TITLE

Coreferential relations in Basque: the annotation process

Klara Ceberio. IXA Group, UPV/EHU, klara.ceberio@gmail.com

Itziar Aduriz. Department of General Linguistics, Universitat de Barcelona, itziar.aduriz@ub.edu

Arantza Díaz de Ilarraza, IXA Group. UPV/EHU, a.diazdeilarraza@ehu.es

Ines Garcia Azkoaga. Department of Basque Language and Communication, UPV/EHU ines.garciaazkoaga@ehu.es

ABSTRACT

In this work we present the coreferential tagging of part of the EPEC Corpus of Basque. Although coreference is a pragmatic linguistic phenomenon highly dependent on the situational context, it shows some language-specific patterns that vary according to the features of each language. Due to the fact that Basque is not an Indo-European language, it differs considerably in grammar from the languages spoken surrounding areas. We will explain these features and the decisions made in each case.

The first steps of this work began with the research focused on pronominal anaphora (Aduriz et al. 2005 and Ceberio et al. 2008), and we extended the topic to coreferential relations in a broader sense, including other types of referential structures, such as proper names, nominal and adverbial anaphora.

After describing the criteria defined for coreferential tagging in Basque, we will explain the annotation process. Our annotation is based on a morphologically and syntactically annotated corpus that provides us with a manageable environment, in which the specific structures that are part of a reference chain can be more easily identified.

A part of the corpus was tagged by two annotators who marked up the same text independently, and by another annotator that acted as judge, solving problems in case of disagreement.

All this process has been automatized as a result of previous studies carried out in this field. The automatic detection of mentions (Soraluze et al. 2012) has provided us with a better working environment, and given us the possibility to build a first significant corpus for a later computational treatment of automatic coreferential resolution.

1. Introduction

It is widely accepted that coreference relations occur when different linguistic structures refer to the same person, object or event (Recasens 2010).

These types of relations are an important part of the understanding of discourse. In fact, language allows users to identify elements or events that are connected to the same entity, thus facilitating the correct understanding and continuity of the text.

Coreference resolution is one of the core tasks of Natural Language Processing (NLP). The main goal of coreference resolution is to identify all the linguistic expressions that refer to the same entity in a text, and many other NLP applications can benefit from an improvement in the results obtained in this task. Information Extraction (Mc-Carthy and Lehnert 1995), Text Summarisation (Steinberger et al. 2007), Question Answering (Vicedo and Ferrández 2006), Machine Translation (Peral et al. 1999), Sentiment Analysis (Nicolov et al. 2008) and Machine Reading (Poon et al. 2010), need this type of information for a better performance of their tasks.

However, coreference resolution is generally considered a high level and difficult task, because it requires previously solve other NLP problems, such as morphosyntactical analysis, lemmatization and syntactic function identifier, named entity recogniser and wordsense disambiguation.

Coreference resolution was seen as natural next step in the automatic processing of the language, and therefore it was introduced in the main strategy of IXA research-group[1]. As in many other languages (Mitkov 2002), we first began studying the subject of anaphora resolution, where the goal was to find the right antecedent of anaphors, linguistics expressions that usually point back to a previously mentioned expression in the discourse (Aduriz et al. 2005). The next step was to extend the topic of our research to coreference relations in a broader sense (Ceberio et al. 2008), including other types of referential structures, such as proper names, nominal and adverbial anaphora.

The aim of this paper is to present our ongoing research and it is structured as follows. We first give an overview of the phenomenon of coreference that will help us in our research in section 2. After that, the criteria defined for coreferential tagging is defined; addressing some linguistic issues and taking into account the features of the Basque

---

[1] http://ixa.eus/Ixa

language. In section 4 the annotation process is explained, first, the automatically part and some inter-annotator disagreements will be outlined. The paper concludes with an overview and summary of this tagging and we briefly explain how this annotation will help in the automatic coreference annotation of Basque.

2.   What is coreference?

A text may be grammatically correct, but incomprehensible. This is because, apart from being grammatically correct, a text needs to be coherent, in other words, the topics of the discourse have to be disposed in the right order, using the right mechanisms of cohesion.

If we consider the text as a whole, cohesion assures the development of the topics and the subtopics. In this work, we use the term *topic*, for the entities that appear in the discourse. In an oral or written production (text), there are some entities and discourse objects, and consequently, referential expressions, that can be linked to each other by different types of relations. Our study will be focused on *coreferential relations* and the very close related *anaphoric expression.* Let's explain the phenomenon with an example[2]:

(1)   [Alfredo Salazar Baskoniako orduko idazkari teknikoak]$_i$ lan handia egin zuen. Batik bat gauza bat deskubritu zuen [hark]: Argentinako harrobia. Hain zuzen, orain dela hamar urte egin zuen Argentinarako lehen bidaia [Salazarrek]$_i$. (...)
"Gaur egun txapelketa sendoak dira hangoak. Jokalari gazte ugarik parte hartzen duela da ezberdintasun nabariena Andres Nocioni da horren adibide argienetakoa: hamabost urte zituenerako egin zuen debuta Argentinako txapelketan; orain hogei urte ditu". [Salazarrek]$_i$ normaltzat du Nocionik hain gazte egitea debuta, lehen esan bezala Argentinako saskibaloia ez delako hain serioa ere, eta, [Baskoniako idazkari tekniko ohiaren]$_i$ ustez, zaila da horrelakorik gertatzea ACBn, lehia handiko liga delako.

'[Alfredo Salazar the technical secretary of Baskonia]$_i$ worked hard. [He]$_i$ specially discovered one thing: the youth system (the pool of young) in Argentina. Indeed, [Salazar]$_i$ travelled to Argentina for the first time ten years before .
(…)
Nowadays, the league has become more stable there (in Argentina). The big difference

---

[2] Most of the examples of this paper come from the EPEC corpus explained in section 4.

is that a lot of young players participate there, Andres Nocioni is the clearest example: he made his debut when he was 15 in the Argentinian League; now, he is 20". For [Salazar]ᵢ it is normal that Nocioni made his debut so young, as said before the Argentinian basketball is not so serious, and, according to [the former technical secretary of Baskonia]ᵢ, it would be unlikely for it to happen in the ACB league, because it is a more competitive league.

Generally speaking, it is said that reference is the property of linking a linguistic expression with something in the real world (Saeed 2009). This 'something' of the real world will be the reference or referential entity. In addition, this referential entity can be personal, spatial or temporal.

Besides, we speak about coreference, following Moirand's definition (1990): "two elements will be coreferent if they represent the same referential entity of the universe".

In this way, in the example above, all the linguistic expressions in square brackets refer to the same person, *Alfredo Salazar*. All of them (['Alfredo Salazar the technical secretary of Baskonia'], ['He'], ['Salazar'], ['the former technical secretary of Baskonia']) are coreferent and part of a coreference chain.

The notion of anaphora appeared together with the study of the discourse and texts. It is one of the most important elements for creating cohesion. The referential connections between the elements of a text are very closely related to the way it is interpreted: "the anaphora puts in relation an element whith another that has been mentioned before or after it (*cataphora*)" (Garcia-Azkoaga 2003: 77).

Moreover, in the case of anaphoric relations anaphoric expressions are semantically dependent on the antecedent (Cornish 1999), and the information of the antecedent is necessary to make the right interpretation of the anaphoric element. For example the pronoun in the first text ['he'] has no meaning itself; we do not know which person it refers to until we identify the antecedent ['The technical secretary of Baskonia, Alfredo Salazar']. This phenomenon is the so called anaphoric relation.

The relation between an antecedent and the referential expression that retrieves it can be anaphorical, but that does not mean that this relation is always coreferent (Garcia Azkoaga *in press*). We accept that the relation between the antecedent and the anaphoric expression it is not symmetrical. Sometimes the anaphor and the antecedent will have the same referent (['the former technical secretary of Baskonia'] and ['he']), but the empirical analysis of real texts shows it is not always the case.

There are some cases, for example the associative anaphors ((2)) or evolutive referent cases ((3)) (Garcia-Azkoaga 2003).

(2) [Herri batera]$_i$ iritsi ginen. [Eliza]$_i$ tontor batean zegoen.

'We arrived [in a town]$_i$. [The church]$_i$ was situated on a hill.'

(3) Har itzazu [lau sagar]$_i$. Zuritu eta zatitu. Eduki egosten ordu erdiz. Txiki-txiki egin. Hoztu ondoren, zerbitzatu [konpota hori]$_i$ gailetatxoekin.

'Take [four apples]$_i$. Peel and cut them. Cook them in boiling water for 30 minutes and pound them. After cooling, [the compote]$_i$ can be served with cakes.'[3]

In the same way, it is possible that linguistic expressions can be coreferent but they do not establish any anaphorical relation. In the case of named entities or NP repetitions (['Salazar'] in the first example), they can be interpreted without going back to the antecedent, they are linguistic expressions that can be interpreted independently. In this case we speak of two linguistic expressions that can be interpreted independently, which are connected with coreferential relations (Kleiber 1994: 22).

All these concepts have been analized from a Natural Language Processing perspective (Botley & McEnery 2000, Mitkov 2002, Stede 2011), and as it is the case in the field, these phenomena have been studied using real corpora.

For this purpose, corpora have been tagged at referential level in many languages. One of the objectives of this work is to describe and define what type of expressions are candidates for coreferential chains (3.1.), as well as to explain the coreferential chains formed by these referential expressions (3.2.).

As Mitkov (2002:130) stated "the annotation of corpora is an indispensable, albeit time-consuming, preliminary work to anaphora resolution (and to most NLP tasks or applications)". The system will learn through these annotated corpora how each phenomenon works, and it will be able to extract some patterns to detect them later in raw corpora. With the aim of building a corpus annotated at coreferential level, we studied different theories of coreference and many works developed for other languages (Aduriz et al. 2005, Ceberio et al. 2008).

---

[3] In generall, the examples in English may have more linguistics expressions that refer to the same entity but we only mark the equivalents of the elements annotated in Basque. See section 3.1 for a detailed explanation.

Although coreference is a pragmatic linguistic phenomenon highly dependent on the situational context, it shows some language-specific patterns that vary according to the features of each language.

Basque is not an Indo-European language and differs considerably in grammar from the languages spoken in the surrounding regions. It is an agglutinative language, in which grammatical relations between components within a clause are represented by suffixes. The next section will show these linguistic features and how we adapted the annotation of coreference to the Basque language.

3.   Linguistic features for automatic coreference resolution

In all languages, we divide the elements of a text into two groups: in the first group we have the so called *referring expressions* and in the second one the *non-referring expressions* (Saeed 2009).

On the one hand, there are linguistic expressions that do not have the property of referring to something, this is the case of some adverbs (*often*)[4] oin oharra gehitu aurrerago azalduko dugula esanez, lekuzko adberbioen kasua, verbs *(show)* or some particles (*as*). This type of words are part of the discourse and they have their meaning, but they do not refer directly to an entity of the real world, and in this sense we call them non-referring expressions. On the other hand, we have the referring expressions; all the people who speak English, for example, know that ['the horse'] in example ((4)) is referring to a concrete animal. Nouns are almost always considered referring expressions.

(4) Ikaragarri sufritu zuen [zaldiak] lasterketan.

'[The horse] suffered during the race.'

In other words, all the referring expressions have specifc grammatical features. As we mentioned, this referrering expressions are possible *mentions* (*zaldi* 'horse'). These are linguistic expressions that have the possibility to refer to an object or set of objects in the world (as defined by ACE, Doddington et al. 2004).

---

[4] Due to the Basque language features, the local adverbs may take this referring property (see Section 3.3.5)

In the next section we will explain what type of grammatical features should these referring expressions have.

3.1.   Mentions: morphosyntactical features

Taking into account elements that have a cohesive function, coreference occurs mainly at noun phrase level. For that reason we decided to focus our study on coreferential relations between noun phrases. In this section we will explain the decisions we adopted regarding the annotation, and what morphosyntactical information is relevant to decide if a noun phrase is a *mention* or a *markable*.

Basque is described as a three-way pro-drop language (Ortiz de Urbina & Hualde 1989): the subject, direct object and indirect object (marked ergative, absolutive and dative) are not always explicit, but the verb gives the recipient all the information about them:

> (5) Ø (ERG) Ø (DAT) Ø (ABS) eman d-i-zki-da-zu.
>   (PRS-ROOT-3PL (ABS)-1SG.DAT-2SG.ERG).
>   'You gave me them.'

The meaning of the three empty elements in (5) is equivalent to that of personal pronouns *zuk* 'you' (2nd singular ergative), *niri* 'me' (1st singular dative) and *horiek* 'them' (3rd plural absolutive) respectively.

Besides, another feature of Basque is the lack of grammatical gender in the morphological system. This is a remarkable difficulty for the automatic resolution of coreference.

The classification we propose is based on the Basque grammar (Euskaltzaindia 2002), and the latest annotations carried out in other languages (Nilsson Björkenstam 2013, Pradhan et al. 2007, Recasens 2010, Rodríguez 2010).

3.1.1.   Pronouns

Being a pro-drop language, personal pronouns are not as frequent as in other languages. They refer to persons and they form the noun phrase, without any other element.

As the Descriptive Grammar of the Basque Royal Academy points out (Euskaltzaindia 1985), in Basque the 'true' personal pronouns are the (*ni* 'I', *gu,* 'we') and second person (*hi* 'you', *zu* 'you', *zuek* 'you (plural)') pronouns.

(6) [Gu] ere gogor defenditzen saiatuko gara, hor erabakiko baita norgehiagoka.

'[We] will defend fiercely, because this will decide the match.'

In Basque, there are no distinct forms for third person pronouns (Laka 1996), and demonstratives are used as third person pronominals *(hau* 'this'*, hori* 'that' (nearer than *hura), hura* 'that').

There is a third series of intensive demonstratives, which are formed adding a prefix to the demontratives, and which has special relevance *(berau, berori, bera* 'he himself/she) because they are always used anaphorically.

(7) [Jacken] bila joko dute, nahiz eta [hark] nahiago duen elurretan jolasten segitu.

'They will look for [Jack], although [he] would like to continue playing in the snow.'

We considered the **reciprocal pronouns** in this group: *elkar* (*'each other')*. There is a variant of this reciprocal *bata bestea (*literally *'the one the other')*, and they always refer to a previous mentioned element in the same sentence.

(8) [Hiru automobilek]$_i$ [elkar]$_i$ jo zuten Gorlizko Andra Mari auzoan.

'[Three cars]$_i$ crashed ([to each other]$_i$) in the Andra Mari quarter of Gorliz.'

In the case of Basque reflexive pronouns make reference to a body part, *X-ren burua* (literally 'my own head'). This is the pronoun corresponding to English 'myself'.

(9) Gizon hark ez zuen [bere burua] besterik maite.

'That man only loved [himself].'

3.1.2. Possessives

We include possessives and demonstrative genitives in this category.

The **possessive pronouns** form by themselves a noun phrase as in the case of personal pronouns and reciprocals, but they can refer to two things at the same time: they refer to an entity and express the owner of this object.

(10) Escuderok euskal musika tradizionala eraberritu eta indartu zuen. [Harenak] dira, esate baterako, *Illeta*, *Pinceladas Vascas* eta *Eusko Salmoa* obrak.

'Escudero renewed and gave prominence to traditional Basque music. The works *Illeta, Pinceladas Vascas* and *Eusko Salmoa*, for example, are [his].'

The possessive pronoun *harenak* ('his'), refers to the ownership of the musician, but at the same time it is referring to the musician's musical works (*Illeta, Pinceladas Vascas eta Eusko Salmoa obrak*, 'The works *Illeta*, *Pinceladas Vascas* and *Eusko Salmoa*'). Possessives can only be determiners, and we have said that we were only taking noun phrases into account, but in the case of possessive determiners the embedded possessives will be considered as mentions:

(11) Epitieren kasuan, [[bere] helburua] lortu dezakeela dirudi eta baliteke denboraldia Lehen Mailan hastea.

'In the case of Epitie, it seems that he could achieve [[his] aim] and possibly start the football season in the Premier League.'

The demonstrative *bere* ('his') refers to the footballer *Epitie,* so that we have to consider this type of possessive determiners or we would be missing the corresponding coreferencial relation.

3.1.3. Noun Phrases

When speaking about noun phrases, we generally take into account noun phrases whose head is a noun. But let us explain what we are talking about when speaking about noun phrases in Basque.

- Regular noun phrases

For coreference annotation we distinguish three types of noun phrases: noun phrases that end with indefinite articles (12), general definite NPs (13) and NPs ending with the demonstrative determiner (14).

(12) Inguruko etxeetan [zenbait beira] hautsi arren, kristalek ez zuten inor zauritu.

'Although [some windows] were broken in the surrounding area, the glasses did not hurt anyone.'

(13) Txileko Gorte Gorenak baimena eman dio Juan Guzman epaileari [Pinocheten erregimenak ustez Argentinan egindako hilketak] ikertzeko.

'The Supreme Court of Chile has authorized the judge Juan Guzman to investigate [the killings committed by the Pinochet regime in Argentina].'

(14) Zenbait hilabetetako ikerketa behar izan zuten [auto hura] David Harrisek lapurtu zuela jakiteko.

'It took months of investigation for the police to discover that [that car] had been stolen by David Harris.'

We divided the noun phrases into two groups based on the definitness of the noun phrases: indefinite noun phrases, which usually add new information in the discourse; and definite or demonstrative noun phrases, if they have been previously metioned in the text. Zabala (1996:35) says "in coreference chains both the definite determiner and the demonstratives, are complementary to other indefinite articles".

- *Noun phrases headed by proper nouns*

When a proper noun is the head of the noun phrase, the referent is directly named and we therefore say that there are names that have an exact reference (*London* is the capital of Great Britain, there is only one). In Basque, in general, the proper names do not need any determiners, but there are some exceptions that will take determiners *(Frantziako Gobernua* 'France Government').

(15) Epailea [Londonera] joango da.

'The judge will go to [London].'

- *Noun-phrases with a subordinate clause*

The head of these mentions is always a noun complemented by a subordinate clause. In (15) the head-noun is complemented by a subordinate clause of the type that is called,

for Basque, a complementary clause. We take the whole stretch of the NP (both the subordinate clause and the head noun) as a mention.

(16) [[[DINAk] [Argentinan] egindako krimenak] ikertzeko baimena] eman du Txileko Gorte Gorenak.

'The Supreme Court of Chile has given [[[permission to investigate [the crimes] [DINA] committed in [Argentina]].'

The head of the noun-phrase is *baimena* ('permission'), but the subordinate clause has other noun-phrases that will be considered as mentions.

- [[permission to investigate [the crimes] [DINA] committed in [Argentina]]
- [the crimes DINA commited in Argentina]
- [DINA]
- [Argentina]

In addition, relative clauses can add information to nouns as in example (17). In that case the boundaries of the mention are set from the beginning of the relative clause to the end of the NP, and include all noun-phrases contained in the NP.

(17) [[Igandeko partiduak] duen garrantzia] dela eta, lasai egotea beharrezkoa dutela esan zuen Lotinak.

'Lotina said that it is necessary to stay calm because of [the importance that [Sunday's match] has].'

- *Coordinated noun-phrases*

In the case of coordination, the nominal groups of a conjoined NP are extracted. We also regard as mentions the nested NPs (*siesta* 'a nap' and *atsedena* 'a rest') and the whole coordinated structure (*siesta eta atsedena* 'a nap and rest').

(18) Bazkal ondoren [[siesta] eta [atsedena]] besterik ez zuten egin.

'After lunch they did nothing but have a [[nap] and [rest]].'

- *NPs as part of complex postpositions*

Basque has a postpositional system. In this type of languages postpositions play a role similar to that of prepositions in languages like English or Spanish, so that postpositions suffixes are attached to the last element of the phrase. They are defined as "forms that represent grammatical relations among phrases appearing in a sentence" (Euskaltzaindia, 1994). There are two main types of postpositions in Basque; the first one is a suffix appended to a lemma *(etxean, '*at home*')* and, the second one is a suffix followed by a lemma (main element) that can also be inflected, the so called complex postposition *(etxetik kanpo, '*outside home*').*

The last type of elements has been termed as complex postposition. This term is used to name the whole sequence of two words involved, and not just to refer to the second element. (Zabala & Odriozola 2004, Arriola et al. 2013).

We only take into account as a mention the first part of the postpositional NP, the noun and the attached suffix that goes before the second lemma of the complex postposition. In (19) the second part of the complex postposition is *kanpo* ('outside'), and we annotate the noun that precedes it *(etxetik* 'home [from]').

(19) Zaragozaren beherakada ere horrelaxe iritsi da: [etxetik] kanpo eskas aritu arren, [etxean] partida guztiak irabazten hasi zen, azken boladan etxeko partidetan ere kale egiten hasi diren arte.

'This is why Zaragoza went down; although they played badly [away], they won all the matches [at home], in the last season they are failing at home too.'

- *Verbal nouns*

As in many other languages in Basque verbs can be nominalised. When the nominalised verb works as the head of the mention and takes the corresponding case marking suffix, the whole clause governed by the verbal noun has to be annotated.

(20) [Instalazio militarrak ixtea] eskatuko dute.

'They will ask for the [closing the military installations].'

But sometimes it is not easy to recognize what is the real function of such type of verbal nouns. They show both features of nouns and verbs and it is said they have a double nature. That is why we established two conditions to tag the verbal nouns as nouns: a)

the subject of the verbal noun takes the genitive case as in example (21) or b) they are followed by an independent determiner (22).

(21) [EBren zabaltzeak] arazo asko konpontzera behartuko du.

'[The opening of the EU] will force them to solve many problems.'

(22) Onik baizik ez dezake ekar [zabaltze horrek].

'[The opening] can only bring good things.'

- Ellipsis

In Basque ellipsis is a broad phenomenon. At morphosyntactical level, a noun-ellipsis occurs when the suffixes attached to the word correspond to a noun, although the noun is not explicit in the word. We consider this type of ellipsis in the case of verbs that take suffixes indicating noun-ellipsis, as in example (23). The presence of the ellipsis implies the existence of both the verb (sailkatu zen-, 'finished') and the ellipsis (-Ø-ak (sailkatu zen-Ø-ak 'Ø who finished'). All the information corresponding to both units is stored an treated as a noun.

(23)   [Bigarren sailkatu zenak] segundo bakarra kendu zion.
'[Ø who finished in second place] only had a second's advantage.'

As mentioned before, at sentence level, the subject, direct object and indirect object of a clause (represented by the ergative, absolutive, dative cases) can be elided in Basque. The morphological information about these elements, as the number or person, is always given by the verb (even if it is explicit in the sentence). These elliptical pronouns are not marked in our tagging, as the previous morphosyntactic analiser did not take into account this type of ellipsis.

(24)   [Ø] Ez zuen podiumean izateko itxaropen handirik.
'(*He*) Ø did not have much hope of being on the podium.'

3.1.4. Adverbs

Although we have been talking about noun-phrases, sometimes adverbs have a cohesive function, and, in the corpus there are a lot of local adverbs that have the cohesive function. In Basque there are three 'true' place adverbs: *hemen* 'here', *hor* 'there (near here)', *han* 'over there'. These adverbial forms, derived from the demonstratives *hau* 'this', *hori* 'that (near here)' and *hura* 'that' (Hualde & Ortiz de Urbina 2003), show three degrees of proximity and establish an intratextual reference. Their interpretation is anaphoric, they usually refer to a place or space previously mentioned in the text. There is another adverb, *bertan*, that can be used in all three cases (*hemen bertan* 'here', *hor bertan* 'there (near here)', *han bertan* 'over there').

(25) Etxeko atezainari giltza bat utziko ziola esan zion, [apartamentura] igo eta [han] itxaroteko.

'She told him, she would leave a key to the doorman, so he could go upstairs ([apartment]) and wait [there].'

3.2. The annotation of coreference chains

We have already mentioned, anaphor and coreference are phenomena that can coincide. Sometimes there are anaphoric relations based on coreference, but there are other cases where the anaphoric relation does not exist. The annotation we present in this paper is restricted to coreference relations.

In this section we will specify which type of coreference relations have we taken into account in our annotation. Since the coreference relation is defined as identity of reference between two or more mentions, when establishing the type of relation betweeen the mentions, we always look for the relation of each mention with its previous mention. That is to say, if we have a coreference chain with three mentions (*A*, *B* and *C*), the annotators should specify the relation between *b* and *a* and the relation between *C* and *B* (and no the relation between *C* and *A*), as can be seen in the next figure:

$$[\text{MentionA}]_1 \quad \leftarrow \quad [\text{MentionB}]_2 \quad \leftarrow \quad [\text{MentionC}]_3$$

(26) [Peter Gentzel atezain suediarra]$_1$ ez da, azkenean, Portland San Antoniora etorriko. Nafarrek egindako eskaintzari ezezkoa eman dio [jokalariak]$_2$, eta Granollers taldearekin jokatuko du hurrengo denboraldian. [Gentzelek]$_3$ atzo

goizean hitz egin zuen Portlandeko entrenatzaile Zupo Ekisoainekin, Granollersek egindako eskaintza onartzeko arrazoiak azaltzeko.

> '[The swedish goalkeeper Peter Gentzel]1 is finally not coming to Portland San Antonio. [The player]$_2$ refused the offer made by the people from Navarre and will play with the team Granollers next season. [Gentzel]$_3$ spoke yesterday to the trainer of Portland Zupo Ekisoain, to explain the reasons to have accepted the offer of Granollers.'

The relations to be specified in the example (26) are: between [The player] and [the Swedish goalkeeper Peter Gentzel]; and between [Gentzel] and [The player]:

[The Swedish goalkeeper Peter Gentzel]$_1$ $\leftarrow$ [The player]$_2$ $\leftarrow$ [Gentzel]$_3$

3.3. Coreferential relations

Over the last years there have been many proposals for annotation schemes (MUC, ACE, OntoNotes) but none of them is considered as standard and can be used for every language and every corpus (Recasens 2010).

Therefore, we will give more details about the coreference classification we proposed for Basque language and our corpus (EPEC Corpus). The following table summarizes the semantic relations we propose:

Type of links			
Coreference	Pronominals		
	Nominals	Repetition	
		Partial repetition	
		No repetition	Synonyms
			Hyponyms
			Hypernyms
			Generic hypernyms
			Inverse instances
			Others
	Proper nouns	Repetition	
		Partial repetition	
		Instances	
		Others	
	Ellipsis		
	Adverbs		
Other relations	Predicative		
	Apposition		

1. Classification of Basque coreference annotation

3.3.1. Pronominals

The relations established between the pronouns and the referring noun phrase are called pronominals. As mentioned before Basque lacks of real pronouns, and demonstratives are used in their place (*hau* 'this', *hori* 'that' (nearer than '*hura*') and *hura* 'that') with all declension cases.

(27) Aimar Olaizolak pazientzia asko dauka. [Kontrarioa]$_i$ estu hartzen saiatzen da beti, [hura]$_i$ nekarazten.

'Aimar Olaizola is very patient. He tries to take [the rival]$_i$ seriously, and make [him]$_i$ tired.'

The corpus to be annotated is composed mostly of newspaper texts, and in the case of interviews, quoted speech is used. In quoted speech the first and second person pronouns become anaphoric, because they can only be interpreted by identifying its antecedent (28).

(28) Egun hauetan Atlantako zilarraz galdetu diogun bakoitzean hura errepikatzea ia ezinezkoa dela erantzun du [Abraham Olanok]$_i$. (...) "Gustura egoteko moduko lana egin dut, baina laugarrengo postuak beti uzten dizu tristura bat, podiumaren atarian gelditzen baitzara. Baina, bai, [nik]$_i$ uste dut nahikoa saio ona egin dudala".

'When asking him about the silver medal of the Olympic Games in Atlanta [Abraham Olano]$_i$ says that it will be practically impossible to repeat it (...) "I am satisfied with the work, but I feel personal sadness with the fourth position because I have almost reached the podium. But, yes, [I]$_i$ think my race was quite good.'

In the section before, we mentioned that possessives have been annotated as pronominals even if they are embedded in a noun-phrase and work as determiners:

(29) [Carlos Prats]$_i$ Txileko Armadako burua eta presidenteordea izan zen Salvador Allenderen Gobernua indarrean zegoenean. 1974an DINAk Buenos Airesen hil zuen atentatuan, [bere]$_i$ emaztearekin batera.

'[Carlos Prats]$_i$ was the head and vice president of the Navy of Chile during the government of Salvador Allende. He was killed by DINA in 1974, together with [his]$_i$ wife.'

16

Possessives pronouns have been annotated as pronominals too:

(30) Aretoa obretan dagoela eta, Bokaleko aretoan iraganen dira Scene Nationale erakundearen partaidetzarekin muntatutako hitzordu nagusiak. [Orkestrak]$_i$, adibidez, martxoa zehar eginen ditu [bereak]$_i$ han.

'Due to the building works, the most important appointments agreed with the Scene Nationale will be at the Bokal hall. [The Orchestra]$_i$ for example, will have there two of [his (appointments)]$_i$.'

3.3.2. Nominals

We speak about nominals when coreference is realized by a noun-phrase. The noun-phrases can have a variety of features and the coreferential relation will be established depending on the antecedent the annotator has chosen.

- Repetition

The noun-phrase is a repetition of the previous mention. Even if the declension mark is not identical we annotate it as identical since it does not give any new semantic information.

We make a distinction between exact repetition (31), when exactly the same words are used and partial repetition (32), when the head of the noun phrases is repeated but the NP includes more -or less- information than the antecendent.

(31) Taldeetako zuzendariak eta [txirrindulariak]$_i$ erabaki horren aurka agertu ziren, batik bat erabaki hori babesik gabeko gune batean hartu zelako, eta [txirrindulariek]$_i$ hotz handia pasatu behar izan zutelako.

'The managers and [the cyclists]$_i$ did not agree with this decision, especially because they took the decision in an area without protection, and because it was very cold for [the cyclists]$_i$.'

(32) Clerc nagusi [esprintean]$_i$. (...) Eztabaidak alde batera utzita, etapa garaipena [azken esprintean]$_i$ erabaki zen.

'Clerc was the first [in the sprint]$_i$ (…) Discussions aside, the triumph of the stage was decided [in the last sprint]$_i$.'

- No repetition

The nominal expressions referring to the same entity do not repeat any word of the previous one. In these cases we have to distinguish different kinds of semantic relations between the noun-phrases of the coreference chains.

Synonyms: the second part of the coreference chains is a synonyms of the first one.

(33) Peruko Gobernuak esan du OEAk ezin dituela indargabetu [hauteskundeak]$_i$ (...) Peruri gerta dakiokeen okerrena [bozak]$_i$ indargabetzea litzatekeela esan zuen De Trazegniesek.

'The Peruvian government said that OAS could not revoke [the elections]$_i$ (...) The worst thing that could happen is the abolition of [the polls]$_i$ said De Trazegnies.'

Hypernyms: the semantic relation between the two noun-phrases is realized by a hypernym. The second noun-phrase refers to the previous one with a broad meaning noun-phrase, but the reference is clearly to the same referent.

(34) Errepide ugari moztu zituzten [manifestariek]$_i$, eta denda eta lantegi asko itxi (...) [Herritarrei]$_i$ laguntza eskaini zien agintariengandik kolpatuak izanez gero.

'The roads have been cut off by [the demonstrators]i, and many shops and have been closed (...)[The citizens] i have been offered help if they were injured by the authorities.'

Generic hypernyms: the head of the second noun phrase is actualized by a generic noun like *thing*, *bussiness*, *event, issue, subject*...

(35) Peruko Gobernuak esan du OEAk ezin dituela indargabetu [Peruko hauteskundeak]$_i$. (...) [Herrialde barruko gaia]$_i$ zela azpimarratu zuen De Trazegniesek.

'The Peruvian government said that OAS could not revoke [the Peruvian elections]i (...) It is [an issue] to discuss inside the country emphasized De Trazegnies.'

Hyponyms: The semantic relation between the two noun phrases is established by means of an hyponym.

(36) Txileko Gorte Gorenak baimena eman dio Juan Guzman epaileari [Pinocheten erregimenak]$_i$ ustez Argentinan egindako hilketak ikertzeko (...) 1991n, ordea, Retting komisioa eratu zen, [Pinocheten diktadurapean]$_i$.

'The Supreme Court of Chile has given permission to the judge Juan Guzman to investigate the crimes allegedly committed in Argentina by the [Pinochet regime]$_i$ (…) But in 1991, Retting commission was founded under [the Pinochet dictatorship]$_i$.'

Inverse instance: The previous element is a noun phrase and the second element adds a characteristic of the first one.

(37) Baina ostegunean [Meles Zenawik]$_i$ adierazi zuen amaitzear zegoela orain bi urte] hasi zen gatazka. Suetena eztabaidatzeko prest dago Addis Abebako Gobernua, [Etiopiako lehen ministroak]$_i$ esan zuenez.

'But on Thursday [Meles Zenawi]$_i$ communicated the conflict that began two years before was ending. The government of Addis Abeba is ready to discuss about a cease-fire, said the [Prime Minister of Ethiopia]$_i$.'

3.3.3. Proper nouns

When the second element of the coreference chain is a proper noun, we defined four relations:

- *Repetition*

The proper noun is repeated, but it can take a different case marking suffix.

(38) Etxera itzultzean tirokatu zuten [Zugic]$_i$, asteazken gaueko 23:15ean. [Zugicek]$_i$ autoa aparkatu ondoren...

'When coming back home [Zugic]$_i$ was shot dead, on Wednesday at 23:15. [Zugic]$_i$ after parking the car...'

- *Partial repetition*

A part of the proper noun is repeated. The second element can be shorter as in example (39) or can contain more information about the proper noun.

(39) [Milo Djukanovicek]$_i$ lehendakaritzarako hauteskundeak irabazi zituenean (...) Kosovoko krisian eta gerran, [Djukanovicek]$_i$ babesa eman zien Kosovotik ihes egin zuten albaniarrei.

'When [Milo Djukanovic]$_i$ won the presidential elections (...) During the crisis and war of Kosovo [Djukanovic]$_i$ gave protection to the Albanians who escaped from Kosovo.'

- Instances

When the previous noun phrase describes a feature of the proper noun.

(40) [Lehendakari]$_i$ hautatu zutenetik, [Djukanovicek]$_i$ aldaketa handia eman dio bere ildo politikoari.

'Since he was elected for [President]$_i$, [Djukanovic]$_i$ changed his political line'

3.3.4. Ellipsis

As mentioned before, in Basque noun phrases whitout an implicit noun are frequent. We consider an elliptical relation when the second element of the chain has a hidden noun and refers to the noun of the previous noun-phrase.

(41) [Denboraldi honetan]$_i$ ordaindu du ACBra egokitu ahal izateko ordaindu beharrekoa. Ikusiko dugu [datorrenean] $_i$ zer eman diezaiokeen taldeari.

'[In this season]$_i$ he paid what is needed to pay to adapt to the ACB league. We will see [next (season)]$_i$ what he can give to his team.'

3.3.5. Place adverbs

We briefly mentioned the place adverbs in Basque in section (3.1.4). In this case we are speaking about place adverb referring to a previous noun phrase that indicates a place.

(42) Federazioak aurreprestakuntzari arreta berezia jarri dio azken urtetotan, eta horretarako [Clairefontaineko Institutua]$_i$ sortu zuen, Paris inguruan dauzkan instalazioetan. [Hara]$_i$ 12 urterekin iristen dira mutikoak, kalitate kontrola pasatu ostean'.

'In the last few years, the federation has paid particular attention to development; [the Claire Fontaine Institute]$_i$ was created for this purpose in the installations near Paris. The 12-year-old boys arrive [there]$_i$ after having a test.

3.4.   Other relation types

Contrary to what other coreference schemas (MUC, ACE) do, we do not consider nominal predicates and appositional phrases as coreferent but we take them into account and annotate them. The information of predicative and appositive relations can be very useful for coreference resolution (Recasens 2010) because they are often used to describe the same individual in the world (Borthen 2004).

3.4.1. Apposition

Two noun phrases are placed side by side and the second expression identifies or supplements the first. They usually appear separated by a comma.

(43) Zeregin horretan, [Alfredo Salazarrek]$_x$, [Baskoniako orduko idazkari teknikoak]$_x$ lan handia egin zuen.

'In this task, [the technical secretary of Baskonia]$_x$, [Alfredo Salazar]$_x$, worked hard.'

3.4.2. Predication

Predicative structures are very frequent in texts. This type of structures are often used to describe the features of the subject.

(44) [Bulatovic]$_x$ [Jugoslaviako Gobernu Federaleko lehen ministroa]$_x$ izan zen.

'[Bulatovic]$_x$ was [the Prime Minister of the Federal Republic of Yugoslavia]$_x$.'

4. Automatic preprocessing of the corpus

The coreferential annotation begins with an annotated corpus, the EPEC Corpus (Reference Corpus for the Processing of Basque) (Aduriz et al. 2006a), that provides us with a more manageable environment in which to work, and allows us to focus on the specific structures that can be part of a reference chain.

First of all, the corpus has been morphosyntactically analysed by means of MORFEUS (Alegria et al. 1996). After that, two automatic taggers (rule-based and stochastic taggers) disambiguate at lemmatization level. Entities, chunks and complex postpositions are identified by means of the following tools: i) EIHERA, which identifies entities (Institution, Person and Location) (Alegria et al. 2006); ii) IXATI Chunker (Aduriz et al. 2006b), which identifies verb chains, noun phrase units, and complex postpositions. Finally, a mention detector based on a deep linguistic analysis of mentions in Basque is applied. As many authors have stated, the mention detection task is crucial to the performance of a coreference resolution system (Stoyanov et al. 2009, Zhekova & Kübler 2010). Therefore, a detailed linguistic study of mentions has been carried out to define the linguistic features of mentions (Soraluze et al. 2012).

The tagged corpus we present in this paper contains 26,000 words and the MMAX2 application (Müller & Strube 2006) was used for the tagging (adapted to the established

requirements in the annotation). The coreference tagging process was carried out manually.

5. Disagreement cases

As mentioned before, a part of the corpus (10 %) was tagged by two annotators. We outline briefly some of the main disagreement points. We observed that the most of the disagreeements came from different semantic interpretations of the text.

The next example shows us how the same proper noun can be interpreted twofold: 1) as a sport team or 2) as the city:

(45) [Granollers taldearekin]$_i$ jokatuko du hurrengo denboraldian. Gentzelek atzo goizean hitz egin zuen Portlandeko entrenatzaile Zupo Ekisoainekin, [Granollersek $_i$ egindako eskaintza onartzeko arrazoiak azaltzeko. "[Granollersen]$_?$ lagunak ditu bere emazteak, eta hara joatea nahiago izan dute".

'He will play with the [team Granollers]$_i$. Gentzel spoke yesterday to the Portland trainer Zupo Ekisoain, to explain the reasons for having accepted the offer from [Granollers]$_i$. "His wife has many friends in [Granollers]$_?$, and they prefer going there".'

A similar interpretation difference occurs with hypernyms; sometimes it is not so easy to establish the limits of this type of relations.

(46) Bonba bat lehertu da Londresen, Ipar Irlandako Gobernuaren lehen bilera egunean [Hammersmith zubian]$_i$, [Londres mendebaldean]$_?$ (...). Atentatua 4:30etan izan zen (Londresko ordua), [Hammersmith zubian]$_i$ (...) Atentatuak, ordea, zubia eta inguru osoa ixtera behartu zuen Polizia, eta trafiko arazo handiak izan ziren egun osoan [Londres mendebaldean]$_?$.

'A bomb exploded in London, in the first meeting day of the Government of Northern Ireland, [on Hammersmith bridge]$_i$, [in West London]$_?$ (…) The explosion happened at 4:30 AM (London time), [on Hammersmith bridge]$_i$. The explosion forced to Police to close the bridge and the surrounding area for all day, and it caused big traffic problems [in West London]$_?$.'

After studying the disagreements there was a second phase where the annotation guidelines have been improved in order to cover more cases and to resolve possible ambiguities. After that, our goal is to check and correct the annotation in order to find accidental annotation mistakes.

6. Concluding remarks and future work

This paper sums up the research work done in the last years and describes the new classification for the annotation of coreferential links for Basque.

Previously, we established the first steps to annotate the corpus automatically at mention level (Soraluze et al. 2012). The work was essential to facilitate us the manual annotation of coreferential links.

Apart from that, the work presented in this paper will be helpful for the development of an application for a semiautomatic tagging of coreferential relations. It will annotate the coreferential links automatically and the annotators will only need to validate or correct the proposed links.

We are aware that the corpus annotated is limited and our main objective is to annotate a bigger corpus, with the aim of creating a valid gold standard corpus for an end to end coreference resolution system.

Bibliografia

Aduriz, Itziar, Klara Ceberio & Arantza Díaz de Ilarraza. 2005. Euskarazko anafora pronominala: ikuspuntu konputazionala eta corpus baten garapena. In *GOGOA* (Vol. 5-1- 2005). 91-116.

Aduriz, Itziar, Maria Jesus Aranzabe, Jose Mari Arriola, Aitziber Atutxa, Arantza Díaz de Ilarraza, Nerea Ezeiza, Koldo Gojenola, Maite Oronoz, Aitor Soroa & Ruben Urizar. 2006a. Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for the automatic processing". In Wilson A., Rayson P. Archer D. (eds.), *Corpus Linguistics Around the World. Book series: Language and Computers*. (Vol. 56) 1- 15. Netherlands: Rodopi.

Aduriz, Itziar, Maria Jesus Aranzabe, Jose Mari Arriola & Arantza Díaz de Ilarraza. 2006b. Sintaxi Partziala**.** In Beatriz Fernández & Itziar Laka (eds.), *Andolin gogoan: Essays in Honour of Professor Eguzkitza*. 31-49 or. Bilbo: UPV/EHU Publishing Services.

Aduriz, Itziar, Klara Ceberio & Arantza Díaz de Ilarraza. 2007. Pronominal Anaphora in Basque: Annotation issues for later computational treatment". In *6th Discourse Anaphora and Anaphor Resolution Colloquium. DAARC2007,* Lagos, Portugal.

Alegria, Iñaki, Xabier Artola, Kepa Sarasola & Miriam Urkia. 1996. Automatic morphological analysis of Basque. In *Literary & Linguistic Computing*, 11(4). 193–203.

Alegria, Iñaki, Nerea Ezeiza & Izaskun Fernandez. 2006. Named Entities Translation Based on Comparable Corpora Multi-Word-Expressions in a Multilingual Context In *Workshop on EACL06,* 1-8. Trento (Italy).

Aranzabe, Maria Jesus, Jose Mari Arriola, Aitziber Atutxa, Irene Balza & Larraitz Uria. 2003. Guía para la anotación sintáctica manual de Eus3LB (corpus del euskera anotado a nivel sintáctico, semántico y pragmático). Technical Report, UPV/EHU/LSI/TR-13.

Arriola, Jose Mari, Itziar Aduriz, Izaskun Aldezabal, Maria Jesus Aranzabe, Klara Ceberio, Ainara Estarrona, Mikel Iruskieta, Mikel Lersundi, Elisabete Pociello, Larraitz Uria & Ruben Urizar. 2013. Reusing the CG-2 Grammar for Processing Basque Complex Postpositions. In Alberto Diaz, Iñaki Alegria & Julio Villena (eds.), In *Actas del XXIX Congreso de la Sociedad Española del Procesamiento del Lenguaje Natural (SEPLN 2013)*, 20-27, Madrid (España).

Borthen, Kaja. 2004. Predicative NPs and the annotation of reference chains. In *Proceedings of Coling 2004*, 1175–1178. Geneva, Switzerland.

Botley, Simon & Tony McEnery (eds.). 2000. *Corpus-based and Computational Approaches to Discourse Anaphora*. Amsterdam: John Benjamins.

Ceberio, Klara, Itziar Aduriz, Arantza Díaz de Ilarraza & Ines Maria Garcia-Azkoaga. 2008. Erreferentziakidetasunaren azterketa eta anotazioa euskarazko corpus batean. In Xabier Artiagoitia & Joseba Andoni Lakarra (eds.), *Gramatika Jaietan. Patxi Goenagaren omenez*, (*ASJU*, 51), 153-172. Bilbo: UPV/EHU & Gipuzkoako Foru Aldundia.

Cornish, Francis. 1999. *Anaphora, Discourse and Understanding: Evidence from English and French*. Oxford: Clarendon.

Doddington, George, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel & Ralph Weischedel. 2004. The Automatic Content Extraction (ACE) program – Tasks, data, and evaluation. In *Proceedings of LREC 2004*, 837–840, Lisbon.

Euskaltzaindia. 1985. *Euskal Gramatika: Lehen urratsak-I*. Bilbo: Euskaltzaindia.

Euskaltzaindia. 2002. *Euskal Gramatika Laburra*: *Perpaus Bakuna*. Bilbo: Euskaltzaindia. (bigarren edizioa).

Garcia-Azkoaga, Ines Maria. 2003. *Kohesio anaforikoa hiru testu generotan. Adinaren araberako azterketa.* Bilbo: EHU-UPV Publishing Services.

Garcia-Azkoaga, Ines Maria. (in press). Erreferentziakidetasuna eta anafora: Diziplinartekotasunerako zedarriztatze erabilgarri baterantz.

Goenaga, Iakes, Olatz Arregi, Klara Ceberio, Arantza Díaz de Ilarraza & Amane Jimeno. 2012. Automatic Coreference Annotation in Basque". In *TLT11 Proceedings of the Eleventh International Workshop on Treebanks and Linguistic Theories*. Lisboa: Edições Colibri.

Hirschman, Lynette & Nancy Chinchor. 1997. MUC-7 coreference task definition. In *MUC-7 Proceedings,* Science Applications International Corporation.

Hualde, José Ignacio & Jon Ortiz de Urbina (eds.). 2003. *A Grammar of Basque*. Berlin/New York: Mouton de Gruyter.

Kleiber, Georges. 1994. *Anaphores et pronoms*. Louvain-la-Neuve: Duculot.

Laka, Itziar. 1996. *A Brief Grammar of Euskara, the Basque Language*. EHU/UPV Publishing Services: Leioa (Spain). http://www.ehu.eus/eu/web/eins/a-brief-grammar-of-euskara (accesed March 2015).

McCarthy, Joseph F. & Wendy G. Lehnert. 1995. Using decision trees for conference resolution. In *Proceedings of the 14th international joint conference on Artificial intelligence* – (Volume 2, IJCAI 1995), 1050–1055, San Francisco, CA, USA.

Milner, Jean-Claude. 1982. *Ordres et raisons de la langue*. Paris: Seuil.

Mitkov, Ruslan. 2002. *Anaphora resolution.* London: Longman.

Moirand, Sophie. 1990. *Une grammaire des textes et des dialogues*. Paris: Hachette.

Nicolov, Nicolas, Franco Salvetti & Steliana Ivanova. 2008. Sentiment Analysis: Does Coreference Matter? In *AISB 2008 Convention Communication, Interaction and Social Intelligence*, 37–40.

Nilsson Björkenstam, Kristina. 2013. SUC-CORE: A Balanced Corpus Annotated with Noun Phrase Coreference. In *Northern European Journal of Language Technology (NEJLT)* (3), 19-39.

Ortiz de Urbina, Jon. 1989. *Parameters in the Grammar of Basque: A GB approach to Basque Syntax*. Dordrecht: Foris.

Peral, Jesús, Manuel Palomar & Antonio Ferrández. 1999. Coreference-oriented interlingual slot structure & machine translation. In *Proceedings of the Workshop on Coreference and its Applications*, (CorefApp 1999), 69–76, Stroudsburg, PA, USA.

Pradhan, Sameer S., Lance Ramshaw, Ralph Weischedel, Jessica MacBride & Linnea Micciulla. 2007. Unrestricted coreference: Identifying entities and events in OntoNotes. In *Proceedings of ICSC 2007*, 446–453, Irvine, California.

Recasens, Marta. 2010. *Coreference: Theory, Annotation, Resolution and Evaluation*. PhD thesis, University of Barcelona, Spain.

Rodriguez Kepa. 2010. *Resources for linguistically motivated multilingual anaphora resolution*. PhD thesis, University of Trento, Italy.

Saeed, John I. 2008. *Semantics* (3rd ed.). New York: Wiley-Blackwell.

Stede, Manfred. 2011. *Discourse Processing*. San Rafael, California: Morgan & Claypool Publishers.

Steinberger, Joseph, Massimo Poesio, Mijail A. Kabadjov & Karel Jeek. 2007. Two uses of anaphora resolution in summarization. In *Information Processing and Management*, 43(6), 1663–1680.

Soraluze, Ander, Olatz Arregi, Xabier Arregi, Klara Ceberio & Arantza Díaz de Ilarraza. 2012. Mention Detection: First Steps in the Development of a Basque Coreference Resolution System. In *Proceedings of KONVENS 2012*. Vienna. 128-136.

Stoyanov, Veselin, Natan Gilbert, Claire Cardie & Ellen Riloff . 2009. Conundrums in Noun Phrase Coreferenc Resolution: Making Sense of the State of-the-Art. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 656–664, Suntec, Singapore.

Thatcher, Nicole. 1999. La coréference lexicale. In *Journal of French Language Studies*, (9/1), 91-104.

Vicedo, José Luis & Antonio Ferrández. 2006. Coreference In Q&A. In *Advances in Open Domain Question Answering*, Volume 32 of Text, Speech and Language Technology, 71–96. Berlin/New York: Springer.

Zabala, Igone & Juan-Carlos Odriozola. 2004. Los complejos posposicionales en vasco. In Perez Gaztelu Elixabete, Igone Zabala I. & Gràcia Sole L. (ed.), *Las Fronteras de la Composición*, 281-315. Donostia (Spain): University of Deusto.

Zabala, Igone. 1996. Testu-lotura: lotura tematikoa eta erreferentzia-sareak testu teknikoetan. In Igone Zabala (ed.), *Testu-loturarako baliabideak: Euskara Teknikoa*, 15-44. Bilbo: EHU-UPV Publishing Services.

Zhekova, Desislava & Sandra Kübler. 2010. UBIU: A language-independent system for coreference resolution. In *Proceedings of the 5th International Workshop on Semantic Evaluation* (*SemEval 2010*), 96–99, Stroudsburg, PA, USA.