Methodological issues in the building of the Basque WordNet: quantitative and qualitative analysis

Eneko Agirre, Olatz Ansa, Xabier Arregi, Jose Mari Arriola,
Arantza Diaz de Ilarraza, Eli Pociello and Larraitz Uria
IXA NLP Group
University of the Basque Country
649 pk. 20.080 - Donostia. Spain
eneko@si.ehu.es

Abstract

This paper describes the methodology we have adopted to ensure the quality of the Basque WordNet in terms of coverage, correctness, completeness and adequacy. The Basque WordNet follows the EuroWordNet framework and, basically, it is produced using a semi-automatic method that links Basque words to the English WordNet. We have found that in order to ensure proper linguistic quality and avoid excessive English bias, a double manual pass on the automatically produced Basque synsets is desirable: a first concept-to-concept pass to ensure correctness of the Basque words linked to the synsets, and a word-to-word pass to ensure the completeness of the word senses linked to the words. By this method, we expect to combine quick progress (as allowed by a development based on the English WordNet) with quality (as provided by a development based on a native dictionary). We have completed the concept-to-concept review of the automatically produced links for the nominal concepts, and are currently performing the word-to-word review.

Introduction

This paper presents work on the Basque WordNet (BasqWN). Our team had an increasing need to construct an extensive and complete Lexical Knowledge Base for Basque (LKBB). To this end, our point of departure is the English Wordnet developed at Princeton University (Fellbaum, 1998), which is consolidating as a 'de facto' standard for the lexical-semantic representation for English. Considering this English Wordnet as a reference, new WordNets in some other languages have been built, especially in the framework of the EuroWordNet project[1]. EuroWordNet (EuroWN) basically adds multilingual links across WordNets.

We present the current state of the Basque WordNet, but this paper rather focuses on the methodology that we chose to build the Basque WordNet. The emphasis of this work is on two points:

- The need to check the data produced with external lexical resources for Basque.
- The need to perform a thorough quality check of the data produced, with special incidence on the necessity of concept-to-concept and word-to-word reviews.

One of the motivations of this paper is the lack of literature on methods to guarantee and assess the correctness, completeness and adequacy of the information produced. EuroWordNet, for instance, published an evaluation of the produced WordNets that consisted basically on a comparison of a number of ratios (variants per concept, senses per entry, overlaps across wordnets) and an evaluation of errors in cross-language equivalence relations (Vossen et al., 1998a; Vossen et al., 1998b).

The paper is organized as follows. We first present the overall principles for the design and methodology, together with the current state of the Basque WordNet. Section 2 presents the automatic generation and concept-to-concept review processes that lead to the preliminary BasqWN 0.1 release. Sections 3 and 4 review the quantitative and qualitative analysis of BasqWN 0.1. Section 5

[1] http://www.hum.uva.nl/~ewn

summarizes the conclusions drawn from the analysis. Section 6 presents the method for the word-to-word review. Finally, section 7 presents some conclusions and future work.

1 Design and methodology

The design of the Basque WordNet (BasqWN) follows that of EuroWN. In EuroWN each language WordNet has its own relations and set of synsets or concepts[2]. The link across the different language WordNets is made using the Inter-Lingual Index (ILI), which is a list of concepts. Most of the concepts in the ILI come form the English WordNet version 1.5. Words are linked to concepts. If we take the concepts linked to one word, each link (i.e. each linked concept) expresses the word senses of the source word. If we take the words linked to one concept, each link (i.e. each linked word) corresponds to a variant of the concept. The variants for a given concept are synonyms of each other. An example of the terminology used in the paper is the following:

- The words *maverick* and *rebel* are variants for the nominal concept/synset 06228559 glossed as *someone who exhibits great independence in thought and action*
- Concept 06228559 above corresponds to the first sense for *maverick*, and the third sense for *rebel*.

The method to be used in order to build the Basque WordNet was a matter of debate in our research group. We could either take a sense inventory and a concept hierarchy independent from WordNet (derived from a dictionary or our own lexicographic work) and link it to the ILI, or take the English WordNet 1.5 concepts and hierarchy and link Basque words to them.[3]

The advantage of the former approach is that we control the sense inventory and hierarchies, building them according to our criteria, at the cost of overtaking expensive lexicographic works and having to devise a way to link the Basque WordNet to the ILI. The advantage of the second approach is that we have the link to the ILI for free, and that most of the work is just reduced to link Basque words to the ILI. We ultimately aimed to have the best of both approaches, but we decided to take the English WordNet 1.5 as the starting point. In any case, monolingual Basque dictionaries will give us a reference for the quality of the sense distinctions made, and we also plan to use the hierarchies and lexico-semantic relations extracted from dictionaries to enrich or even transform the Basque WordNet. In fact, we are studying a more complex representation for lexico-semantic relations, which could depart the Basque WordNet from the actual EuroWordNet design.

From another point of view, the method to build the Basque WordNet was influenced by the following general criteria: we want the Basque WordNet to have wide coverage and we also want it to be of high quality. Coverage involves concepts, entries, parts of speech, word senses and synonyms. Quality can be divided in correctness, completeness and adequacy:

1. Correctness of variants of a synset and word senses of a word.
2. Completeness of variants of a synset and word senses of a word.
3. Adequacy of the specificity level for variants in the synset (also wrt. to English equivalents), and for word senses.

We had an additional requirement on the Basque WordNet which conflicts with the above list in the short term: we needed fast development with limited resources. This led us to first put the stress on coverage, and leave quality enforcement for later. For the same reason, we adopted a semi-automatic procedure to build the nominal part (cf. section 2).

The above considerations lead us to devise a three-stage process for the development of Basque WordNet:

- First stage: development of the core Basque WordNet.
  For each part of speech:
  1. Link all Base Concepts manually, checking quality.
  2. Automatically generate Basque equivalents for the English concepts.

---

[2] In this paper concept and synset refer to the same thing. We are aware that in lexical resources other than WordNet, concepts and synsets can be different things.

[3] In EuroWordNet (Vossen et al. 1998a) the former approach is referred to as the *merge* model, and the latter as the *expand* model.

3. Concept-to-concept review of the equivalents generated in previous step. The goal is the fast development of the core WordNet for this part of speech. The stress is on coverage and speed. Linguists focus on correctness of variants in a synset, but also cover completeness of variants in the synset.
4. Word-to-word review of word senses. The goal is twofold: ensure quality across word senses and try to cover main senses for most frequent/relevant words. The stress is on quality. Linguists focus on the correctness and completeness of word senses for a word.
5. Check adequacy of specifity level.

- Extend coverage to all the vocabulary.
- Second stage: map the Basque WordNet to a Basque monolingual dictionary. Design of an enriched representation and extraction of varied lexico-semantic representations from the dictionary.

We want to stress the need of both the concept-to-concept and word-to-word reviews. Both work on the same data (word-concept links) but from different complementary angles.

1.1 Current state of the Basque WordNet

At present, the Basque WordNet is in the following development state (see section 3 for figures on number of synsets and entries):

- Verbs: Base Concepts and a core subset (similar to S1 in EuroWN) have been manually done (step 1 of first stage).
- Nouns: Base Concepts have been manually done (step 1). The automatic method was executed, and the concept-to-concept review completed (steps 2 and 3) At this stage, we released the preliminary version, BasqWN 0.1 which is of similar size and characteristics to the final EuroWN releases (Vossen et al., 2001). The word-to-word review (step 4) is being performed at present, as well as the check on the specifity level.

Regarding the human effort involved, the concept-to-concept review (step 3) took 1,640 hours at approximately 16 concepts per hour. The word-to-word review has already taken 502 hours, at a rate of approximately 2 words per hour, which has involved reviewing around 16 concepts per hour.

The Basque WordNet is implemented as a relational database with a web interface that allows browsing and editing (Benitez et al. 1998). Linguists have access to a set of integrated online lexical resources comprising the Spanish and English WordNets, a monolingual Basque dictionary (Sarasola, 1996), a synonym dictionary (UZEI, 1999), two bilingual English-Basque dictionaries (Morris 1998; Aulestia & White, 1990), a bilingual Spanish-Basque dictionary (Elhuyar, 1998), and a terminology database (UZEI, 1987). The linguists also use the Oxford Spanish-English bilingual dictionary off-line (OUP, 1994). The project currently involves three persons in the supervision committee, one person for linguistic coordination, two persons for linguistic work and one person for computer support.

2 Automatic generation and concept-to-concept review

In order to help the linguists in their task, we automatically generated Basque noun concepts from machine-readable versions of Basque-English bilingual dictionaries (Morris, 1998; Aulestia & White, 1990). All translation pairs in the dictionaries were extracted in the form of English term-Basque term (*term* here refers to a single word or multiword phrase). These pairs were combined with WordNet synsets and the resulting combinations were analyzed following the class methods (Atserias et al. 1997). The algorithm produces triples like Basque word – synset – confidence ratio. The confidence ratio is assign depending on the results of the hand evaluation, which is organized around different submethods. The pairs produced by class methods with a confidence rate lower than 62% were discarded. Next section presents the figures for this step.

The results of the previous process were validated by hand. The linguists reviewed the synsets that had a Basque equivalent one by one, checking whether the Basque words were correctly assigned and adding new words to the synonym set if needed. This process can be sped up if the linguist treats related synsets, i.e. when the next synset to be treated is related to the previous one. In order to facilitate the manual work, we used an interface linked to a database (Benitez et al., 1998) that allowed for simultaneous updates and accesses. In order to offer the synsets following hyperonym

chains, we added an extra button that lead to the next synset in the hyperonym chain that was yet undone. In addition, there was the possibility to mark a synset as "dubious". This was used when further lexicographic analysis was needed. In this case, the linguists would meet with the linguistic coordinator. Next section presents the figures for this step.

We have to note that following the above procedure no new concepts are added to the ILI, i.e. we do not add concepts that are not already available in the English WordNet.

3 Quantitative state of BasqWN 0.1

Table 1 reviews the amount of synsets, entries, etc. of the Basque WordNet compared to WordNet 1.5 and the EuroWordNet final release (Vossen et al. 2001). The first two rows show the number of Base Concepts, which were manually done. For nouns in the Basque WordNet 0.1, the *Nouns(auto)* row shows the figures as produced by the raw automatic algorithm , and the *Nouns(man)* row the figures after the manual concept-to-concept review. The number of entries was manually reduced down to 50%, and the number of senses down to 15%. This high number of spurious entries and senses is caused primarily by a high number of orthographic and dialectal variants that were introduced by the older bilingual dictionary, which does not follow the standard rules adopted in the last years.

There is an extra set of 3,000 synsets that are currently under review. This set comprises mainly multiword terms with dubious lexicalization in Basque, i.e., the Basque equivalent is a definition rather than a term:

maverick (unbranded range animal esp. a stray calf): *markatu_gabeko_txahal*
breakthrough (making an important discovery): *garrantzi_handiko_aurkikuntza*

		Synsets	***No. of senses***	***Sens./ syns.***	***Entries***	***Sens./ entry***
Basque WordNet	**Nominal BC**	228	-	-	-	-
	Verbal BC	792	-	-	-	-
Basque WordNet 0.1	**Nouns (auto)**	27641	291011	10.52	46164	6.3
	Nouns (man)	23486	41107	1.75	22166	1.8
	Verbs (man)	3240	9294	2.86	3155	2.95
WN1.5	**Nouns**	60557	107484	1.77	87642	1.23
	Verbs	11363	25768	2.27	14727	1.75
Dutch Wordnet	**Nouns**	34455	54428	1.58	45972	1.18
	Verbs	9040	14151	1.57	8826	1.60
Spanish Wordnet	**Nouns**	18577	41292	2.22	23216	1.78
	Verbs	2602	6795	2.61	2278	2.98
Italian Wordnet	**Nouns**	30169	34552	1.15	24903	1.39
	Verbs	8796	12473	1.42	6607	1.89

Table 1: figures for the Basque WN release 0.1 compared to WN 1.5 and the final EWN release.

The senses per entry figures are higher than those from WN 1.5 and most of the WordNets, but similar to the Spanish WordNet. The fact that the nouns and verbs included are in general more polysemous can explain this fact. We also performed an analysis of the distribution for the variants in each synset and the number of word senses per entry. From this analysis we found that many words had a polysemy degree higher than 15, a figure that is not observed in the English WordNet. This fact is analyzed in more detail in section 4.2.

All in all, the amount of synsets and entries for the Basque WordNet 0.1 is comparable to those for the WordNets produced in EuroWordNet, but lower than the WN 1.5 release. The coverage of concepts in WN 1.5 is 38% and the coverage of entries in a Basque monolingual dictionary is 100%.

4 Qualitative analysis of BasqWN 0.1

Somehow, we were not satisfied by the quantitative analysis and the results of the concept-to-concept review. On the one hand, the quantitative analysis only shows the state of the coverage of concepts and entries, as long as they are compared to reference figures from WordNet (concepts) and Basque reference dictionaries (entries). It is rather difficult to assess the coverage of the number of word

senses and synonyms[4], as these can only be compared to WordNet, but there are no reference figures for the Basque WordNet itself. We think that the coverage of word senses and synonyms can be more reliably estimated measuring by hand the completeness of the word senses of a sample of words and the variants for a sample of concepts.

On the other hand, the concept-to-concept review only enforces the correctness and completeness of the variants in the synset. As the stress of the first stage was on quickly producing a first version, correctness was more important than completeness, and we were not completely satisfied with the completeness of the variants.

As already listed in section 1, these are the correctness, completeness and adequacy requirements that were not covered by the quantitative analysis:

a. Correctness of word senses of a word. This is indirectly enforced by the concept-to-concept review, but the linguist tends to focus on the list of variants itself. Additional cross-validation could be needed.
b. Completeness of word senses of a word. This can only be enforced by comparison to lexical resources listing sense inventories. This could be used to provide an estimate of the coverage for word senses.
c. Completeness of variants of a concept. Further review could be necessary, which could provide an estimate of the coverage for synonyms.
d. Adequacy of the specifity level for variants in a concept, i.e. all variants of a concepts are of the same specificity level.
e. Adequacy of the specifity level for word senses, i.e. granularity of word senses.

In order to assess points **a**, **b** and **e**, we performed a manual comparison and mapping of the word senses given by BasqWN 0.1 with those of a monolingual dictionary and a bilingual dictionary. This assessment is presented in the next subsection.

We have also manually checked the correctness and completeness (**c**) of the variants for a concept, using a synonym dictionary for this purpose. The results were highly satisfactory, but we decided to explicitly include the use of the synonym dictionary in all subsequent reviews and updates of the Basque WordNet (see next section).

Subsection 4.2 presents some preliminary assessment of the adequacy of the specificity level for variants in a concept (**d**).

4.1 Manual mapping of word senses from BasqWN and Basque dictionaries.

The sense partition of Basque monolingual dictionary reflects a suitable native sense partition, and needs not to be of the same granularity as of WordNet. In principle, both sense partitions could even be incompatible, in the sense that it could involve many-to-many mappings.

We chose to use the *Euskal Hiztegia* (EH) dictionary (Sarasola, 1996), as it is a general purpose monolingual dictionary, and it covers standard Basque. It contains 33,111 entries and 41,699 senses. Besides, it is the dictionary used for the Basque task in the Senseval 2001 competition. One drawback of this dictionary is that it mainly focuses on literature tradition, and it lacks many entries and word senses which are more recent. For this reason, we decided to include also a bilingual Basque-English dictionary (Morris, 1998). Moreover, if the linguist thought that some other word sense was missing he/she was allowed to include it.

The linguist was also required to leave out some senses from EH, particularly those tagged as nuances of other senses corresponding to rare usages.

[4] Coverage of synonyms is linked to the number of variants in each synset. As explained in section 1, we use the terms *variant* and *synonym* to name the words that lexicalize a synset.

Word	***Number of senses***	***Senses Added From Bilingual Dictionary***	***Senses Added From Monolingual Dictionary***	***Senses Added By Linguist***
HERRI: country, nation, town	9	0	3	0
TENTSIO: tension	9	0	0	1
BIDE: way	22	3	4	0
GAI: material, matter	16	0	0	1
EGUN: day	9	1	1	0
ELIZA: church	4	0	1	0
MASA: mass	4	1	0	0
UR: water	5	2	0	0
KANAL: canal	8	1	0	0
KOROA: crown	9	0	1	0
IBILBIDE: course, way	8	3	0	0
KANTU: song	3	0	1	0
KAPITAIN: captain	5	0	1	0
LANTEGI: factory	5	1	0	0
ENPLEGU: job	1	1	1	1
Total	117	13	13	3

Table 2: word senses missing from BasqWN 0.1

Table 2 shows the result of the comparison of the sense inventories. The first column lists the Basque words alongside some of their possible translations. The second column shows the number of word senses per word in BasqWN. The third, fourth and fifth columns list the number of word senses added according to the bilingual dictionary, monolingual dictionary and the linguist's introspection.

All in all, both bilingual and monolingual dictionaries contribute equally to the new senses. An average of 1.9 new senses are added for each word, which makes an average of 0.24 new senses for each existing sense. This makes an idea of the completeness of the word senses for words. All word senses were found to be correct. These figures can be interpolated to estimate that the coverage of word senses for the entries currently in BasqWN is around 80%.

Regarding the mapping between the word senses of BasqWN and the monolingual dictionary, most of the times it was one-to-one or many-to-one. The granularity of the word senses in BasqWN is much finer. We have even found a sense in the Basque monolingual dictionary that accounts for 9 BasqWN word senses. We have not found many-to-many mappings. We have to note that some of the word senses in BasqWN were not present in the Basque dictionary, as illustrated in the following word senses in BasqWN:

CROWN (*koroa*) ➔ an English coin worth 5 shillings
CAP (*koroa*) ➔ an artificial crown for a tooth
DATE (*egun*) ➔ the particular year that an event occurred

This analysis has also uncovered some other difficulties such as the incompatibility concerning parts of speech. For instance, one of the senses of the word *herri* can only be used in compounding and the English equivalent would be that of the adjective *public.* In order to map both senses, we need to find a nominal synset equivalent to *public* (or introduce a new nominal synset with this meaning) and introduce a cross-PoS equivalent relation between the nominal and adjectival concepts. These kind of phenomena cannot be detected in concept-to-concept review.

4.2 Adequacy of the specificity level of variants in synsets

As already mentioned in the quantitative analysis, we found out that some words had an unusually high number of word senses. Quick hand inspection showed that for some concepts the variants were of heterogeneous specifity. In fact, we suspected that some words were placed in too many concepts. An example follows:

> The concept r*eligious* glossed as *a member of a religious order* has the following Basque variants: *erlijioso, serora, lekaide* (respectively translated as *religious, nun, monk*)

Two of the Basque variants are "correct" in some way, but they are wrongly placed in the hierarchy.

In fact, a program that searches for words that have two word senses, one hyperonym of the other found out that there are 4500 such pairs out of 41107 word senses. This is a very high figure compared to the English WordNet, and indicates that we need to check those word senses.

In other cases, the problem is more difficult to detect. For instance:

> The concept *superior* glossed as *the head of a religious community* has the following Basque variants: *nagusi, buru* (respectively translated as *boss, head*)

In this case, there is not direct equivalent for *superior* in Basque, and *buru* is correctly used to lexicalize this concept. From a specificity point of view, we could say that *buru* is the Basque hyperonym for *superior*, but from another point of view, it is a perfectly plausible translation equivalent in Basque.

5 Conclusions of the quantitative and qualitative analysis of BasqWN 0.1

A summary of the quality assessment for the nominal part of the Basque WordNet is the following:

- Coverage of concepts: 38% of the concepts in WordNet 1.5.
- Coverage of entries: 22,166 entries, which makes 25% of WordNet 1.5, but accounts for all *Euskal Hiztegia* entries.
- Coverage of senses: estimated as 80% of the senses for the entries already in BasqWN 0.1.
- Coverage of synonyms: estimated as complete for the present concepts.
- Correctness of variants of a synset: estimated to be correct for all variants.
- Completeness of variants of a synset: estimated as all variants being present.
- Adequacy of the specificity level of the variants of a synset: we have some evidence that in some instances one of the variants are at the wrong level of the hierarchy, usually too high.
- Correctness of word senses of a word: estimated as all senses being correct.
- Completeness of word senses of a word: estimated as 20% of the word senses being missing.
- Adequacy of the specificity level of word senses of a word (granularity): the specifity level of BasqWN is much finer than that of the reference dictionary. This does not need to be a problem.

A number of other issues have also been observed:

- We have around 3000 nominal concepts with dubious lexicalizations.
- Some words have a very high polysemy degree.
- We found incompatibility of part of speech for a number of word senses.

6 Word-to-word review

Most of the shortcomings detected in the previous section can be overcome following an additional review of the current BasqWN 0.1. In this review we want to ensure that the coverage of word senses is more complete, trying to include the estimated 20% of word senses that are missing. In this case, the review is to be done studying each word in turn and taking attention to the following issues:

- Coverage of senses: add main word senses of basic words.
- Correctness of word senses of a word: delete inadequate word sense when necessary.
- Completeness of word senses of a word: add main word senses.
- Adequacy of the specificity level of word senses of a word (granularity): check that granularity of word senses is balanced.

The need to build a core WordNet lead us to define a subset of the nominal entries to be covered: on the one hand, the top 400 words from a frequency analysis were treated, on the other hand, the entries in a basic bilingual Basque-Spanish dictionary (Elhuyar, 1998) which tries to define the core vocabulary of Basque (13,000 nouns). The word senses are provided by the monolingual dictionary

(EH) and the bilingual dictionary. The bilingual dictionary includes modern words and word senses which are not in EH.

7 Summary and further work

This paper has presented a methodology that tries to integrate the best of development methods based on the translation of the English WordNet and development methods based on a native dictionary. We first have developed a quick core WordNet comparable to the final EuroWordNet released using semi-automatic methods that includes a concept-to-concept manual review, and later performed an additional word-to-word review based on native lexical resources that guarantees the quality of the WordNet produced.

We are currently extending the coverage of the noun entries and word senses to those in a basic vocabulary of Basque. In the future we plan to apply the methodology to verbs and adjectives and to extend the coverage to a comprehensive set of noun entries. In addition, we would like to check the coverage for entries and word senses of nouns in a Cross-lingual Information Retrieval application. In parallel, we are planning to map the word senses in a Basque monolingual dictionary to the Basque WordNet, which would allow to import lexico-semantic relations from the Basque dictionary (Agirre & Lersundi, 2001) to EuroWordNet.

Acknowledgments

This research was partially funded by the European Commision under the Feder program (project 2FD1997-1503), the Spanish Ministry of Science and Technology (project Hermes TIC2000-0335-C03-03), the University of the Basque Country (UPV 141.226-G19/99) and the Provincial Government of Gipuzkoa (project Berbasare OF 206/00).We would like to thank the research team of the Technical University of Catalonia for sharing their algorithms and tools with us.

References

Agirre, E. and Lersundi, M. Extracción de relaciones léxico-semánticas a partir de palabras derivadas usando patrones de definición. In Proceedings of SEPLN 2001. Jaén (Spain). September, 2001.

Atserias, J.; Climent, S.; Farreras, J. Rigau, G. & Rodríguez, H. 1997. Combining Multiple Methods for the Automatic Construction of Multilingual WordNets. In proceedings of Conference on Recent Advances on NLP. (RANLP'97). Tzigov Chark, Bulgaria 1997.

Benítez, L. ;Escudero, G. ; Farreras, J. & Rigau, G. 1998. WWI: A Multilingual WordNet Interface using the Web. Technical Report LSI-98-6-T. LSI Department, Universitat Politécnica de Catalunya.

Fellbaum, C.1998. WordNet: An electronic Lexical Database. The MIT Press,Cambridge, Massachusetts. London, England.

UZEI. 1987. Euskalterm. http://www.uzei.com/en/euskalter.htm (20 Sep. 2001).

Vossen, P. ; L. Bloksma; S. Climent; M. Anonia Marti; G. Oreggioni; G. Escudero; G. Rigau; H. Rodriguez; A. Roventini; F. Bertagna; A. Alonge; C. Peters; W. Peters. 1998. The Reestructured Core wordnets in EuroWordnet: Subset1. EuroWordNet(LE-4003) Deliverable D014/D015, University of Amsterdam.

Vossen, P.; L. Bloksma; S. Climent; M.A. Marti; M. Taule; J. Gonzalo; I. Chugur; M. F. Verdejo; G. Escudero; G. Rigau; H. Rodriguez; A. Alonge; F. Bertagna; R. Marinelli; A. Roventini; L. Tarasi. 1998. EuroWordNet Subset2 for Dutch, Spanish and Italian, EuroWordNet (LE-4003) Deliverable D027/D028, University of Amsterdam.

Vossen, P.; L. Bloksma; S. Climent; M.A. Marti; M. Taule; J. Gonzalo; I. Chugur; M. F. Verdejo; G. Escudero; G. Rigau; H. Rodriguez; A. Alonge; F. Bertagna; R. Marinelli; A. Roventini; L. Tarasi., W. Peters. 2001. Final Wordnets for Dutch, Spanish, Italian and English, EuroWordNet (LE2-4003) Deliverable D032/D033, University of Amsterdam.

Dictionaries

Aulestia, G. & White , L. 1990. English-Basque Dictionary, University of Nevada Press, Reno.

Elhuyar. 1998. Hiztegi txikia.

Morris, M. 1998. Morris Hiztegia.

OUP, 1994. The Oxford Spanish Dictionary, Oxford University Press, 1994

Sarasola, I. 1996. Euskal Hiztegia.

UZEI. 1999. Sinonimoen Hiztegia.