Pronominal Anaphora in Basque: Annotation issues for later computational treatment

Aduriz I.*, Ceberio K. and Díaz de Ilarraza A.

IXA Group
(http://ixa.si.ehu.es)
Department of Computer Languages and Systems
University of the Basque Country
Manuel Lardizabal 1, 20018 Donostia
mailto:jibcebec@ehu.es

*Department of General Linguistics
Universitat de Barcelona
Gran Vía de les Corts Catalanes 585, 08007 Barcelona
itziar@fil.ub.es

Abstract

This paper describes the process followed in the annotation of pronominal anaphora in the Eus3LB Corpus[1] of Basque. Our aim is to use this annotation as the basis for later computational treatment of Basque language. We present the criteria defined for the tagging, the problems we found with the ambiguity and some relevant linguistic conclusions about the features of the antecedents needed to link them correctly to their anaphoric elements.

1. Introduction

Anaphora resolution is crucial in real-world natural language processing applications e.g. machine translation. Although it has been a wide-open research field in the area since 1970s, this is the first work dealing with the subject for Basque especially in the task of tagging corpora.

Anaphora resolution, like other types of language's automatic treatment, needs corpus annotation. Mitkov (2002) highlights the importance of an annotated corpus for research purposes: *"The annotation of corpora is an indispensable, albeit time-consuming, preliminary to anaphora resolution (and to most NLP tasks or applications), since the data they provide are critical to the development, optimization and evaluation of new approaches"*.

Taking this statement into account, we began annotating Eus3LB Corpus with anaphoric information. This Corpus is part of the general 3LB project (Palomar et al., 2004), whose main objective is to build three syntactically, semantically and pragmatically annotated corpora for Spanish, Catalan and Basque.

This paper describes the way followed in the anaphoric tagging of Eus3LB. First, some general features of Basque language are explained. In section 3, we settle on the main subject of our study, the pronominal anaphora, and the annotation scheme. The fact that there is not a specific theoretical work for Basque in this area, made the previous study larger than we expected. Section 4 explains the annotation process and finally, in section 5, we present some conclusion and future work.

2. Main Features of Basque

Basque is not an Indo-European language and differs considerably in grammar from the languages spoken in other regions around. It is an agglutinative language, in which grammatical relations between components within a clause are represented by suffixes. This is a distinguishing feature since the morphological information that words contain is richer than in surrounding languages. Given that Basque is a head final language at the syntactic level, the morphological information of the phrase (number, case, etc.), which is considered to be the head, is in the attached suffix. That is why morphosyntactic analysis is essential.

3. Pronominal Anaphora in Basque

According to Hirst (1981): "*anaphora, in discourse, is a device for making an abbreviated reference (containing fewer bits of disambiguating information, rather than being lexically or phonetically shorter) to some entity (or entities)*".

This reference can appear before the anaphoric element (*[Lisa]$_i$ could see the stars in the sky. [She]$_i$ was very lucky*) or after it (*The elevator opened for [him]$_i$ on the 14th floor, and [Alec]$_i$ stepped out quickly*) in this case we call it cataphora (Mitkov, 2002).

Anaphora has been classified in several ways (Mitkov, 2002): depending on its grammatical category, the position of its referent in the text, etc. (Ferrández, 1998).

In this work we have specifically focused on the pronominal anaphora; concretely, the demonstrative determiners only when they behaved as pronouns. In Basque there are not different forms for third person pronouns and demonstrative determiners are used as third person pronominals (Laka, 2000). There are three degrees

[1] The Eus3LB corpus is a part of the corpus of the 3LB project (Palomar et al., 2004).

of demonstratives that are closely related to the distance: *hau* (this/he/she/it), *hori* (that/he/she/it), *hura* (that/he/she/it).

Besides, the gender is not a valid feature to detect the antecedent of a pronominal anaphora because there is no gender distinction in the Basque morphological system.

3.1 The annotation schema

3.1.1. Previous work

In the literature, we have found bibliographic references to corpora annotated both anaphorically and coreferentially. Some of the sources for this study were: The Lancaster Anaphoric Treebank (UCREL) (Garside et al., 1997), the MUC Coreference Task (MUC-7) (Hirschman, 1997), the corpus of the University of Wolverhampton (Mitkov, 2000), part of the Penn Treebank Corpus (Ge, 1998), DRAMA scheme (Passoneau and Litman, 1997) and the MATE/GNOME scheme (Poesio, 2004). All of the aforementioned corpora annotations have been carried out in and for the English language.

Moreover, we consulted resources for other languages, such as the TIGER Project (Kunz & Hansen-Schirra, 2003) for German. Similar work has been carried out at the University of Prague (Hajič & Urešová, 2004), where the corpus has been annotated at a pragmatic level, including the annotation of coreferential elements.

Finally, Navarro et al., (2003) carried out the study for Spanish. The Cast3LB[2] Corpus has been tagged pragmatically with the help of an annotation tool. This tool has marked the anaphoric and coreferential relationships (including ellipsis) as well as the corresponding referents.

3.1.2. The Eus3LB Corpus

Eus3LB was created as a reference corpus in the framework of 3LB project. In this section, we specify what we have already tagged in the Eus3LB Corpus and we explain the criteria defined for the annotation of the pronominal anaphora.

The 50.000 words corpus we worked with consists of journalistic texts parsed before.

The parsing process starts with the outcome of the morphosyntactic analyser MORFEUS (Aduriz et al., 1998), which was created following two-level morphology (Koskenniemi, 1983) and it deals with the parsing of all the lexical units of a text, both simple words and multiword units.

From the obtained results, grammatical categories and lemmas are disambiguated. The disambiguation process is carried out by means of linguistic rules written following the CG grammar formalism (Karlsson, 1995) and stochastic rules based on markovian models (Ezeiza et al., 1998) with the aim of improving the parsing tags in which the linguistic information obtained is not accurate enough.

Once morphosyntactic disambiguation has been performed, we should, ideally, be working on a morphosyntactically fully disambiguated text when assigning syntactic functions.

The aim of the syntactic disambiguation rules is to assign a single syntactic function to each word. In the next noun and verb chains stage the corpus will be marked by a chunker and therefore verb and noun chains make use of the syntactic functions provided by each word-form. The next figure sums the architecture of the system.

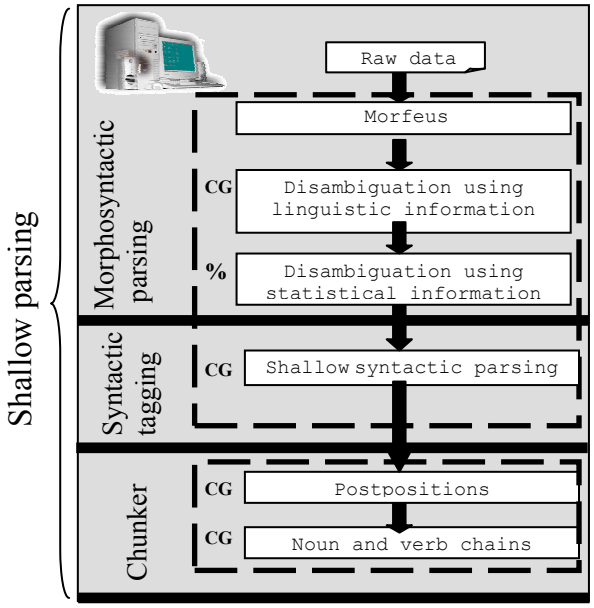

Fig. 1. Syntactic processing of Basque.

The tags we have taken into account for the anaphoric annotation are those referring to noun phrases. Regarding these phrases, we recognise simple and coordinated noun chains, for which these three function tags have been established:

- % NCH: this tag is attached to words with main syntactic function tags that constitute a phrase unit by themselves.
- % INIT_NCH: this tag is attached to the initial element of a phrase unit.
- % FIN_NCH: this tag is attached to the final element of a phrase unit.

The example in figure 2 shows this morphosyntactic analysis. Let us explain how the system works whith the following example:

(1) ***Ertzaintzako patruila bat*** *bertaratu zen eta* ***hark*** *itzali zuen sua.*
*(**A police car** came to the place and **(it)** put out the fire.)*

/<Ertzaintzako>/ "<INIT_CAP>" "Ertzaintza" Proper Noun @GENITIVE> %INIT_NCH	police
/<patruila>/ ""COMMON NOUN @NOCASE	car
/<bat>/ "bat" DEF DET … @SUBJ %FIN_NCH	a
/<bertaratu>/ "bertaratu" VERB… @-MAINVERB %INIT_VCH	to come
/<zen>/ "izan" AUX VERB… @+AUXVERB %FIN_VCH	to be (aux.)
/<,>/<PUNT_KOMA>/	
/<eta>/ "eta" CONJUNCTION @CJNCT	and
/<**hark**>/ "**hura**" DEM DET ERG @SUBJ %**NCH**	it
/<itzali>/ "itzali" VERB … @-MAINVERB %INIT_VCH)	to put out
/<zuen>/ "*edun" AUX VERB @+AUXVERB %FIN_VCH	to do (aux.)
/<sua>/ "su" COMMON NOUN … @SUBJ %NCH)	the fire
/<.>/<PUNT_PUNT>/	

Fig. 2. Analysis of chains. English translation on the right

[2] Cast3LB and Eus3LB are part of the 3LB project (Palomar et al., 2004).

The elements the annotators had to mark were the demonstrative determiners like *hark* (it ergative), when they behaved only as a pronoun. In other words, the demonstratives which had the %NCH tag. In the next section we explain what we tagged and how this tagging was carried out.

3.2. Annotation format

The human taggers were said to mark the mentioned demonstrative determiners as anaphoric elements, using the tag [ANAnum][3]. After that, they looked for the correct antecedents, and the anaphora's corresponding referents in the text were tagged with the [REFnum]. If referents consisted of more than one element, they were marked with the tag [REF-Bnum] at the beginning of the phrase and [REF-Enum] at the end of the phrase.

This is exemplified in Figure 3: *hark* (it) is the anaphora and *Ertzaintzako patruila bat* (a police car) its antecedent.

```
/<Ertzaintzako>/ "<INIT_CAP>"                police
  "Ertzaintza" Proper Noun @GENITIVE>
  %INIT_NCH [REF-B3]
/<patruila>/                                 car
""COMMON NOUN @NOCASE
/<bat>/                                      a
"bat" DEF DET … @SUBJ %FIN_NCH [REF-E3]
/<bertaratu>/                                to come
"bertaratu" VERB… @-MAINVERB %INIT_VCH
/<zen>/                                      to be (aux.)
"izan" AUX VERB… @+AUXVERB %FIN_VCH
/<,>/<PUNT_KOMA>/
/<eta>/                                      and
"eta" CONJUNCTION @CJNCT
/<hark>/                                     it
"hura" DEM DET ERG @SUBJ %NCH [ANA3]
/<itzali>/                                   to put out
"itzali" VERB … @-MAINVERB %INIT_VCH)
/<zuen>/                                     to do (aux.)
"*edun" AUX VERB @+AUXVERB %FIN_VCH
/<sua>/                                      the fire
"su" COMMON NOUN … @SUBJ %NCH)
/<.>/<PUNT_PUNT>/
```

Figure 3. Example of the anaphoric tagging.

This time we used this annotation format to present the identification of the anaphoric element and its referent. However our general annotation schema for representing linguistic information (Artola et al., 2004) is inspired on stand-off annotation following the recommendations of Text Encoding Initiative (Burnard and Sperberg-McQueen, 2002).

4. Annotation process

4.1. General guidelines

Here we will explain the steps carried out for establishing the general guidelines. The same corpus had been tagged twice with different purposes. The objective of the first annotation was to refine the general guidelines we had and it was performed by only one annotator. After that, we detected some problems along the annotation process and therefore the guidelines were changed and improved.

In the second annotation step, we saw the importance of the corpus being annotated the by different human taggers.

Following we present the general guidelines for this annotation step:

- As mentioned before, we focused on the tagging of the demonstratives *hau* (this/he/she/it), *hori* (that/he/she/it), *hura* (that/he/she/it) only when they behaved as a pronoun. In example *(2)*, the demonstrative *honek*[4] is a candidate to be tagged, while *honek*, in example *(3)*, is rejected as an anaphoric element, because it is a determinant.

  *(2) Lehen fasean **[Espainiari]**$_i$ irabazi zioten **[honek]**$_i$ lehen postuan ziurtatuta zuenean.*

  (In the first phase, they beat **[Spain]**$_i$ when **[it]**$_i$ had assured in the first place).

  *(3) Bihartik aurrera derrigorrezkoa izango da segurtasun gerrikoa jartzea. Neurri **honek** hainbat aldaketa ekarriko ditu.*

  (From tomorrow it will be obligatory to wear the seatbelt while driving. **This** measure will bring a lot of changes).

- Demonstratives can take several declension forms. The annotators had to annotate all of them excepting the genitive case (because the demonstrative does not form a noun chain by itself).

  *(4) **[Gidaria]**$_i$ auto barruan zegoen itota eta suhiltzaileen laguntza behar izan zen gorpua auto barrenetik ateratzeko. Mediku batek **[haren]**$_i$ heriotza ziurtatu zuen gero (...).*

  (**[The driver]**$_i$ was drowned in the car and firemen's assistance was necessary to take the corpse out of the car. Then, a doctor confirmed **[his]**$_i$ death (...)).

  In the example above, *haren* is not tagged as an anaphoric element because it is part of a noun chain. Some exceptions to this rule have been taken into account as it will be seen in the next point.

- Demonstratives in genitive belonging to complex postposition[5] are tagged. For example: ***haren aurka*** 'against him/her/it' is a complex postposition that plays an equivalent role to prepositions in other languages. We therefore decided to mark them as anaphoric elements.

  *(5) Gainera, **[Estradaren]**$_i$ abokatu batek **[haren]**$_i$ aurkako zantzuak ezabatzeko eskatu zion.*

  (In addition, one of the **[Estrada's]**$_i$ lawyers asked him to delete the traces against **him**$_i$).

The annotators were said not to tag neither the demonstratives belonging to cohesive elements nor those present in predicative sentences. More details will be given in section 4.4.

After establishing the mentioned guidelines two annotators began a second annotation.

[3] '*ANA*' for anaphora and *(num)* for its corresponding number, depending on its appearing order.

4 The demonstrative *honek* is the ergative form of *hau* (this/she/he/it).

[5] The annotators were provided with a list of these complex postpositions.

4.2. Ambiguity cases

Apart from identifying the linguistic information needed for later computational treatment of the pronominal anaphora, our goal in this work is to detect ambiguity cases.

Once the second annotation step has been finished, a third annotator examined the results produced mainly focusing on the disagreements between the annotators and the reasons for these disagreements. Table 1 shows the level of agreement when annotating the anaphora and its antecedent.

	HAU this/he/she/it	**HORI** that/he/she/it	**HURA** that/he/she/it
Anaphora	83%	94%	87%
Antecedent	62%	60%	76%

Table1. Agreement results.

Let us explain some aspects about the Table 1. The first row shows the agreement between the two annotators in the process of identifying the anaphoric element. The lower agreement in the identification of the demonstrative *hau* is due to the fact that this demonstrative is often part of cohesive elements as it will be explained later on.

The second row represents the agreement identifying the antecedent, once there is an agreement in the anaphoric element. The low figures are associated to the fact that, in this process, no precise guidelines were given to the annotators because we wanted to discover possible ambiguities in this phase of the annotation. The disagreements detected have constituted the basis in the definition of more precise guidelines for future works.

Following we explain the ambiguities detected by means of some examples. They are classified into three groups attending to the linguistic structure of the antecedent: noun phrases, sentences, and others.

1) **Noun phrases**. The disagreement occurs in two ways:
   a. When trying to delimit the boundaries of the referent if an apposition or a relative clause is part of the noun-phrase.

      (6) Realak gaur bertan fitxatu nahi du ***[[John Toshack]$_{i1}$ galestarra]$_{i2}$*** *entrenatzaile posturako. Horretarako, Saint Etiennera joan da Luis Uranga presidentea,* ***[harekin]$_i$*** *hitz egiteaz gain (...)*[6]

      (The Football Club Real wants to sign up today **[the Welsh [John Toshack]$_{i2}$]$_{i1}$** as a coacher. For this, the president of the club, Luis Uranga, has gone to Saint Etienne, apart from speaking with **[him]$_i$** (...)).

      In the example, one of the annotators identifies the whole noun phrase *John Toshack galestarra (*the Welsh John Toshack*)* as the referent of the anaphora *harekin* (with him), while the second one marked only the core of the noun phrase *John Toshack*. For future work we will consider the whole noun phrase as the referent.

   b. Two different noun phrases can be the referent of an anaphoric pronoun.

      (7) Erasoa jaso eta egun gutxira neskak, ustekabean, ***[erasotzailearekin]$_{i1}$*** *topo egin zuen kalean. Horri esker,* ***[gaztea]$_{i2}$*** *identifikatzea lortu zuen Poliziak. Atzo artean, ordea, Poliziak ez zuen* ***hura$_i$*** *harrapatzerik izan.*

      (A few days after suffering the attack, the girl, unexpectedly, bumped into **[the aggressor]$_{i1}$**. Thanks to that, Police could identify **[the young]$_{i2}$**. However, Police did not capture **him$_i$** until yesterday).

      The example illustrates this fact. *Erasotzailearekin* (the aggressor) and *gaztea* (the young) refer to the same person in this context.

2) **Sentences**. Two types of disagreement have been detected in those cases in which a sentence is the referent of the anaphoric element:
   a. To select the whole sentence or part of the sentence (e. g. a subordinate clause).

      (8) Prozedura horiek gainditu ostean, ***[[hitzarmena sinatuko dute]$_{i1}$ Kanputxeako Gobernuak eta NBEk]$_{i2}$.*** *Soilik* ***[horren]$_i$*** *ondoren osatu ahal izango da Khmer Gorriak epaitzeko tribunala.*

      (Once those procedures were overcome, **[[the Canpuche Government and the UN [will now sign the agreement]$_{i1}$]$_{i2}$**. Only after **[that]$_i$**, it will be possible to constitute the court to judge the Red Khmers).

   b. To select different parts of the same sentence although being semantically different. This can be linked to the lack of context.

      (9) ***[Telebista publikoan]$_{i1}$,*** *aldiz,* ***[guduka gogorrak]$_{i2}$*** *egon ziren, baina* ***[hau]$_i$*** *ere Gobernuaren indarrek bereganatu zuten atzo goizerako.*

      *(However, t*here were **[very hard struggles]$_{i1}$ [in the public television]$_{i2}$**, but **it$_i$** was also taken over by the Government forces by yesterday morning).

3) **Others.** Difficulty to select a reference in the near context.

   (10) Zentzu horretan gerra bat pizteak "ondorio larriak" izango lituzke Fronte Polisarioko ordezkariaren iritziz, "lurralde horretako egonkortasunean, ***[Kanariar artxipielagoan]$_{i1}$ [Ø]$_{i2}$*** *barne, ezin baita* ***[hura]$_i$*** *bere inguruko herrialdeetan gertatzen diren liskarrei bizkarra emanda bizi".*

   (In that sense, according to the representative of the Polisario the breaking of a war would bring "serious consequences" "in the stability of that country, including **[the Canary Islands]$_{i1}$ [Ø]$_{i2}$**, because **[they]$_i$** can not live ignoring what happens in the surrounding countries".)

   In example *(10)*, one of the annotators has not recognized any antecedent in the context whereas the other has marked *Kanariar artxipielagoan*

[6] In these examples we use $_{i1}$ for the selection of one annotator and $_{i2}$ for the selection of the other.

(the Canary Islands) as the referent to *hura* (they)**.**

4.3. Linguistic Conclusions

Finally, we studied some linguistic features such as the distance between the anaphoric element and its antecedent; the grammatical features of the referent (whether it is noun phrase or sentence). Besides, we detected that, in most of the cases, anaphora and referent agreed in number while information about declension case was not relevant at all. The result of this study is illustrated in Table 2.

	HAU (this/ he/she/it)	**HORI** (that/ he/she/it)	**HURA** (that/ he/ she/it)
Pronouns	**151**	**210**	**215**
Anaphora	**96 %**	**99 %**	**99 %**
Cataphora	4 %	1 %	1 %
Referent in the same sentence	**38 %**	**52 %**	**65 %**
Referent in the previous sentence	34%	42 %	26 %
Others	28 %	6 %	9 %
Referent Noun Phrase	**67 %**	21 %	**99%**
Referent sentence	33 %	**79 %**	1%
Number agreement	99 %	**100 %**	99 %
Number disagreement	1 %	**0 %**	1 %

Table 2. Results of the annotation.

The demonstrative pronoun *hau* (this) appears 151 times in a 50.000 words corpus. With respect to the position of the anaphora's referents, in 96% of the cases they appear before the anaphoric elements, and in only 4% of the cases are cataphora, that is to say, the referent comes after the anaphoric element.

Regard to the distance between the pronoun *hau* (this) and its referent, in 38% of the cases they appear in the same sentence, in other cases they are in the previous sentence and just in fewer cases they come in the two or three previous sentences.

Finally, we have detected that in 67% of the cases the antecedent, is a noun phrase, while 33% it is a sentence. Here we have a representative example of the first case:

(11) ***[Ben Amor]***$_i$ *ere ez da Mundiala amaitu arte etorriko Irunera,* ***[honek]***$_i$ *ere Tunisiarekin parte hartuko baitu Mundialean.*

([Ben Amor]$_i$ is not coming to Irun before the world championship is finished, since **[hɨ** $_i$ will play with Tunisia in the World Championship).

The demonstrative *hori* (that) appears 210 times in the same corpus. In 99% of the cases, it is anaphora and in the 52% of cases, the referents appear in the same sentence. Regarding the structure of the referent, most of them (79%) appear in clauses and they make reference to an idea or a proposition, rather than to a specific element or person. For example:

(12) ***[Euskaraz egiten den musika oro mespretxatzea]***$_i$, ***[hori]***$_i$ *dun kezkagarria iruditzen zaidana.*

([Despising all music done in Basque]$_i$**, [that]**$_i$ is what I think is worrying).

The third demonstrative determiner we have analyzed *hura* (that), is occurs 215 times. This demonstrative is almost always anaphora (99%), rather than cataphora. In 65% of the cases it is in the same sentence of the anaphora and only sometimes in the previous sentence (26%). Almost all the references (99%) are noun phrases and more specifically they are usually proper nouns.

(13) Banesto galduta dabil, ***[Miguel Indurain]***$_i$ *erretiratu zenetik,* ***[hura]***$_i$ *ordezkatuko duen norbaiten bila baitabiltza.*

(Since **[Miguel Indurain]**$_i$ retired, Banesto is lost. They are looking for somebody to replace **[him]**$_i$).

4.4. Open questions

One of the advantages of corpus-based study of language phenomenon is the fact that the researchers realize about aspects that are not apparent. During these studies it always remains a list of open questions to be analyzed and refined in more detail in the future:

a) Cohesive elements in Basque often include a demonstrative: *hau da* (that is to say), *harekin eta honekin* (with that and this), *honetaz gain* (apart from this), *horren ondorioz* (as a consequence of that). We considered they do not really have any specific referent because its vagueness (sometimes the referent can be the whole previous paragraph or even the whole text). Our annotation guidelines define these elements as not anaphoric.

(14) Nafarroako zergak nafarron artean ordaintzen ditugu, eta ordainketa hori egiten dugunok eskubide osoa dugu gure iritzi, adostasunedota kexa adierazteko. ***Honekin batera****, Nafar Goabernuari proposatzen diot gaia serio hartzeko (...)*

(The taxes of Navarra are paid by all the inhabitants of Navarra, and those who pay them are within our rights to express our opinion, approval or complaints. **Together with that**, I propose the Government of Navarra to consider this matter seriously).

b) Demonstratives in sentences with predicative noun phrases were not tagged because it was often difficult to decide whether a noun phrase is used predicatively or referentially, like in the next example:

(15) Jolasa baita ***hau****.*

(Because **it** is a game).

5. Conclusions and future work

This is the first study carried out on annotation of pronominal anaphora in Basque. It has been a useful start in defining criteria for anaphora's annotation.

Based on these criteria, we plan to tag a larger corpus in the near future. In order to facilitate this task, we are trying to adapt to our format other existing tools like the 3LB-RAT annotation tool (Saiz Noeda et al., 2004) or the MMAX tool (Müller and Strube, 2003).

In regards to the specific features of the annotated demonstratives, we can say that some of them corroborate the consulted bibliographic statements, while others simultaneously open new perspectives to continue researching other characteristics of the anaphora and its referents. For the moment, we are mainly interested in defining the sources of knowledge needed to identify the referents such as other morphological information, syntactic dependencies or semantic features.

The results obtained from this work will be helpful for allowing the development of an automatic anaphora resolution tool for Basque.

6. References

Aduriz I, Agirre E, Aldezabal I, Alegria I, Ansa O, Arregi X, Arriola J.M, Artola X, Díaz de Ilarraza A, Ezeiza N, Gojenola K, Maritxalar A, Maritxalar M, Oronoz M, Sarasola K, Soroa A, Urizar R, Urkia M. *A Framework for the Automatic Processing of Basque.* In Proceedings of the First International Conference on Language Resources and Evaluation. Granada, Spain. (1998).

Aduriz I., Aranzabe M. J., Arriola J.M., Díaz de Ilarraza A., Gojenola K., Oronoz M., Uria L. *A Cascaded Syntactic Analyser for Basque.* CICLing 2004. Seoul, Korea. (2004).

Artola X., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Sologaistoa A., Soroa A. *EULIA: a graphical web interface for creating, browsing and editing linguistically annotated corpora.* LREC 2004. Lisbon, Portugal. (2004).

Burnard, L. & Sperberg-McQueen C. M. *TEI P4*: *Guidelines for Electronic Text Encoding and Interchange.* Oxford, Providence, Chalottersville, Bergen: Text Encoding Initiative Consortium. (2002).

Ezeiza N., Aduriz I., Alegria I., Arriola J.M., Urizar R. *Combining Stochastic and Rule-Based Methods for Disambiguation in Agglutinative Languages.* COLING-ACL'98, Montreal (Canada). (1998).

Ferrández Rodríguez A. *Aproximación computacional al tratamiento de la anáfora pronominal y de tipo adjetivo mediante gramáticas de unificación de huecos.* PhD thesis. University of Alicante. (1998).

García Azkoaga I. *Erlazio anaforikoak argudiozko testuetan.* In Koherentzia, kohesioa eta konexioa: testuratze baliabideak. Hizkuntzaren azterketa eta irakaskuntza, Larringan & Idiazabal (ed.). Gasteiz: EHU-UPV & Arabako Foru Aldundia. (1998).

Hirst G. *Anaphora in Natural Language Understanding.* Berlin: Springer-Verlag. (1981).

Hirschman, L. *MUC-7 coreference task definition.* Version 3.0. (1997).

Hualde J.I. & Ortiz de Urbina J. *A grammar of basque*. Berlin: Mouton de Gruyter. (2003).

Karlsson F., Voutilainen A., Heikkilä J. Anttila A., ed. *Constraint Grammar: a language-independent system for parsing unrestricted text*. Vol. 4 Natural Language Processing. New York, Berlin Mouton de Gruyter. (1995).

Koskenniemi K. *Two-level Morphology: A general Computational Model for Word-Form Recognition and Production*. University of Helsinki, Department of General Linguistics. Publications 11. (1983).

Laka I. *A Brief Grammar of Euskara, the Basque Language*. HTML document. Euskarako errektoreordetza, EHU. http://www.ehu.es/grammar (2000).

Mitkov R. *Anaphora resolution*. London: Longman. (2002).

Müller C. and Strube M. *Multi-level annotation in MMAX.* In Proc. of the 4th SIGDIAL. Sapporo, Japan. (2003).

Navarro B., Civit M., Martí M. A., Marcos R., Fernández B. *Syntactic, semantic and pragmatic annotation in Cast3LB.* In Proceedings of the Shallow Processing of Large Corpora. A Corpus Linguistics Workshop, Lancaster, UK. (2003).

Palomar M., Civit M., Díaz A., Moreno L., Bisbal E., Aranzabe M. J., Ageno A., Martí M.A. y Navarro B. *3LB: Construcción de una base de datos de árboles sintáctico-semánticos para el catalán, euskera y español.* XX. Congreso SEPLN, Barcelona. (2004).

Poesio M. *The MATE/GNOME scheme for anaphoric annotation, revisited.* Proc.. of SIGDIAL, Boston. (2004).

Saiz Noeda M., Izquierdo R. *3LB-RAT: una herramienta para la anotación referencial.* Iberamia 2004. Puebla, México. (2004).