A Distributed Database System for Developing Ontological and Lexical Resources in Harmony

[1] Faculty of Informatics
Masaryk University
Botanická 68a, 602 00 Brno
Czech Republic
hales@fi.muni.cz, xrambous@fi.muni.cz
[2] Faculteit der Letteren
Vrije Universiteit van Amsterdam
e Boelelaan 1105, 1081 HV Amsterdam
The Netherlands
Piek.Vossen@irion.nl

Abstract. In this article, we present the basic ideas of creating a new information-rich lexical database of Dutch, called Cornetto, that is interconnected with corresponding English synsets and a formal ontology. The Cornetto database is based on two existing electronic dictionaries - the Referentie Bestand Nederlands (RBN) and the Dutch wordnet (DWN). The former holds FrameNet-like information for Dutch and the latter is structured as the English wordnet. In Cornetto, three different collections are maintained for lexical units, synsets and ontology terms.

The database interlinks the three collections and aims at clarifying the relations between them. The organization and work processes of the project are briefly introduced.

We also describe the design and implementation of new tools prepared for the lexicographic work on the Cornetto project. The tools are based on the DEB development platform and behave as special dictionary clients for the well-known DEBVisDic wordnet editor and browser.

1 Introduction

Lexical data and knowledge resources has rapidly developed in recent years both in complexity and size. The maintenance and development of such resources require powerful database systems with specific demands. In this paper, we present an extension of the DEBVisDic environment [1] for the development of a lexical semantic database system for Dutch that is built in the Cornetto project. The system holds 3 different types of databases that are traditionally studied from different paradigms: lexical units from a lexicological tradition, synsets within the wordnet framework and an ontology from a formal point of view. Each of these databases represents a different view on meaning. The database system is specifically designed to create relations between these databases and to allow to

A. Gelbukh (Ed.): CICLing 2008, LNCS 4919, pp. 1–15, 2008.

edit the information in each. It represents a complex editing environment but also a research tool to study the relations between language, as defined in a lexicon and wordnet, and knowledge, as defined in an ontology.

The paper is further structured as follows. In the Section 2, we will describe the Cornetto project in terms of the design of the database structure and the major editing actions. The Section 3 introduces the DEB platform and the new features that have been introduced for the Cornetto project. Finally, we describe the specific client interface for viewing and editing the data in the Section 4.

2 The Cornetto Project

Cornetto is a two-year Stevin project (STE05039) in which a lexical semantic database is built, that combines Wordnet with FrameNet-like information [2] for Dutch. The combination of the two lexical resources will result in a much richer relational database that may improve natural language processing (NLP) technologies, such as word sense-disambiguation, and language-generation systems. In addition to merging the Wordnet and FrameNet-like information, the database is also mapped to a formal ontology to provide a more solid semantic backbone.

The database will be filled with data from the Dutch Wordnet [3] and the Referentie Bestand Nederlands [4]. The Dutch Wordnet (DWN) is similar to the Princeton Wordnet for English, and the Referentie Bestand (RBN) includes frame-like information as in FrameNet plus additional information on the combinatoric behaviour of words in a particular meaning.

An important aspect of combining the resources is the alignment of the semantic structures. In the case of RBN these are lexical units (LUs) and in the case of DWN these are synsets. Various heuristics have been developed to do an automatic alignment. Following automatic alignment of RBN and DWN, this initial version of the Cornetto database will be extended both automatically and manually.

The resulting data structure is stored in a database that keeps separate collections for lexical units (mainly derived from RBN), synsets (derived from DWN) and a formal ontology (SUMO/MILO plus extensions [5]). These 3 semantic resources represent different view points and layers of linguistic, conceptual information. The alignment of the view points is stored in a separate mapping table. The database is itself set up so that the formal semantic definition of meaning can be tightened for lexical units and synsets by exploiting the semantic framework of the ontology. At the same time, we want to maintain the flexibility to have a wide coverage for a complete lexicon and encode additional linguistic information. The resulting resource will be made available in the form of an XML database.

The Cornetto database provides a unique combination of semantic, formal semantic and combinatoric information.

2.1 Architecture of the Database

Both DWN and RBN are semantically based lexical resources. RBN uses a traditional structure of form-meaning pairs, so-called Lexical Units [6].

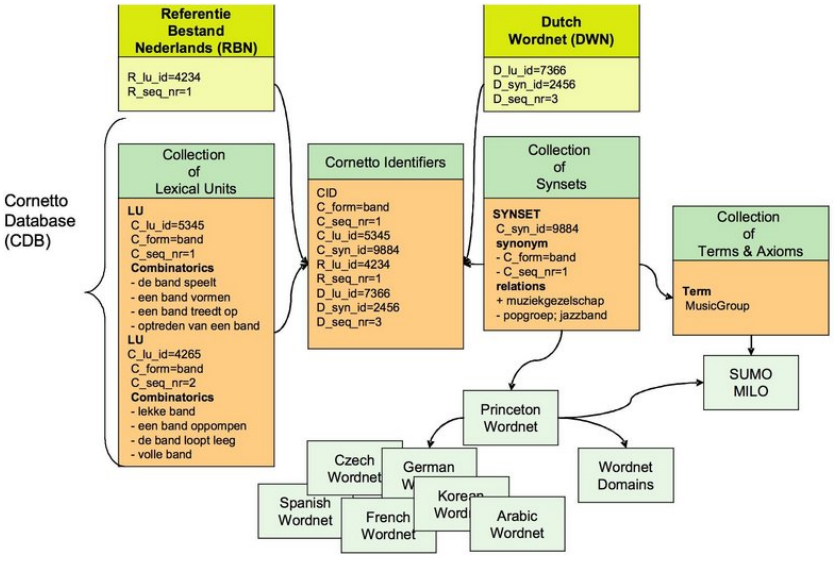

Fig. 1. Data collections in the Cornetto database

The Cornetto database (CDB) consists of 3 main data collections:

1. Collection of Lexical Units, mainly derived from the RBN
2. Collection of Synsets, mainly derived from DWN
3. Collection of Terms and axioms, mainly derived from SUMO and MILO

The Lexical Units are word senses in the lexical semantic tradition. They contain all the necessary linguistic knowledge that is needed to properly use the word in a language. The Synsets are concepts as defined by [7] in a relational model of meaning. Synsets are mainly conceptual units strictly related to the lexicalization pattern of a language. Concepts are defined by lexical semantic relations. For Cornetto, the semantic relations from EuroWordNet are taken as a starting point [3].

Outside the lexicon, an ontology will provide a third layer of meaning. The Terms in an ontology represent the distinct types in a formal representation of knowledge. Terms can be combined in a knowledge representation language to form expressions of axioms. In principle, meaning is defined in the ontology independently of language but according to the principles of logic. In Cornetto, the ontology represents an independent anchoring of the relational meaning in Wordnet. The ontology is a formal framework that can be used to constrain and validate the implicit semantic statements of the lexical semantic structures, both the lexical units and the synsets. In addition, the ontology provides a mapping of a vocabulary to a formal representation that can be used to develop semantic web applications.

In addition to the 3 data collections, a separate table of so-called Cornetto Identifiers (CIDs) is provided. These identifiers contain the relations between the lexical units and the synsets in the CDB but also to the original word senses and synsets in the RBN and DWN.

The Figure 1 shows an overview of the different data structures and their relations. The different data can be divided into 3 layers of resources, from top to bottom:

- The RBN and DWN (at the top): the original databases from which the data are derived;
- The Cornetto database (CDB): the ultimate database that will be built;
- External resources: any other resource to which the CDB will be linked, such as the Princeton Wordnet, wordnets through the Global Wordnet Association, Wordnet domains, ontologies, corpora, etc.

The center of the CDB is formed by the table of CIDs. The CIDs tie together the separate collections of LUs and Synsets but also represent the pointers to the word meaning and synsets in the original databases: RBN and DWN and their mapping relation.

Furthermore, the LUs will contain semantic frame representation. The frame elements may have co-indexes with Synsets from the wordnet and/or with Terms from the ontology. This means that any semantic constraints in the frame representation can directly be related to the semantics in the other collections. Any explicit semantic relation that is expressed through a frame structure in a LU can also be represented as a conceptual semantic relation between Synsets in the Wordnet database.

The Synsets in the wordnet are represented as a collection of synonyms, where each synonym is directly related to a specific LU. The conceptual relations between Synsets are backed-up by a mapping to the ontology. This can be in the form of an equivalence relation or a subsumption relation to a Term or an expression in a knowledge representation language.

Finally, a separate equivalence relation is provided to one ore more synsets in the Princeton Wordnet.

The work is divided in 4 steps:

1. Automatic alignment of the word meanings of the two resources
2. Import of the result of the alignment into the database
3. Import of the SUMO ontology and WordNet domains to the synsets of the Dutch wordnet
4. Manual revision of the lexical units, the synsets and the ontological mapping

In the next paragraphs, we will discuss these steps briefly.

2.2 Aligning Word Meanings

To create the initial database, the word meanings in the Referentie Bestand Nederlands (RBN) and the Dutch part of EuroWordNet (DWN) have been automatically aligned. The word *koffie* (coffee) for example has 2 word meanings in RBN (*drink* and *beans*) and 4 word meanings in DWN (*drink, bush, powder* and *beans*). When we try to automatically align these meanings, we can get a complete match, no match or a partial match between these meanings. This then results in 4, 5, or 6 distinct meanings in the Cornetto database depending on

the degree of matching across these meanings. Note that this alignment is different from aligning WordNet synsets because RBN is not structured in synsets. We can for example not use the overlap of synonyms because RBN has no synonyms. For measuring the match, we used all the semantic information that was available in both resources: e.g. definitions and domain labels.

To match word meanings with the same domain label, we first had to normalize the labels. We first cleaned the labels manually (e.g., *pol* and *politiek* can be merged). Next, we measured the overlap in vocabulary associated with each domain. So if the label *oorlog* (war) in RBN is associated with the same words as the label *geweld* (violence) in DWN, we can make these labels equivalent. The overlap was expressed using a correlation figure for each domain in the matrix with each other domain. Domain labels across DWN and RBN do not require an exact match. Instead, the scores of the correlation matrix can be used for associating them.

Overlap of definitions was based on the overlapping normalized content words relative to the total number of content words. For other features, such as part-of-speech, we manually defined the relations across the resources. We only consider a possible match between words with the same orthographic form and the same part-of-speech. The strategies used to determine which word meanings can be aligned are:

1. The word has one meaning and no synonyms in both RBN and DWN
2. The word has one meaning in both RBN and DWN
3. The word has one meaning in RBN and more than one meaning in DWN
4. The word has one meaning in DWN and more in RBN
5. If the broader term (BT or hypernym) of a set of words is linked, all words which are under that BT in the semantic hierarchy and which have the same form are linked
6. If some narrow term (NT or hyponym) in the semantic hierarchy is related, siblings of that NT that have the same form are also linked.
7. Word meanings that have a linked domain, are linked
8. Word meanings with definitions in which one in every three content words is the same (there must be more than one match) are linked.

Each of these heuristics will result in a separate score for all possible mappings between word meanings. In the case of *koffie* (coffee), we thus will have 8 possible matches: RBN1-DWN1, RBN1-DWN-2, RBN1-DWN-3, RBN1-DWN4, . . . , etc, . . . RBN2-DWN-4. For the match RBN meaning 1-DWN meaning 1, we will thus get 8 scores, one for each heuristics. The number of links found per strategy is shown in the Table 1.

To weigh the heuristics, we manually evaluated each heuristics. Of the results of each strategy, a random sample was made of 100 records (800 samples in total). Each sample was checked by 8 persons (6 staff and 2 students). For each record, the word form, part-of-speech and the definition was shown for both RBN and DWN (taken from VLIS). The testers had to determine whether the definitions described the same meaning of the word or not. The results of the tests were averaged, resulting in a percentage of items which were considered good links. The averages per strategy are shown in the Table 1.

Table 1. Results for aligning strategies

	Conf.	Dev.	Factor	LINKS	
1: 1 RBN & 1 DWN meaning, no synonyms	97.1	4.9	3	9936	8.1 %
2: 1 RBN & 1 DWN meaning	88.5	8.6	3	25366	20.8 %
3: 1 RBN & >1 DWN meaning	53.9	8.1	1	22892	18.7 %
4: >1 RBN & 1 DWN meaning	68.2	17.2	1	1357	1.1 %
5: overlapping hypernym word	85.3	23.3	2	7305	6.0 %
6: overlapping hyponyms	74.6	22.1	2	21691	17.7 %
7: overlapping domain-clusters	70.2	15.5	2	11008	9.0 %
8: overlapping definition words	91.6	7.8	3	22664	18.5 %

The minimal precision is 53.9 and the highest precision is 97.1. Fortunately, the low precision heuristics also have a low recall. On the basis of these results, the strategies were ranked: some were considered very good, some were considered average, and some were considered relatively poor. The ranking factors per strategy are:

- Strategies 1, 2 and 8 get factor 3
- Strategies 5, 6 and 7 get factor 2
- Strategies 3 and 4 get factor 1

A factor 3 means that it counts 3 times as strong as factor 1. It is thus considered to be a better indication of a link than factor 2 and factor 1, where factor 1 is the weakest score. The ranking factor is used to determine the score of a link. The score of the link is determined by the number of strategies that apply and the ranking factor of the strategies. The final score is normalized to a value between 0 and 1.

In total, 136K linking records are stored in the Cornetto database. Within the database, only the highest scoring links are used to connect WordNet meanings to synsets. There are 58K top-scoring links, representing 41K word meanings. In total 47K different RBN word meanings were linked, and 48K different VLIS/DWN word meanings. 19K word meanings from RBN were not linked, as well as 59K word meanings from VLIS/DWN. Note that we considered here the complete VLIS database instead of DWN. The original DWN database represented about 60% of the total VLIS database. VLIS synsets that are not part of DWN can still be useful for RBN, as long as they ultimately get connected to the synset hierarchy of DWN.

As a result of the alignment, a new list of lexical units and synsets is generated. All the relevant data for these lexical units and synsets are copied from the RBN and DWN, respectively.

2.3 Importing External Data

DWN was linked to WordNet 1.5. WordNet domains are mapped to WordNet 1.6 and SUMO is mapped to WordNet 2.0 (and most recently to WordNet 2.1). In

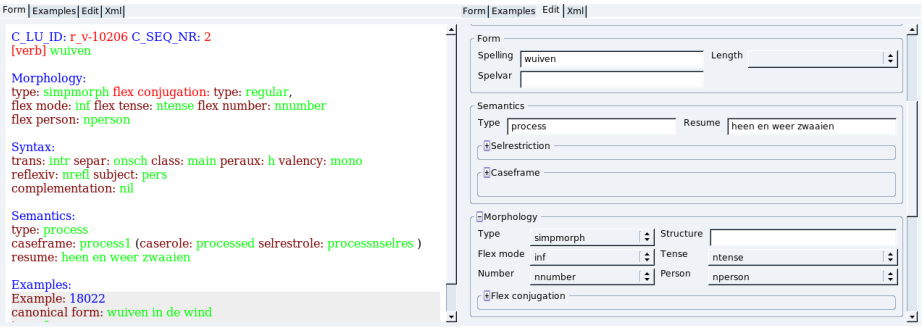

Fig. 2. Cornetto Lexical Units, showing the preview and editing form

order to apply the information from SUMO and WordNet domains to the synsets, we need to exploit the mapping tables between the different versions of Wordnet. We used the tables that have been developed for the MEANING project [8,9]. For each equivalence relation to WordNet 1.5, we consulted a table to find the corresponding WordNet 1.6 and WordNet 2.0 synsets, and via these we copied the mapped domains and SUMO terms to the Dutch synsets.

The structure for the Dutch synsets thus consists of:

- a list of synonyms
- a list of language internal relations
- a list of equivalence relations to WordNet 1.5 and WordNet 2.0
- a list of domains, taken from WordNet domains
- a list of SUMO mappins, taken from the WordNet 2.0 SUMO mapping

The structure of the lexical units is fully based in the information in the RBN. The specific structure differs for each part of speech. At the highest level it contains:

- orthographic form
- morphology
- syntax
- semantics
- pragmatics
- examples

The above structure is defined for single word lexical units. A separate structure will be defined later in the project for multi-word units. It will take too much space to explain the full structure here. We refer to the Cornetto website [10] for more details.

2.4 Manual Editing

The aligned data is further manually edited through various cycles of editing. For this purpose, special editing clients have been developed. We will discuss the editing clients in more detail below.

The editing process itself consists of a number of steps, where we will focus on different types of information. In the first cycle, we will manually verify the alignment of word-meanings. For this purpose, selections of words and word meanings are made. This selection involves the following criteria:

- Frequent nouns and verbs
- Words with many meanings
- Lexical units with a mapping to a synset with a low score
- Lexical units without a mapping with a synset

During this work, we typically carry out the following actions:

- Confirm or delete a mapping
- Create another mapping
- Split a single lexical unit in two lexical units
- Merge two lexical units into one
- Add lexical units or delete lexical units
- Split a synset unit in two synsets
- Merge two synsets into one
- Add synsets or delete synsets
- Add or delete synonyms to synsets

At the end of these actions, we will get a new and revised list of senses and mappings to synsets. This will be the new sense and synset structure of the Cornetto database.

The second phase of the editing involves the relation of the synsets to the ontology. The initial mapping is based on a projection of the SUMO labels to the synsets via the equivalence relations. These assignments will be revised, where we foresee two possible relations:

- a synset is a name for a SUMO term: there is direct equivalence
- a synset is defined through a KIF expression [11] that involves one or more SUMO terms.

In the last case, the synset does not name a disjunct ontological type but a lexicalization of a certain conceptualization of such a type or relation between types. For example, the next Dutch words do not require creating a new type in the ontology but will be defined as instances of the type Water in or used for specific purposes:

- zwemwater = water that is good for swimming
- drinkwater = water that is good for drinking
- zeewater = water from the sea
- rivierwater = water from a river
- theewater = water for making tea
- koffiewater = water for making coffee
- bluswater = water that is or can be used for extinguishing fire

The precise semantic implications for these concepts will be expressed by a list of triplet relations in the database, as in the following KIF expression for *rivierwater*:

```
(instance, 0, Water)
(instance, 1, River)
(origin, 0, 1)
```

This expression can be paraphrased as "there is an instance of Water and there is an instance of River such that the former originates from the latter." The numbers in this expression represent variables and we assume that the variable 0 corresponds to the referent of the defined synset.

During this phase, it may also be necessary to revise the hypernym relations in DWN to form a proper semantic hierarchy that is in line with the ontological decisions. During this phase, we also will formulate constraints on ontological mappings and synset relations. These constraints will be applied to all the assigned ontology relations. Any violation will be flagged and edited. Violations can follow from direct assignments or from assignments that are inherited downward through hyponymy relations.

In the final phase of the project, the editing will focus on the correlations between the frame-structures in the lexical units and the synsets. This process is called micro-level alignment. The frame structures for the lexical units, specify argument slots for verbs. These slots are now specified using semantic labels that are defined in RBN. In addition, we provided positions for pointers to synsets and pointers to SUMO labels. An example for a case frame for *genezen* (to cure) is given below:

```
<semantics_verb>
  <sem-type>action</sem-type>
  <sem-caseframe>
    <caseframe>action2</caseframe>
    <args>
      <arg>
        <caserole>agent</caserole>
        <selrestrole>agentanimate</selrestrole>
        <synset_list/>
      </arg>
      <arg>
        <caserole>theme</caserole>
        <selrestrole>themenselres</selrestrole>
        <synset_list/>
      </arg>
      </args>
  </sem-caseframe>
  <sem-resume>beter maken</sem-resume>
</semantics_verb>
```

The correlations with synsets need to be created manually. They need to be compatible with the given semantic labels and with other role relations that are listed in DWN and the matched ontological process.

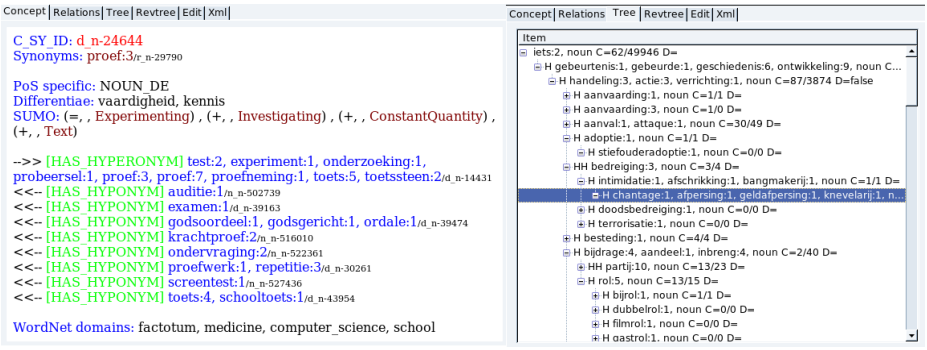

Fig. 3. Cornetto Synsets window, showing a preview and a hyperonymy tree

3 The DEB Platform

The Dictionary Editor and Browser platform [12,1] offers a development framework for any dictionary writing system application that needs to store the dictionary entries in the XML format structures. The most important property of the system is the *client-server* nature of all DEB applications. This provides the ability of distributed authoring teams to work fluently on one common data source. The actual development of applications within the DEB platform can be divided into the server part (the server side functionality) and the client part (graphical interfaces with only basic functionality). The server part is built from small parts, called *servlets*, which allow a modular composition of all services. The client applications communicate with servlets using the standard HTTP web protocol.

For the server data storage the current database backend is provided by the Berkeley DB XML [13], which is an open source native XML database providing XPath and XQuery access into a set of document containers.

The user interface, that forms the most important part of a client application, usually consists of a set of flexible forms that dynamically cooperate with the server parts. According to this requirement, DEB has adopted the concepts of the Mozilla Development Platform [14]. Firefox Web browser is one of the many applications created using this platform. The Mozilla Cross Platform Engine provides a clear separation between application logic and definition, presentation and language-specific texts.

3.1 New DEB Features for the Cornetto Project

During the Cornetto project the nature of the Cornetto database structure has imposed the need of several features that were not present in the (still developing) DEB platform. The main new functionalities include:

– *entry locking* for concurrent editing. Editing of entries by distant users was already possible in DEB, however, the exclusivity in writing to the same dictionary item was not controlled by the server. The new functions offer the

entry locking per user (called from the client application e.g. when entering the edit form). The list of all server locks is presented in the DEB administration interface allowing to handle the locks either manually or automatically on special events (logout, timeout, loading new entry, ...).
- *link display preview caching.* According to the database design that (correctly) handles all references with entity IDs, each operation, like structure entry preview or edit form display, runs possibly huge numbers (tens or hundreds) of extra database queries displaying text representations instead of the entity ID numbers. The drawback of this compact database model is in slowing down the query response time to seconds for one entry. To overcome this increase of the number of link queries, we have introduced the concept of *preview caching*. With this mechanism the server computes all kinds of previews in the time of saving a modified entry in special entry variables (either XML subtags or XML metadata). In the time of constructing the preview or edit form, the linked textual representations are taken from the preview caches instead of running extra queries to obtain the computed values.
- *edit form functionalities* – the lexicographic experts within the Cornetto project have suggested several new user interface functions that are profitable for other DEB-based projects like collapsing of parts of the edit form, entry merging and splitting functions or new kinds of automatic inter-dictionary queries, so called AutoLookUps.

All this added functionalities are directly applicable in any DEB application like DEBVisDic or DEBDict.

4 The New DEBVisDic Clients

Since one of the basic parts of the Cornetto database is the Dutch WordNet, we have decided to use DEBVisDic as the core for Cornetto client software. We have developed four new modules, described in more details below. All the databases are linked together and also to external resources (Princeton English WordNet and SUMO ontology), thus every possible user action had to be very carefully analyzed and described.

During the several months of active development and extensive communication between Brno and Amsterdam, a lot of new features emerged in both server and client and many of these innovations were also introduced into the DEBVisDic software. This way, each user of this WordNet editor benefits from Cornetto project.

The user interface is the same as for all the DEBVisDic modules: upper part of the window is occupied by the query input line and the query result list and the lower part contains several tabs with different views of the selected entry. Searching for entries supports several query types – a basic one is to search for a word or its part, the result list may be limited by adding an exact sense number. For more complex queries users may search for any value of any XML element or attribute, even with a value taken from other dictionaries (the latter is used mainly by the software itself for automatic lookup queries).

The tabs in the lower part of the window are defined per dictionary type, but each dictionary contains at least a preview of an entry and a display of the entry XML structure. The entry preview is generated using XSLT templates, so it is very flexible and offers plenty of possibilities for entry representation.

4.1 Cornetto Lexical Units

The Cornetto foundation is formed by Lexical Units, so let us describe their client package first. Each entry contains complex information about morphology, syntax, semantics and pragmatics, and also lots of examples with complex substructure. Thus one of the important tasks was to design a preview to display everything needed by the lexicographers without the necessity to scroll a lot. The examples were moved to separate tab and only their short resumé stayed on the main preview tab.

Lexical units also contain semantic information from RBN that cannot be published freely because of licensing issues. Thus DEBVisDic here needs to differentiate the preview content based on the actual user's access rights.

The same ergonomic problem had to be resolved in the edit form. The whole form is divided to smaller groups of related fields (e.g. morphology) and it is possible to hide or display each group separately. By default, only the most important parts are displayed and the rest is hidden.

Another new feature developed for Cornetto is the option to split the edited entry. Basically, this function copies all content of edited entry to a new one. This way, users may easily create two lexical units that differ only in some selected details.

Because of the links between all the data collections, every change in lexical units has to be propagated to Cornetto Synsets and Identifiers. For example, when deleting a lexical unit, the corresponding synonym has to be deleted from the synset dictionary.

4.2 Cornetto Synsets

Synsets are even more complex than lexical units, because they contain lots of links to different sources – links to lexical units, relations to other synsets, equivalence links to Princeton English WordNet, and links to the ontology.

Again, designing the user-friendly preview containing all the information was very important. Even here, we had to split the preview to two tabs – the first with the synonyms, domains, ontology, definition and short representation of internal relations, and the second with full information on each relation (both internal and external to English Wordnet). Each link in the preview is clickable and displays the selected entry in the corresponding dictionary window (for example, clicking on a synonym opens a lexical unit preview in the lexical unit window).

The synset window offers also a tree view representing a hypernym/hyponym tree. Since the hypero/hyponymic hierarchy in Wordnet forms not a simple tree but a directed graph, another tab provides the reversed tree displaying links

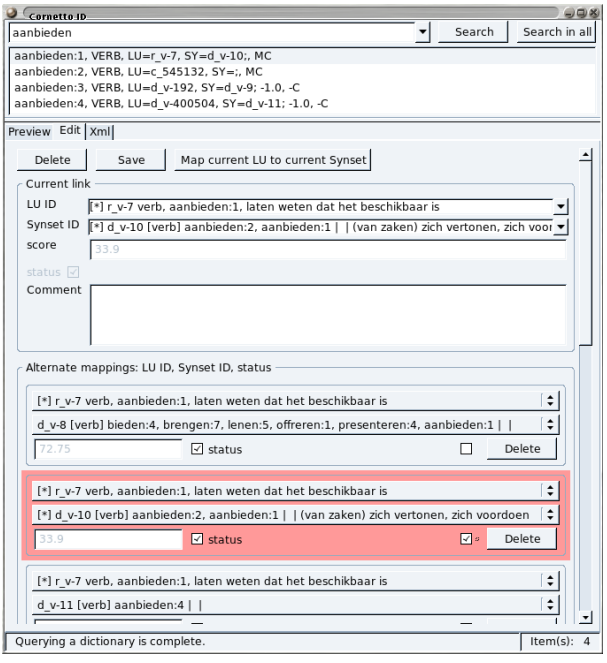

Fig. 4. Cornetto Identifiers window, showing the edit form with several alternate mappings

in the opposite direction (this concept was introduced in the VisDic Wordnet editor). The tree view also contains information about each subtree's significance – like the number of direct hyponyms or the number of all the descendant synsets.

The synset edit form looks similar to the form in the lexical units window, with less important parts hidden by default. When adding or editing links, users may use the same queries as in dictionaries to find the right entry.

4.3 Cornetto Identifiers

The lexical units and synsets are linked together using the Cornetto Identifiers (CID). For each lexical unit, the automatic aligning software produced several mappings to different synsets (with different score values). At the very beginning, the most probable one was marked as the "selected" mapping.

In the course of work, users have several ways for confirming the automatic choice, choosing from other offered mapping, or creating an entirely new link. For example, a user can remove the incorrect synonym from a synset and the corresponding mapping will be marked as unselected in CID. Another option is to select one of the alternate mappings in the Cornetto Identifiers edit form. Of course, this action leads to an automatic update of synonyms.

The most convenient way to confirm or create links is to use *Map current LU to current Synset* function. This action can be run from any Cornetto client

package, either by a keyboard shortcut or by clicking on the button. All the required changes are checked and carried out on the server, so the client software does not need to worry about the actual actions necessary to link the lexical unit and the synset.

4.4 Cornetto Ontology

The Cornetto Ontology is based on SUMO and so is the client package. The ontology is used in synsets, as can be seen in the Figure 3. The synset preview shows a list of ontology relations triplets – relation type, variable and variable or ontology term.

Clicking on the ontology term opens the term preview. A user can also browse the tree representing the ontology structure.

5 Conclusions

In the paper, we have described the Cornetto project workflow using the new lexicographic tools developed for this project. We have presented how a combination of automatic scored strategies with the human lexicographic work can be used for merging large databases of previous dictionaries to obtain a new qualitative language resource with complex morphological, syntactic and semantic information.

The presented project tools are, however, not a single purpose programs but they fit in the general framework of the Dictionary Editor and Browser (DEB) platform used for developing other publicly available language data tools.

Acknowledgments

The Cornetto project is funded by the Nederlandse Taalunie and STEVIN. This work has also partly been supported by the Ministry of Education of the Czech Republic within the Center of basic research LC536 and in the Czech National Research Programme II project 2C06009.

References

1. Horák, A., et al.: First version of new client-server wordnet browsing and editing tool. In: Proceedings of the Third International WordNet Conference - GWC 2006, Jeju, South Korea, Masaryk University, Brno, pp. 325–328 (2006)
2. Fillmore, C., Baker, C., Sato, H.: Framenet as a 'net'. In: Proceedings of Language Resources and Evaluation Conference (LREC 2004), Lisbon, ELRA, vol. 4, pp. 1091–1094 (2004)
3. Vossen, P. (ed.): EuroWordNet: a multilingual database with lexical semantic networks for European Languages. Kluwer Academic Publishers, Dordrecht (1998)
4. Maks, I., Martin, W., de Meerseman, H.: RBN Manual (1999)

5. Niles, I., Pease, A.: Linking lexicons and ontologies: Mapping WordNet to the suggested upper merged ontology. In: Proceedings of the IEEE International Conference on Information and Knowledge Engineering, pp. 412–416 (2003)
6. Cruse, D.: Lexical semantics. University Press, Cambridge (1986)
7. Miller, G., Fellbaum, C.: Semantic networks of english. Cognition (October 1991)
8. Daudé, J., Padró, L., Rigau, G.: Wordnet mappings, the Meaning project (2007), http://www.upc.es/~nlp/tools/mapping.html
9. Daudé, J., Padró, L., Rigau, G.: Validation and tuning of wordnet mapping techniques. In: Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2003), Borovets, Bulgaria (2003)
10. Vossen, P.: The Cornetto project web site (2007), http://www.let.vu.nl/onderzoek/projectsites/cornetto/start.htm
11. Genesereth, M.R., Fikes, R.E. (eds.): Knowledge Interchange Format, Version 3.0, Reference Manual. Stanford University, Stanford CA, USA (1992), http://www-ksl.stanford.edu/knowledge-sharing/kif/
12. Horák, A., et al.: New clients for dictionary writing on the DEB platform. In: DWS 2006: Proceedings of the Fourth International Workshop on Dictionary Writings Systems, Italy, pp. 17–23. Lexical Computing Ltd, U.K (2006)
13. Chaudhri, A.B., Rashid, A., Zicari, R. (eds.): XML Data Management: Native XML and XML-Enabled Database Systems. Addison-Wesley, Reading (2003)
14. Feldt, K.: Programming Firefox: Building Rich Internet Applications with Xul. O'Reilly (2007)