Multilingual linguistic resources: from monolingual lexicons to bilingual interrelated lexicons

Marta Villegas†*, Nuria Bel†, Alessandro Lenci#, Nicoletta Calzolari#, Nilda Rumy#, Antonio Zampolli#, Teresa Sadurní*, Joan Soler*

† GILCUB (Grup Investigació Lingüística Computacional Universitat Barcelona)
{tona,nuria}@gilcub.es
#Istituto di Linguistica Computationale. CNR {lenci,glotollo,nilda,parole}@ilc.pi.cnr.it

†Institut d'Estudis Catalans {mvillegas,tsadurni,jsoler}@iec.es

Abstract

This paper describes a procedure to convert the PAROLE-SIMPLE monolingual lexicons into bilingual interrelated lexicons where each word sense of a given language is linked to the pertinent sense of the right words in one or more target lexicons. Nowadays, SIMPLE lexicons are monolingual although the ultimate goal of these harmonised monolingual lexicons is to build multilingual lexical resources. For achieving this goal it is necessary to automatise the linking among the different senses of the different monolingual lexicons, as the production of such multilingual relations by hand will be, as all tasks related with the development of linguistic resources, unaffordable in terms of human resources and time spent. The system we describe in this paper takes advantage of the SIMPLE model and the SIMPLE based lexicons so that, in the best case, it can find fully automatically the relevant sense-to-sense correspondences for determining the translational equivalence of two words in two different languages and, in the worst case, it will be able to narrow the set of admissible links between words and relevant senses. This paper also explores to what extent semantic encoding in already existing computational lexicons such as SIMPLE can help in overcoming the problems arisen when using monolingual meaning descriptions for bilingual links and aims to set the basis for defining a model for adding a bilingual layer to the SIMPLE model. This bilingual layer based on a bilingual relation model will be the basis indeed for defining the multilingual language resource we want PAROLE-SIMPLE lexicons to become.

1. Introduction

Re-utilization of existing lexical resources and automatic production of more information to enrich them so that these become the basis for a broad range of HLT applications is the main objective of the work presented in this paper. Thus, the objective was to study the feasibility of reusing SIMPLE monolingual semantic lexicons to build a multilingual lexical resource.

SIMPLE is a follow up of the PAROLE project (see www.ub.es/gilcub/SIMPLE/simple.html) that has added a semantic layer to the already existing morphological and syntactic layers developed by PAROLE, being these layers an harmonized common model for computational lexicons encoding relevant information. The semantic lexicons produced (about 10,000 semantic units for each of the 12 PAROLE languages) follow an harmonized common model that encodes structured semantic types and frames, linked to syntactic and morphological information.

The ultimate aim of the work we are reporting is to define a new layer of information that supplies a model for encoding word to word links paired via sense-to-sense correspondences between two, or more, monolingual computational lexicons. This model has to provide the means to create bilingual, in a first step, and multilingual, at the end, links among the words contained in the different lexicons. This paper is however mainly concerned with the procedures that will allow automatic creation of links among words based on their translational equivalence.

The starting point has been to profit of traditional bilingual dictionaries as they are the obvious and most extensive repository of bilingual knowledge. Being, though, for human consultation, the only information we should rely on are the word to word correspondences, as traditional bilingual dictionaries bear little systematic information about constraints on the input and target senses for these words to be related[1]. Thus an entry for the Spanish word *manzana* in a Spanish-Catalan bilingual dictionary may look like:

(1) manzana: 1. (*Fruit*) poma ('apple'), 2. (*of houses*) illa ('block')

Once having extracted the words which can be considered translational equivalents in at least one case, the key point is then to determine under what sense is this correspondence based, so as to consider the combination of 'word+sense' as an element of a fully translational equivalent pair for both languages.

The most obvious argument supporting the need for this sense identification is to ensure bi-directionality between bilingual dictionary entries. For example, while in (1) above we know that the correspondence *manzana-poma* is true bi-directionally, in the correspondence *manzana-illa* bi-directionality does not hold, as the Catalan entry *illa* can also refer to an island. This case of partial equivalents is the most frequent case in bilingual dictionaries, due to the polisemy of most words.

[1] Sometimes there is no information at all, or it is non-systematically expressed in terms of (i) a semantic descriptor or hyperonym; (ii) an example; (iii) a reference to a domain; etc.

The objective of our experiment is, hence, to evaluate to what extent the sense encoding done in SIMPLE, allows pairing the relevant sense, for example, of the Spanish word *manzana* with the relevant senses of one, or more than one, Catalan words (those that appear in the traditional bilingual dictionary) so that bi-directionality of the word-to-word correspondences is guaranteed when pointing to that sense.

(2) manzana → poma poma→ manzana
manzana →illa illa → manzana

Besides, closure of the different senses found for the involved lexical units must also be guaranteed. That is, we want to ensure that the Catalan word *illa* is also put in correspondence with the corresponding Spanish word *isla* under the pertinent sense (i.e. the corresponding 'island'):

(3) manzana →poma poma →manzana
manzana →illa illa → manzana
isla → illa illa → isla

Last but not least, when moving from monolingual descriptions to bilingual descriptions we have also had to deal with the well known differences between monolingual sense division and bilingual meaning discrimination. Thus, for instance, in a Catalan monolingual dictionary the 'food' meaning of *peix* ('fish') is considered a sub-sense inside the prime sense 'animal' as represented in (4), whereas in a bilingual Catalan-Spanish dictionary, the 'food' sense is promoted to be a sense, as represented in (5), and in a Catalan-English dictionary there is no reference at all to any 'food' meaning component, as represented in (6):

(4) Peix: 1 ….. 2 1 'name given to animals which exclusively live in water'. 2 'the meet and certain products of fishes'. 3 …
(5) peix: (*ichthyology*) pez || (*gastronomy*) pescado
(6) peix: fish || *pl*. (*zodiac*) Piscis

Summing up, the system we describe takes advantage of the SIMPLE model and lexicons in order to (i) discriminate senses involved in bilingual word-to-word equivalences as extracted from bilingual dictionaries in order to establish the correct correspondences between senses in two monolingual computational lexicons; and (ii) to provide means for moving (semi-)automatically from monolingual descriptions to bilingual descriptions by defining and modeling the set of admissible correspondences.

The ultimate goal of this exercise is to be able to define the relations that hold between fully translational equivalents made of 'word+sense', so that we can trace multilingual paths of such fully translational equivalents. This multilingual linking among so many languages would be almost impossible if not carried out automatically. Besides, for those correspondences that cannot be determined as fully translational equivalents, the model will have to provide descriptive mechanisms that allow to handle the meaning differences.

2. Description of the experiment

The input to the procedure are word-to-word correspondences extracted out of bilingual dictionaries, and the procedure output is modeled sense-to-sense correspondences among two monolingual computational lexicons. The sample data for this preliminary exercise only includes nouns and the words has been taken out of the 500 most frequent ones in Catalan (and their translations into Spanish and Italian). For this experiment we have dealt only with Catalan, Spanish and Italian word to word correspondences that have been extracted out of traditional bilingual dictionaries.

The system includes four modules: monolingual computational lexicons, word-to-word correspondences, set of candidate senses, and resolution algorithm as described below.

2.1. Monolingual computational lexicons

The system we describe takes advantage of the SIMPLE[2] model (Lenci et al., 2000) and SIMPLE lexicons. The SIMPLE model is based on the recommendations of the EAGLES[3] Lexicon/Semantic Working Group and on extensions of the Generative Lexicon theory. SIMPLE aims at capturing the various dimensions of word meaning and relies on an extension of 'qualia structure' (Pustejowsky 1995) for structuring the semantic/conceptual types as a representational aspect of word meaning. The semantic types in SIMPLE are defined as clusters of structured information and form a general ontology which is organized following the principles of orthogonal organization of types, as formalized in the Generative Lexicon. The SIMPLE model makes a crucial use of the Template notion which is introduced to satisfy two different needs: (i) making the encoding task more easy; and, (ii) enhancing the general consistency of the lexicons by providing structured sets of information. Templates are defined so as to mirror semantic types.

2.2. Word to Word Correspondences

Word-to-word correspondences (from now on WWC) are binary relations between the input word and the target word given by bilingual dictionaries as translational equivalents. These words are checked against PAROLE SIMPLE lexicons in order to extract all relevant information: (i) all morphological units whose `Lemma'[4] matches with the words extracted, plus (ii) the syntactic and semantic units linked to these morphological units. This means that the resulting modeled sense-to-sense correspondences (from now on, SSC) will eventually be independent of the criteria applied when establishing

[2] Semantic Information for Multifunctional Plurilingual Lexicons is a project sponsored by the EC DGXIII in the framework of the Language Engineering programme.
[3] http.//www.ilc.pi.cnr.it/EAGLES/rep2.
[4] Due to the different criteria used in traditional lexicons and in a multipurpose computational lexicon such as SIMPLE, we prefer to use 'word' as to refer to lexical units, and 'lemma' to what in SIMPLE is actually 'Spelling'.

lexical units in both paper bilingual dictionaries and PAROLE/SIMPLE lexicons. This is the case of the Catalan word *cort* (meaning 'pigsty' and 'royal court') which, due to etymological matters, has two different entries in paper dictionary but only one morphological unit in the PAROLE lexicon. As we will see, the system will establish the correct sense to sense correspondences despite the criteria applied.

For the sake of the exercise, WWCs only include the first translational equivalent supplied by the bilingual dictionary. Thus, for instance, despite of the fact that in the bilingual Catalan-Spanish dictionary the word *paraula* is assigned the Spanish words *palabra*, *vocablo*, *término* and *voz* as almost sinonymous, the set of WWC derived for the experiment only includes the first candidate *palabra*, which generally corresponds to the 'preferred' equivalent.

Two different WWCs sets are generated for each pair of involved languages, according to the ordering of being in one case source language, and in the other case target language. This is explained in terms of closure so as to cover all possible bi-directional correspondences, specifically in the case of disambiguating a one to many correspondences. That is, WWCs are to be bi-directional, thus if a Catalan word *X* is linked to the Spanish word *Y*, we know that the Spanish word *Y* will be also linked to the Catalan word *X*. We can see this by comparing the two examples below:

(7) CAT to SP WWCs: SP to CAT WWCs:
(*ala,ala*) ⟷ (*ala,ala*)

(8) CAT to SP WWCs: SP to CAT WWCs:
(poma,manzana) → *(manzana,poma)*
(manzana,illa)
(illa,manzana) ← *(illa,isla)*

In (7), the set of WWCs reflect the fact that the Catalan word *ala* ('win') is related to the Spanish word *ala*, and that this relation is fully bi-directional (for every sense involved). In (8), the Catalan word *poma* is related to the Spanish word *manzana*, but such word-to-word relation is not fully bi-birectional as the Spanish word *manzana* is also related with the Catalan word *illa*. In this case, the set of WWCs needs to include all (*illa,X*) WWCs in order to guarantee the closure of the system.

Note, however, that not all bi-directional WWCs are included in a first instance. This is the case of the example below:

(9) CAT to SP WWCs: SP to CAT WWCs:
(paraula,palabra) *(palabra,paraula)*
(mot,palabra) ...
...

In this case, the Spanish to Catalan WWC (*palabra,mot)* is not included, as such correspondence is not the 'preferred' one (the first candidate in the bilingual dictionary). As we will see, this does not mean that we obviate such correspondence but rather helps in typing correspondences in terms of [+/- preferred].

2.3. Set of involved senses

For each set of binary WWCs, we get the related senses for both all source and target words in PAROLE/SIMPLE lexicons. So, in the trivial examples in (7) and (8) above, the Sets of Involved Senses[5] include (10) and (11) respectively.

(10) Catalan Senses: Spanish Senses:
ala_BodyPart { *ala_BodyPart*
ala_Artifact { *ala_Artifact*
ala_Part { *ala_Part*

(11) Catalan Senses: Spanish Senses:
poma_Fruit { *manzana_Fruit*
illa_Area { *manzana_ArtifactualArea*
illa_ArtifactualArea

The objective is to discriminate and to model the correct correspondences between the input senses in bold (i.e. all the senses of the source word in the bilingual dictionary) and the target senses (i.e. all the senses of all target words in the bilingual dictionary). As we will see, in order to resolve the case of *manzana* and *poma/illa* in (3) above, in order to disambiguate which one of the two target senses (*Fruit* and *Artifactual Area*) correspond to the input sense, the system has to check all the material derived from closure. This aspect will be further explained in 2.4.

2.4. Resolution algorithm

Once we have all involved senses, the system has to establish the correct SSCs between the input senses and the target senses. Ideally the system should find a candidate for all input senses.

The system acts according to three different admissible scenarios:

Trivial Case (TC): involving one source sense and one target sense. In these cases, the system checks the correctness of the only candidate sense and the [+/- preferred] status of the correspondence. These trivial cases merely serve to determine the harmony between lexicons.

Non Trivial Case1 (NTC1): for cases involving one input sense and >1 target senses (for instance, the Catalan to Spanish WWC(*poma,manzana)* in (8) above). The objective consists on identifying the best candidate among the target senses and checking the [+/- preferred] status of the correspondence.

Non Trivial Case2 (NTC2): for cases involving >1 input senses and >1 target senses (for instance the Catalan to Spanish WWC(*ala,ala*) in (7) above). The objective in this case is to identify the relevant candidates for all source senses.

[5] For the sake of clarity, we represent 'senses' as *'word_Template'*

The system includes two main procedures: *Check best Candidate* and *Check Closure* defined as follows:

Check best Candidate is expressed in terms of ponderation according to the following terms:

1. Identity of Template Labels for source and target senses, as in the case of (12a, b, c) where the template labels are identified.

2. Subsumption relations between input and target Templates according to the SIMPLE ontological hierarchy they are structured. This will be the case for the Spanish *caballo_Animal* ('horse') being related to the Italian *cavallo_EarthAnimal*.

3. Identity of Semantic Type between input and target senses. This allows us to overcome differences in encoding between input and target lexicons as illustrated in the following example for the WWCs corresponding to the Spanish *cocina* and the Italian *culinária* and *cucina* (the senses refer to a: 'cooking', b: 'kitchen', c: 'cooker' and d: 'furniture of the kitchen')

(12)

Spanish Senses		Italian Senses
a. cocina1_Domain	→	*culinária1_Domain*
b. cocina2_Building	→	*cucina1_Building*
c. cocina3_Instrument	→	*cucina2_Instrument*
d. cocina4_Group	→	*cucina3_Furniture*
&[Sem_Type: *Furniture*]		&[Sem_Type: *Furniture*]

In (12d), despite the Template assigned to the Spanish sense, *cocina4* does not match with the Template assigned to the Italian *cucina3*, but the correspondence can be correctly established by means of the Semantic Type[6].

4. Subsumption relations between input and target Semantic Type. The hierarchy that structures Semantic Types also allows to determine that the Spanish *cuchillo_tool* ('knife') is to be related to the Catalan *ganivet_Instrument*.

5. Matching between features and Template information. As mentioned in 2.1, SIMPLE model includes a set of Template types organized as an orthogonal hierarchy. Templates are defined in terms of clusters of information describing the various dimensions of meaning participating in a given word sense. Sometimes, dimensions in word meaning compete and may derive into conflicting results. The system, therefore, predicts on these conflicting results by checking the different dimensions involved. Thus, for instance, we can predict that input and target senses defined as [+edible] are related despite the Template assigned in each case.

6. Checking co-occurrence of other additional features such as Domain, Connotation, etc. for disambiguation purposes. The system can also disambiguate correspondences by means of relevant features. Thus, in the following example the Italian sense *acqua2* ('amniotic liquid') is rejected as it bears the Domain feature Obstetrics:

(13)

Spanish Senses:	Italian Senses:
agua1_Substance	*acqua1_NatSubstance* *acqua2_NatSubstance* &[Domain: *Obstetrics*]

7. Syntactic Information. As the semantic units in SIMPLE are linked to the corresponding syntactic unit, the system can also have access to further information that can help in disambiguation, if required. This will be the case for (14). The Spanish *campo* ('field' or 'country') and the target Italian *campo*, where the system can add to the subsumption information, information regarding their countable nature to determine the SSC.

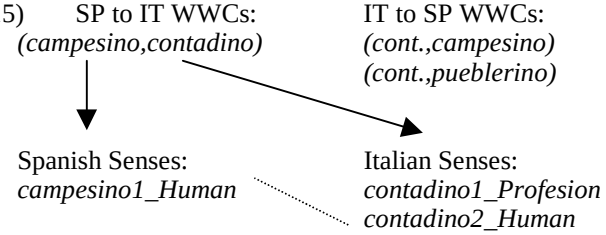
(14)

Spanish Senses:	Italian Senses:
campo_ArtifArea &[COUNTABLE] *campo_Location* &[UNCOUNTABLE]	*campo_Area* &[COUNTABLE]

Check Closure of WWCs serves a double purpose: first, it allows to discriminate between 'preferred' correspondences and 'non-preferred' correspondences as described above and, second, it is required for disambiguation when a one to many relations or many to many relations are possible.

A combination of all these checking is what allows the resolution of the the Spanish to Italian WWC (*campesino,contadino*) ('peasant') as in (15). In this example, if closure of WWC were obviated, we would have one input sense and two target senses and, according to the Template Identity in 1 above, the Spanish *campesino1_Human* sense would be wrongly related to the Italian *contadino2_Human* ('yokel') sense:

(15)

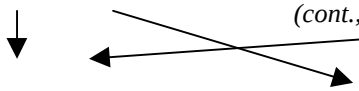


Note, however, that if senses derived from closure are included, we obtain the following Set of Involved Senses:

[6] Additional information will be represented here by brackets.

(16) SP to IT WWCs: *(campesino,contadino)*

IT to SP WWCs: *(cont.,campesino.)* *(cont.,pueblerino)*



Spanish Senses:
campesino1_Human → *contadino1_Profession*
pueblerino1_Human → *contadino2_Human*
&[Connotation:*negative*]

Italian Senses:
&[Connotation:*negative* +Metaphor]

Here the occurrence of Connotation features for *pueblerino* and *contadino2* and the metaphoric status of *contadino2* allows the system to establish the correct correspondences.

3. Quantitative results

The results are given in groups according to Cases in 2.4 above. For each case we supply: the number of WWCs; the input and target words involved; the input and target senses involved; the number of resolved WWCs for the source language and, for them, the number of resolved SSCs.

Resolved WWCs are those WWCs with at least one resolved SSC correspondence between input/target senses involved. Resolved input senses are those input senses that are correctly assigned a SSC with a target sense.

WWC	95	
Input words	95	
Target words	95	
Input Senses	95	
Target Senses	95	
Resolved WWC	93	97.89%
Resolved SSC	93	97.89%

Table 1: Results for Catalan to Spanish TC

All correspondences but one are bi-directional, therefore SSCs are suggested as 'preferred'.

WWC	26	
Input words	26	
Target words	26	
Input Senses	26	
Target Senses	71	
Resolved WWC	26	100%
Resolved SSC	26	100%

Table 2: Results for Catalan to Spanish NTC1

WWC	212	
Input words	185	
Target words	212	
Input Senses	546	
Target Senses	563	
Resolved WWC	240	98.58%
Resolved SSC	433	79.30%

Table 3: Results for Catalan to Spanish NTC2

WWC	68	
Input words	68	
Target words	68	
Input Senses	68	
Target Senses	68	
Resolved WWC	65	95.58%
Resolved SSC	65	95.58%

Table 4: Results for Spanish to Italian TC

WWC	35	
Input words	35	
Target words	35	
Input Senses	35	

Target Senses	88	
Resolved WWC	34	99.02%
Resolved SSC	34	99.02%

Table 5: Results for Spanish to Italian NTC1

WWC	106	
Input words	106	
Target words	136	
Input Senses	272	
Target Senses	302	
Resolved WWC	99	93.39%
Resolved SSC	192	70.50%

Table 6: Results for Spanish to Italian NTC2

4. Conclusions

The main point of this exercise was to assess whether the lexicons and the model used to encode them were useful to identify the sense on which the correspondence relative to fully translation equivalents is made. In this sense, the results of the exercise are highly satisfactory.

Besides, the examples clearly show that sense-to-sense correspondences cannot be based on ontological terms only. Not only languages differ as far as lexicalisation of concepts is concerned, but also, and probably more critically if we want to provide automatic means for linking monolingual lexicons, we have to be able to cope with the fact that the criteria used for encoding using ontological labels might differ from one lexicon to another. This is the point where SIMPLE model proves crucial. SIMPLE Template Type system is not a mere collection of ontological labels. SIMPLE Templates are a generalisation on clusters of atomic elements of information. The procedure we have described analyses the clusters of information supplied by the input sense and looks for the target senses that better suit each of the input one.

Besides the relevance of the figures above, the results of the experiment lead us to suggest a classification of the relations that hold for the different correspondences in the following terms:

[+/- EQUIVALENT], for one-to-one correspondences *vs*. one-to-many correspondences.

One-to-many correspondences may derive from:

(i) **the semantic encoding** reflecting the differences between monolingual sense division and bilingual meaning discrimination –the target language contains two more fine grained descriptions for one under-specified input description (see WWC(*agua*,*aigua*) example below).

(ii) **the morphological encoding**: the way entries are split at the morphological layer might derive into a one-to-many correspondence -this is the case of the Spanish to Catalan *doctor vs. doctor/doctora* ('doctor-masculine', 'doctor-feminine') correspondence where the Spanish entry is under-specified for sex but the Catalan entries distinguish between masculine and feminine inheriting the distinction already made at the morphological level.

[+/- PREFERRED], whenever the input-target correspondence is bi-directional we also suggest a ponderation which reflects the preferred WWC. We expect the model to also allow for adding specific-to-general, slang-to-standard information to the correspondences.

4.1. Equivalent or Partial correspondences

Partial correspondences occur whenever in a given language a word is split up into several senses while in the other language the senses are considered as an indivisible meaning. The results of the experiment distinguish between three different situations:

A. Union Case: An 'under-specified' entry in one language subsumes two 'super-specified' entries in another language. This happens whenever a given feature is taken as 'sense discriminating' in only one language and the correspondence can be expressed in terms of set union:

SP*:agua_Substance* & [Telic:[+Edible] &
Constitutive:[+Liquid]]

⇓

CAT:*aigua*:*_Substance* & [Constitutive:[+Liquid]]

∪

CAT:*aigua:_Drink* & [Telic:[+Edible]]

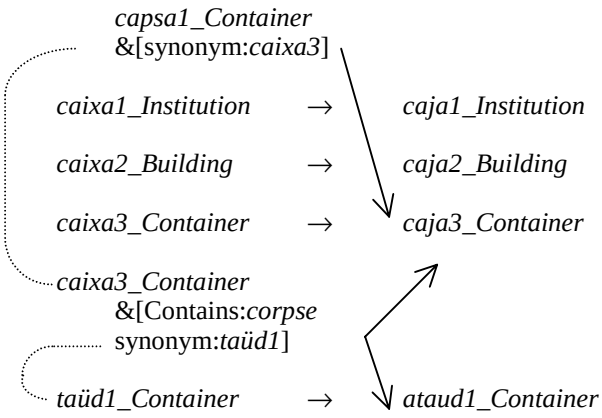
B. Constrained Case: An 'under-specified' entry in one language is related to two entries in the other language but the relation can not be explained in terms of 'set union'. This happens whenever the occurrence of certain information derives into sense discrimination. This is the case of the *doctor/doctora* example mentioned above:

SP: *doctor_Profession*
& [Constitutive: Sex_Underspecified
…]

CAT:*doctor_Profession*
& [Constitutive:Sex_Male
…]

CAT:*doctora_Profession*
& [Constitutive:Sex_Female
…]

Another interesting example of the Constrain Case is the case of the Spanish word *caja* (*box)* and the Catalan word *caixa*, where the Spanish lexicon includes one under-specified *Container* reading whereas the Catalan lexicon includes two *Container* readings one for 'box' and another for 'coffin' where the only difference between *caixa1* and *caixa2* is that the last one bears the additional restrictive information that it is a 'container for corpses'. We can see correspondences involving *caixa* and *caja* in the following diagram:



C. Extended senses: We have found some examples of 'extended' or figurative readings deriving from productive mechanisms which are only encoded in one the languages. As SIMPLE lexicons also include information concerning Polysemy relations in reference to such productive mechanisms, this information, together with those derived for closure, can be used to predict 'figurative' readings.

After having commented on the successful cases, the following step seems to be to ask for the other input senses involved that were not solved. As the figures reflect, we resolve 552 Catalan Input Senses out of 665 (83%) and 289 Spanish senses out of 375 (77.06%). An important issue to notice here is that the system we are defining does not consist on building up a bilingual lexicon but rather adding a modeled bilingual layer to a pair of monolingual lexicons. This means we depend on already existent resources and, therefore, on the differences regarding the coverage and criteria of such resources. This is an important remark because if we examine the unresolved input senses we find out these unresolved derive from:

(i) Word coverage in one of the monolingual lexicons: the target lexicon does not contain the target word and therefore the sense is not available;

(ii) Sense coverage: the target lexicon does not contain the required sense for instance because it belongs to a very specific domain (i.e. the Italian 'amniotic liquid' for *acqua* has not been included in the Spanish one), or because it is a 'figurative or metaphoric' reading, or it is a slang sense (the Spanish *caballo* for 'drug' or *pájaro* for 'guy' which had no counterpart in the target Italian lexicon);

(iii) Input senses which do not have a lexicalised counterpart in the target language. These are what traditionally bilingual dictionaries account with explanatory glosses. This is the case for *castell* or *torre* in Catalan ('human tower in popular festivities'), *bisbe* in Catalan ('short, fat sausage') or *fiesta* in Spanish ('bull festivity');

(iv) Finally, mismatching due to differences in encoding criteria which makes impossible any correspondence (this occurs for 5 WWC out of 333).

5. Future Perspectives of the work

The work reported gives some hints about what we have considered to be the basis for the construction of a multilingual lexical resource. To model the relations that hold between different lexical units in different languages by ensuring that they are [+preferred] and [+equivalent] sense-to-sense correspondences will allow to draw paths of fully translational equivalents for more than two languages. Thus, when having identified WWC's for different languages we can derive the SSC's for more than two languages which ensure the sense for true translational equivalence for all the languages involved. Such a modelisation must foresee the properties of the relations that link two lexical units in two different languages (for instance full equivalent will be a transitive relation).

(18) *contadino1_Profession*⇔*campesino_Human*⇔ *pagès_Profession*

We expect that partial correspondences will also be used to express complex relations based on restrictions. Using other information encoded, such as metaphoric use or synonymy, will be useful for identifying candidates that can be offered to the lexicographer for him to take the last decision in bilingual linking if possible.

However, we have to extend the scope of the experiment so as to include all translational equivalences given by bilingual dictionaries and synonym information

in SIMPLE lexicons in order to extend correspondence assignments within a set of related senses. We predict that this exercise will provide better results and give new clues for further defining and modeling the set of possible correspondences for a multilingual resource.

6. References

Fontenelle, T., 1997 *Turning a bilingual dictionary into a lexical-semantic database*. Tübingen: Max Niemeyer, 1997. (Lexicographica. Series maior; 79).

Heid, U., 1990. Monolingual, bilingual, <<interlingual>> description. Some remarks on a new method for the production of bilingual dictionaries. *Euralex '90: proceedings*: 4th. International EURALEX Congress: Barcelona. Biblograf, 1992. (Vox). p. 167-184.

Klavans, J. & Tzoukermann, E., 1995. Combining corpus and machine-readable dictionary data for building bilingual lexicons. Machine Translation, 10: 185-218.

Kromann, H.P., 1989. Principles of Bilingual Lexicography. In Dictionaries, Wörterbucher, Dictionnaires. Walter de Gruyter, New York.

Lenci, A., F. Busa, N. Ruimy, E. Gola, M.Monachini, N. Calzolari, A. Zampolli, E. Guimier, G. Recourcé, L. Humphreys, U. Von Rekovsky, A. Ogonowski, C. McCauley, W. Peters, I. Peters, R.Gaizauskas, M. Villegas, 2000. *SIMPLE Linguistic Specifications (Project Deliverable D2.2)*

Picchi, E.; Peters, C.; Calzolari, N., 1988. "Implementing a bilingual lexical database system". In *BudaLEX '88: proceedings*: papers from the 3rd International EURALEX Congress: Budapest, 4-9 September 1988. T. Magay and J. Zigány eds. Budapest: Akadémiai Kiadó, 1990, p. 317-329.