Proceedings of the 1st International WordNet Conference, January 21-25, 2002, Mysore, India, pag. 293-302

Emanuele Pianta, Luisa Bentivogli, Christian Girardi

ITC-irst
Centro per la Ricerca Scientifica e Tecnologica
Via Sommarive, 18
I-38050 Povo – Trento, Italy
{pianta, bentivo, cgirardi}@itc.it

MultiWordNet

Developing an aligned multilingual database

Abstract

This paper illustrates the MultiWordNet project, aimed at producing an Italian WordNet strongly aligned with the Princeton WordNet. The main conceptual differences between the MultiWordNet and the EuroWordNet conceptual models are presented first. Then two automatic procedures capable of speeding up the work of lexicographers are described. Finally, we give some details about the adopted data model and we present a graphical user interface that can be used to browse and update the aligned database.

1 Introduction

MultiWordNet is an on-going project at ITC-irst aiming at producing an Italian WordNet strictly aligned with Princeton WordNet (PWN), see Fellbaum (1998). In its first version, it contains around 37,000 Italian words organized into some 28,000 synsets, along with information about the correspondence between Italian and English (PWN) synsets. MultiWordNet adopts a methodological framework distinct from EuroWordNet.

There are at least two models for building a multilingual wordnet. The first model, adopted within the EuroWordNet project, consists of building language specific wordnets independently from each other, trying in a second phase to find correspondences between them (Vossen, 1998). The second model, adopted within MultiWordNet (MWN), consists of building language specific wordnets keeping as much as possible of the semantic relations available in the Princeton WordNet (PWN). This is done by building the new synsets in correspondence with the PWN synsets, whenever possible, and importing semantic relations from the corresponding English synsets; i.e., we assume that if there are two synsets in PWN and a relation holding between them, the same relation holds between the corresponding synsets in the new language.

According to Vossen (1996), the MWN model (or “expand model” in his words) seems less complex and guarantees the highest degree of compatibility across different wordnets. To see this, consider that building any wordnet necessarily implies a large number of subjective (and questionable) decisions. Thus, if two wordnets are built independently for two different languages, they will exhibit differences which depend only partially on divergences between the languages. Some non trivial structural discrepancies will in fact depend on subjective decisions or different building criteria. The MWN model minimizes these discrepancies by strictly adhering to the PWN building criteria and subjective choices.

The MWN model also has potential drawbacks. The most serious risk is that of forcing “an excessive dependency on the lexical and conceptual structure of one of the languages involved”, as Vossen (1996) points out. This risk can be considerably reduced by allowing the new wordnet to diverge, when necessary, from the PWN. Note that at least one wordnet within the EuroWordNet project, Spanish WordNet, is built following the "expand model" (Atserias et al., 1997).

2 Automatic procedures in the construction of MultiWordNet

Another important advantage of the MWN model is that automatic procedures can be devised to speed up both the construction of corresponding synsets and the detection of divergences between PWN and the wordnet being built. In all these procedures PWN itself can be used as a useful resource.

The construction of Italian WordNet, which is the first instantiation of the MWN model so far, is crucially based on two automatic procedures. The first is called the *Assign-procedure*. Given an Italian word sense, the procedure selects a weighed list of the most likely corresponding PWN synsets. This list is then used by lexicographers to actually build the Italian synsets. The second procedure supports the detection of *lexical gaps* (LG-procedure), which are cases when a lexical concept of a language is expressed through a free combination of words in another language (see below, Sect. 2.2).

Both these procedures use, as a crucial linguistic resource, the electronic version of the Collins bilingual dictionary. The bilingual Collins is a medium size dictionary, including 40,959 headwords and 60,901 translation groups in the English section, and 32,602 headwords and 46,545 translation groups in the Italian section. By *translation group* (TGR) we mean a group of *translation equivalents* (TEs) translating one of the senses of a source language word. In bilingual dictionaries, TGRs are usually separated by semicolons. We take them as the relevant sense unit as they correspond to WordNet senses. In the following example *wood* has 5 TGRs as a noun, and 2 TGRs as an adjective:

> **wood** [wUd] **1.** n **a**. *(material)* legno; *(timber)* legname (m) **b.** *(forest)* bosco **c.** *(Golf)* mazza di legno; *(Bowls)* boccia **2.** adj **a.** *(made of wood*) di legno **b**. *(living etc. in a wood*) di bosco, silvestre.

In this example only one TGR includes more than one TE, that is the second TGR of *wood* as an adjective (2.b), which can be translated by either *di bosco* or *silvestre*. Note that a TE can be a simple word (*bosco*) or a phrase (*di legno*), and that each TGR is introduced by a gloss illustrating the sense of *wood* which is being translated.

2.1 The Assign-procedure

Following the MWN model, our aim is to build, whenever possible, Italian synsets which are synonymous (semantically correspondent) with the PWN synsets. If this is not possible, then we have found an English-to-Italian or an Italian-to-English lexical idiosyncrasy.

Italian synonymous synsets can be built following different strategies. The first strategy is based on English-to-Italian TEs. For each PWN synset *S,* we look for the Italian TEs which are cross-linguistic synonyms of the English words of *S*. The union of such TEs is the Italian synonymous synset of S. If we cannot build any Italian synonymous synset for S, we have found an English-to-Italian lexical idiosyncrasy.

The second strategy is based on Italian-to-English TEs. For each sense σ of an Italian word *I*, we look for a PWN synset *S* including at least one English TE of *I* and we establish a link between *I* and *S*. When the procedure has been applied to all Italian word senses, we can build the equivalence class of all sets of Italian words which have been linked to the same PWN synset. Each set in the equivalence class is the Italian synset synonymous with some PWN synset. If, for a set of Italian synonyms there is no PWN synonymous synset, then we have found an Italian-to-English lexical idiosyncrasy.

The best alignment between Italian and Princeton WordNet can most likely be achieved by using both strategies and trying to cross-validate their results. As a matter of fact, so far we only exploited the Italian-to-English strategy. The same holds for Atserias et al. (1997).

Finding links between Italian word senses and PWN synsets is a complex and time consuming task, even if less complex and much quicker than building Italian synsets from scratch, organizing them in a semantic net, and putting them in correspondence with PWN synsets. For each Italian word sense, the lexicographer needs to look up its TEs in a bilingual dictionary, find all the synsets containing such TEs, to look carefully at the meaning of these synsets (synonyms, glosses, semantic relations), and finally to decide which synset is synonymous with the Italian word sense, if any. For certain word senses the lexicographer may need to consider tens of PWN synsets.

To help the lexicographer in her work we devised a procedure that selects, for each sense of an Italian word, the PWN synsets which are most likely to have a comparable meaning, if any. In the best case the procedure selects only the right candidate, and the lexicographer only needs to confirm the selection. In the worst case the procedure finds only wrong candidates or it cannot find any candidate, and the lexicographer has to do all the work manually. In the most common case the procedure finds a restricted set of candidates including the right one, and the lexicographer needs to confirm the right choice and to reject the wrong ones. In other words the algorithm helps the lexicographer to focus on the most promising PWN synsets.

The Assign-procedure takes as input one of the senses of the Italian-to-English section of the Collins dictionary and gives as output a set of candidates, each described by the pair *<PWN synset, confidence score>*, where *confidence score* (CS) measures the degree of confidence in the link between the Italian word sense and the *PWN synset*. Only candidates with a CS greater than a certain threshold are proposed to the lexicographer. Choosing such a threshold is a matter of balancing precision and recall. The greater the threshold, the lower is the probability that wrong candidates are proposed to the lexicographer (high precision), but also the greater the possibility that the right choice is not included in the set of candidates (low recall). See below for the actual evaluation of the algorithm.

For a certain word sense listed in the Italian-to-English dictionary, the Assign-procedure considers the group of English words which are proposed as TEs for that word sense, and finds all the synsets containing at least one such TE. Such synsets constitute the *set of candidates* (CandSet) to be linked with the input Italian word sense. We can rephrase the first step of the algorithm by saying that it computes the CandSet of a certain Italian word meaning. The rest of the algorithm consists of ordering the CandSet by calculating the CS of each of its synsets.

The ordering of the CandSet is based on a number of linking rules. Each rule, when successfully applied to a candidate, raises its CS. Note that the *partial CS* contributed by each rule varies according to factors specific to the rule. Besides the Collins dictionary, various resources may be accessed by the linking rules, such as an Italian monolingual dictionary, Italian nomenclatures, and PWN itself. Also, the Italian gloss that introduces almost all TGRs in the Collins dictionary plays a crucial role in determining the value of the CS.

We can group the linking rules into four main groups depending on the principle on which they are based: *generic probability*, *back translation*, *gloss matching*, and *synset intersection*.

A. Generic probability. The generic probability rule is based on the assumption that only one candidate in CandSet is the right target for the linking of an Italian word sense. As a consequence we can assume that the bigger the cardinality of the CandSet, the lower the probability that each candidate is the right one. The cardinality of the CandSet depends on the degree of ambiguity of the words which are proposed as TEs of the input word sense. If there is only one synset in the CandSet, this means that all the TEs of the input word sense are monosemous. Thus it is highly probable that the only synset in the CandSet is synonymous with the input word sense. Compare the monosemic criteria used by Atserias et al. (1997).

B. Back translation. The back translation rule exploits the following principle. Suppose we link a word sense to the correct target synset through a TE (linking-TE). Then it is probable that at least some of the PWN synonyms of the linking-TE have the input word as English-to-Italian TE. Take for example the Italian word "puntura". When referred to insects, the Collins gives *sting* as TE. However *sting* belongs to 4 synsets of PWN: {sting, stinging}, {pang, sting}, {sting, bite, insect bite}, {bunco, bunco game, ...}. Only the third synset is synonymous with the Italian word. If we look at the synonyms of *sting* in the third synset we find that the English-to-Italian section gives *puntura* as a (back) translation of *bite*. Summing up, the back translation rule considers the PWN synonyms of some linking-TE, and calculates a partial CS that is proportional to the number of synonyms that have the Italian word as English-to-Italian TE.

C. Gloss matching. A set of linking rules exploits the information contained in the Italian gloss that introduces almost all TGRs. The gloss may contain a semantic field specification (e.g. "**sclerosis *n*** *(Med)* sclerosi", where "*Med*" means medicine), a synonym (e.g. "**reason 1. *n* a.** *(motive, cause)* ragione,…"), a hypernym (e.g. "**sole *n*** *(fish)* sogliola"), or a specification of the context of use (e.g. "**handle 1. *n*** … (*of knife*) manico, impugnatura; (*of door, drawer*) maniglia;…"). This information can be used in different ways.

The information about the semantic field is exploited thanks to a resource which has been developed in parallel with MWN, the labelling of all PWN synsets with a semantic field label (see Magnini and Cavaglia', 2000). If the Italian gloss contains a semantic field label, and if this label matches the label attached to a synset in CandSet, then the candidate gets a partial CS. For instance the label *Elettr* contained in "**corrente *n*** ... (*Elettr, di acque*) current" matches the label *Electricity* attached to the synset {current, electric current}. Variations in the form of the labels are handled via a correspondence table.

When the Italian glosses contain words or phrases, we try a match between them and the words contained in PWN glosses. To do this, we extract the lemmas of the Italian and English words of the gloss, and we check whether one of the English words can be the TE of one of the Italian words. The strength of the matching depends on the ambiguity of the translation. The more polysemous the words the lower is the strength.

There are two extensions to this mechanism, based on the fact that glosses often specify the genus of the word they are defining, instead of a synonym. The first extension tries a match between an Italian word and the hypernym of its TE. The second mechanism tries a match between an Italian word and an English word contained in the gloss of a hypernym of the candidate synset. If the match between an Italian and English word is achieved through one of the indirect mechanisms, the partial CS will be lower than in the direct case.

A variant of the previous rule resorts to a more fine-grained analysis of the glosses. The Collins dictionary specifies the context of use of a word by following a restricted number of patterns. For instance to specify the usage of a noun, the Collins dictionary may use the pattern *of+noun*: see for instance "**piega *n*** ... (*della pelle*) fold ...", where *della pelle* means *of the skin*. A similar strategy is sometimes used by the PWN glosses, for instance: "{fold, plica} -- (a folded part (as a fold of skin or muscle))". Thanks to the parallelism between *della pelle* in the Italian gloss and *of skin* in the WN gloss we can enforce the linking between the Italian word sense and the PWN candidate.

This is done in fact by a specific linking rule, which executes a shallow parsing of the Collins and PWN glosses, isolates *of+noun* patterns and selects their nominal heads. To match the two heads the same translation-based technique explained in the previous paragraphs is applied.

Proceedings of the 1st International WordNet Conference, January 21-25, 2002, Mysore, India, pag. 293-302

D. Synset intersection. This rule exploits the fact that TGRs can include multiple TEs, which of course are synonymous. If one of the TEs is ambiguous we can use the other TEs to disambiguate. In practice the rule takes the different sets of candidates which are accessible through the different TEs, and intersects them. The synsets which are in the intersection get a partial CS. For instance, the Italian word *pilastro* is translated in its metaphorical sense as "pillar, mainstay". The word *pillar* belongs to 5 PWN synsets, whereas *mainstay* belongs to 3 synsets. However there is only one synset that includes both of them: "{pillar, mainstay} – a prominent supporter; 'he is a pillar of the community'". This synset gets a partial CS from the rule.

To assess the performance of the Assign-procedure, we carried out an evaluation based on the nouns listed under the letter D in the Italian-to-English section of the Collins dictionary (the letter has been chosen randomly). We took the number of TGRs in this part of the dictionary as an estimation of the number of word senses for which the algorithm should be able to find some candidate synsets. We selected the candidates with a confidence score higher than a fixed threshold, that is the candidates that were actually proposed to the lexicographers. The number of such candidates is 89% of the number of word senses listed in the Collins dictionary. Then, after the candidates were confirmed or rejected by the lexicographers, we calculated precision and recall of the candidates selected by the algorithm. The *precision* amounts to 70%, calculated as ratio between the number of candidates accepted by the lexicographers and the number of candidates proposed by the algorithm. The *recall* amounts to 63%, calculated as ratio between the number of candidates accepted by the lexicographers and the number of word senses listed in the Collins dictionary.

2.2 The Lexical Gaps-procedure

The literature on contrastive analysis shows that, given a source and a target language, various types of idiosyncrasies can occur at the lexical level. However, only some of them are relevant to the information coded in MWN, which strictly follows the PWN building criteria. In MWN, a synset of a language *L1* containing lexical units $w_1, \ldots, w_n$ has a correspondent in another language *L2* if there are one or more lexical units in *L2* which are cross-language synonyms of $w_1, \ldots, w_n$. As a consequence, only two kinds of idiosyncrasies imply a lack of cross-language correspondence in MWN (Bentivogli and Pianta, 2000):

- *lexical gaps*: a language expresses through a lexical unit what the other language expresses with a free combination of words (borrower = *chi prende in prestito*) (Hutchins and Somers, 1992). Following the MWN building criteria only idioms and restricted collocations are considered lexical units and thus can be synonymous with simple or compound words. On the contrary, a free combination of words is not a lexical unit and thus implies a missing synset for that language.
- *denotation differences*: the TE of a source language exists but it is more general or more specific. In the former case the TE is a sort of cross-linguistic hypernym of the source language word and in the latter case it is a cross-linguistic hyponym (bell ≅ (small/electric bell) *campanello* + (church bell) *campana* + (on cats) *sonaglio*).

During the construction of Italian WordNet we developed a procedure for identifying lexical gaps in a semi-automatic way. The procedure is based on the distinction between idioms and restricted collocations on the one hand and free combinations of words (which imply gaps) on the other hand. In practice, the boundaries between idioms, restricted collocations and free combinations of words are not clear-cut. However, in many cases a distinction can be drawn by relying on knowledge contained in dictionaries that explicitly mark idioms and restricted collocations. Also, the three groups exhibit certain structural regularities that can be exploited to automatically distinguish them from each other with a certain degree of confidence. For further details see Bentivogli et al. (2000).

The LG-procedure automatically classifies all TGRs of the Collins bilingual dictionary in three main classes: lexical units, lexical gaps and TGRs that need to be manually checked and classified as lexical units or lexical gaps. The results are the following:

	Lexical units (%)	Lexical gaps (%)	Manual check (%)
English-to-Italian	88.4	1.0	10.6
Italian-to-English	92.1	0.9	7.0

Information about lexical gaps can be used in two ways, depending on whether we are dealing with Italian-to-English gaps or vice versa. The Italian-to-English gaps point to a set of Italian synsets that need to be added manually in the Italian WN: we know for sure and from the beginning that such synsets cannot be built in correspondence to any English synset and thus their construction cannot be based on the results of the Assign-procedure. On the other hand, information about English-to-Italian gaps point to PWN synsets which do not have a correspondent in Italian and can be excluded a priori from those selected by the Assign-procedure.

It is worthwhile to note that the procedure is also significant from a theoretical point of view. In fact it provides as a further result an approximate quantitative evaluation of lexical gaps, showing that the English and Italian lexica are highly comparable and thus giving strong empirical support to the MWN model.

3 The data model of MultiWordNet

The data model underlying the MWN database reflects the main theoretical assumptions of the MWN model. The database is based on the idea that there is a set of data which are common to all languages, and other data which are specific for each language. We take for granted that semantic relations (has-hypernym, has-part, entails, etc.) are common, whereas lexical relations (has-synonym, derives from, etc.) are language-specific. In the current implementation, which contains only English and Italian data, the semantic relations of PWN are contained in a module called COMMON-DB, whereas the lexical relations for Italian and English are contained in other two modules called ITALIAN-DB and ENGLISH-DB. In other words the information about which words belong to which synsets is contained in language-DBs, whereas the information about the relations between synsets which hold for all languages is contained in COMMON-DB. Another crucial piece of information, that is the cross-language correspondence between synsets, is realized by using the same synset identifier in the different languages. All the synsets of different languages which have the same identifier belong in fact to the same *multisynset*. COMMON-DB describes the relations between the multisynsets of MWN. Note that all semantic information which is language-independent can be added to the COMMON-DB. This is in fact what has been done with the semantic field information (see above, Sect. 2.1).

In the above description, we showed how the MWN data model represents the fact that different languages share a good deal of information at the conceptual level. However the data model needs also to represent conceptual divergences between languages (e.g. lexical gaps). Moreover, even if we take the PWN semantic relations as the basis for the COMMON-DB, we want to be able to add new semantic relations or to modify existing ones. The possibility of modifying the PWN semantic relations and of representing conceptual idiosyncrasies in specific languages has been implemented by resorting to add-on modules which overwrite (without physically changing) the original PWN data. The COMMON-DB in fact contains all the original PWN semantic relations plus a COMMON-ADD-ON overwriting part of them. Also, each language-DB contains a language-ADD-ON specifying the semantic relations which are idiosyncratic to that language. Fig. 1 summarises the main features of the MWN data model. The arrows represent the overwriting relations. Within the COMMON-DB, PWN data are overwritten by the COMMON semantic ADD-ON, whereas the COMMON-DB is overwritten by the semantic ADD-ON of each language-DB.

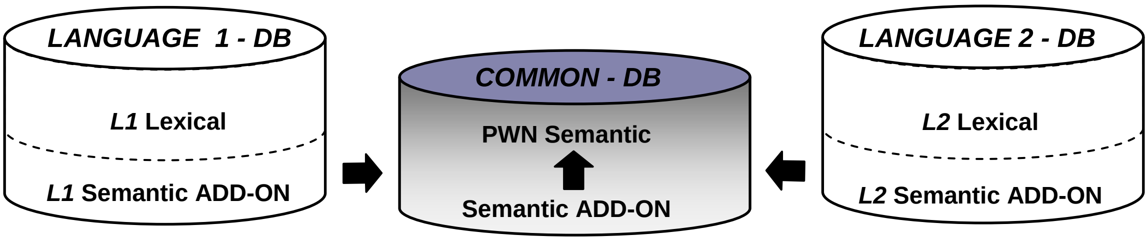

Figure 1:The MultiWordNet data model.

Lexical idiosyncrasies are the typical kind of information which is coded in the language-specific add-ons. If there is positive lexicographic evidence that a certain lexical concept is missing in one language, a special empty-node label substitutes the synonyms in the lexical section of the language-DB. Then, two different strategies are followed to represent denotation differences vs. lexical gaps (see Sect. 2.2). If the empty node corresponds to a denotation difference, one or more *nearest relations* link the empty node to one more general or to many more specific synsets. If the empty node corresponds to a lexical gap, an appropriate translating paraphrase (introduced by the *TE* keyword) is reported in the gloss of the empty node. The *nearest* relations are included in the language-specific add-ons.

Each language-DB contains also a module with lexicographic information about the (possibly wrong) link between word senses and synsets. See the following section for more details on this.

The following table reports the data concerning the first version of Italian WordNet, which contains the most frequent words in the Italian lexicon.

	Nouns	Verbs	Adjectives	Adverbs	Total
Word senses	37,235	8,296	4,511	1,805	51,847
Words	26,463	4,414	4,220	1,417	36,514
Synsets	20,571	4,130	2,413	1,006	28,120

As regards relations, all common semantic relations are imported from PWN and are available as well as the *nearest* relations, i.e. the new semantic language-specific relations that have been added in MWN to represent denotation differences. On the contrary, synonymy is the only lexical relation instantiated so far.

At the application level, procedures have been implemented for data consistency checking, encrypting, and statistics reporting. The MWN data model has been implemented both through a specialized database written in Lisp and as MySql tables. The LISP implementation is used to develop the resource and supports multi-user concurrent access and updating, whereas the MySql solution increases browsing efficiency. Both implementations run under both the UNIX and Windows environments.

4 The graphical interface

A graphical user interface has been designed to help the user in browsing and possibly editing MWN multisynsets. The graphical interface, implemented in Tcl/Tk, acts as a client of the central server and manages all MWN information. Two access modes are defined: the *browse mode* allows an end-user to access all information in the database, excluding the lexicographer cards, whereas the *edit mode* gives access to all available data, and allows the lexicographer to modify and add new data.

The interface (see Fig. 2) consists of three main areas: the search area (1), the synonym area (2), and the relation area (3).

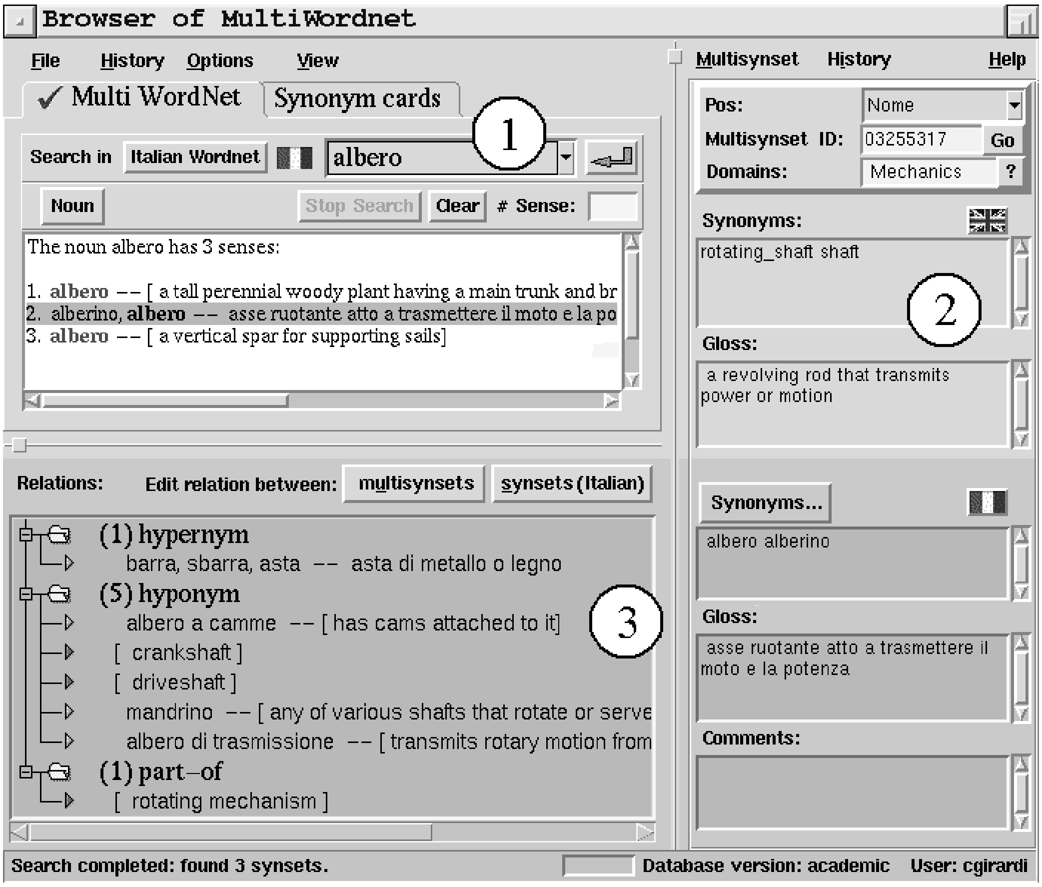

Figure 2: The MultiWordNet interface.

In the *search area* (1) the `MultiWordNet` tabbed pane (ticked off in the figure) has the same functionalities as the PWN interface, including the ability to search by substring. Moreover, unlike the PWN interface, the MWN interface requires the user to select the language of the word she is searching (identified be a little flag near the input field). In addition, when a synset is selected in the display field, it becomes the *focus* of the interface and additional information about that synset is displayed in the other areas of the interface. The example in Fig. 2 shows also that whenever an Italian gloss is not available for an Italian synset, the English gloss of the corresponding PWN synset is displayed instead (in square brackets).

The *synonym area* (2) contains information about the multisynset to which the focus synset belongs. This area is in turn split up in three zones. The lower one contains the synset of the search language, i.e. synonyms, glosses, and comments. The middle zone contains the same information for another language (currently only English is available). The upper zone displays the information common to all the synsets of the multisynset, namely the identification number, the part of speech, and the semantic field labels. The synset area also allows access to lexicographic information. The `Synonyms...` button on top of the lower zone opens a `Synonym window` (see Fig. 3) in which all the lexicographic cards related to the focus synset can be accessed. Each lexicographer card includes an evaluation about the assignment of a word sense to a synset. The word sense may be identified by reference to either a bilingual dictionary or the DISC Italian monolingual dictionary. Information about register of use, examples of use, and comments can be added. Note that a card can also state that a certain word sense *should not* be assigned to a synset (see the `nn` flag). This kind of information is as relevant as the positive assignments for documenting lexicographic work.

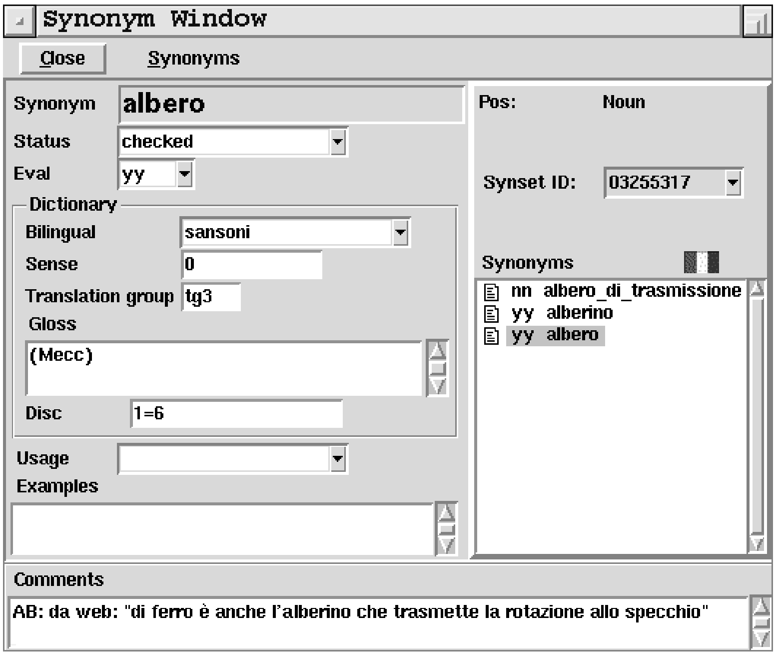

Figure 3: The synonym window.

In some cases, the lexicographer may want to access all the cards related to a certain word (instead of a synset as in the example above), independently from the senses and the linked synsets. To this extent a specific tabbed pane `Synonym cards` is available in the search area (see Fig. 4), as an alternative to the `MultiWordNet` pane described above. By specifying a word in the input field a list of candidate synsets is shown in the display field. Note also that negative assignments are accessible. From the described list the lexicographer can focus on a specific candidate synset as in the previous case.

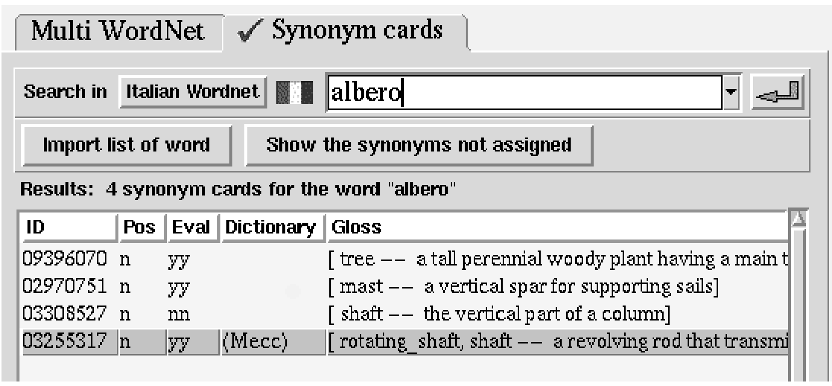

Figure 4: The synonym cards *tabbed pane.*

The third and last main area (3) of the MWN graphical interface displays information about the *semantic relations* of the focus synset. Whereas the `MultiWordNet` pane, following the

PWN approach, shows information about *one* semantic relation at a time, the relation area shows information about *all* the semantic relations of a synset.

Note that the interface gives access to the full MWN hierarchy. The users can also see relations from the focus synset to synsets which do not have any synonyms in the search language. In this case, the keyword `GAP!` is displayed in place of the synset. The `GAP!` keyword is used both to describe lexical gaps and denotation differences (see Sect. 2.2). When a `GAP!` node corresponds to a denotation difference the interface will display the *nearest relations* for that node. If the `GAP!` node corresponds to a lexical gap, the appropriate translating paraphrase is displayed in the gloss. Semantic relations that are idiosyncratic to the search language (such as the nearest relation) are marked by a little flag.

5 Conclusions

In this paper we have discussed a model of an aligned multilingual database distinguished from EuroWordNet. MultiWordNet stresses the usefulness of a strict alignment between wordnets of different languages, while retaining the ability to represent true lexical idiosyncrasies between languages. One of the biggest advantages of the MultiWordNet model is the possibility to use automatic procedures to speed up the lexicographic work needed to build the wordnet of a new language. The Princeton WordNet itself can be used as a crucial resource by these procedures.

A database model relying on MultiWordNet has been described, along with a graphical interface allowing users to browse and update the aligned database. Italian WordNet is the first instantiation of the MultiWordNet model. In its first version, it contains around 28,000 synsets and 52,000 word meanings.

References

Atserias J., Climent S., Farreres X., Rigau G. and Rodriguez H. (1997) *Combining multiple methods for the automatic construction of multilingual WordNets*. Proceedings of the International Conference "Recent Advances on Natural Language Processing" RANLP'97, Tzigov Chark, Bulgaria.

Bentivogli L., Pianta E. and Pianesi F. (2000) *Coping with lexical gaps when building aligned multilingual wordnets*. Proceedings of LREC 2000, Athens, Greece, pp. 993-997.

Bentivogli L. and Pianta E. (2000) *Looking for lexical gaps*. Proceedings of Euralex-2000 International Congress, Stuttgart, Germany.

Fellbaum C., ed., (1998) *WordNet: An electronic lexical database*. The MIT Press, Cambridge (Mass.).

Hutchins W.J. and Somers H.L. (1992) *An introduction to Machine Translation*. Academic Press, London.

Magnini B. and Cavaglia' G. *Integrating Subject Field Codes into WordNet*. Proceedings of LREC 2000, Athens, Greece, pp 1413-1418.

Vossen, P. (1996) *Right or wrong: combining lexical resources in the EuroWordNet project.* Proceedings of Euralex-96 International Congress.

Vossen, P., ed., (1998) *EuroWordNet: A multilingual database with lexical semantic networks*. Kluwer Academic, Dordrecht.

MultiWordNet homepage: http://multiwordnet.itc.it