Database with Lexical Semantic Networks, Dordrecht, Kluwer Academic Publishers, 1998.

[15] P. Vossen, Introduction to the Special Issue on the BalkaNet Project. Romanian Journal of Information Science and Technology, Vol. 7, No.1-2, pp. 5-6, 2004.