

Evaluation of the syntactic annotation in EPEC, the Reference Corpus for the Processing of Basque

Larraitz Uri, Aina Estarrona, Izaskun Aldezabal, Maria Jesús Aranzabe,
Arantza Díaz de Ilarraza, Mikel Iruskieta

IXA Group (Natural Language Processing)
University of the Basque Country
Computer Science Faculty
Manuel Lardizabal Pasealekua 1
E-20018 Donostia

{larraitz.uria, ainara.estarrona, izaskun.aldezabal, maxux.aranzabe, a.diazdeillaraza,
mikel.iruskieta}@ehu.es

Abstract. The aim of this work is to evaluate the dependency-based annotation of EPEC (the Reference Corpus for the Processing of Basque) by means of an experiment: two annotators have syntactically tagged a sample of the mentioned corpus in order to evaluate the agreement-rate between them and to identify those issues that have to be improved in the syntactic annotation process. In this article we present the quantitative and qualitative results of this evaluation.

Key Words: Basque corpus, dependency-based syntactic annotation, evaluation, annotators' agreement-rate, Kappa agreement index.

1. Introduction

This work has been carried out in the framework of the Ixa research group¹, where resources such as data-bases and corpora annotated at different linguistic levels are being developed.

The EPEC corpus [1], considered in the Ixa group a reference corpus for the processing of Basque, is so far annotated at syntactic level, with dependencies' relations; and a part of the semantic annotation (the nominal part) is also finished.

Every annotation process has to be evaluated in order to warranty its quality. In this paper, we present the qualitative and quantitative evaluation of the dependency-based annotation of a sample of EPEC. The aim of this evaluation is twofold: to measure the agreement-rate between the annotators and to identify those issues that have to be improved in the syntactic annotation process.

The paper is organized as follows: in section 2 we explain some features of the EPEC corpus. Section 3 deals with the model adopted for the syntactic analysis and

¹ <http://ixa.si.ehu.es/Ixa>

annotation. In section 4 we present the evaluation carried out: first, the quantitative evaluation, indicating the data obtained from the Kappa agreement index based on [2] is explained, and secondly, the qualitative evaluation, which is illustrated with some representative examples. Finally, some conclusions and future works are outlined in section 5.

2. The EPEC corpus

The EPEC Corpus is a 300,000 words collection of written standard Basque. It is aimed to be a "reference" corpus for the development and improvement of several NLP (Natural Language Processing) tools we are developing for Basque [3].

The corpus has been linguistically annotated at different levels: it was first morphologically analyzed by means of MORFEUS [4] and then manually disambiguated [5]. In the manual tagging, each word-form of the whole corpus was assigned its corresponding analysis at the segmentation level: part-of-speech, number, definiteness and declension case. After the morphological disambiguation, other modules within the chunker IXATI [6], [7] such as complex postpositions, name-entities, multiword lexical units and morphosyntax were applied. The manual dependency-based syntactic annotation started precisely at this stage. This way, we have nowadays a Treebank for Basque of 300,000 words completely and correctly analyzed at dependency level [8], [9]. The semantic annotation has been so far carried out at the nominal part [10], based on Euskal Wordnet [11].

Although this is the process followed when annotating manually the dependency relations, we have also developed grammars and tools for automatic disambiguation [12], including the disambiguation of syntactic functions. For this purpose, we have made use of the Constraint Grammar (CG) formalism [13], [14], and stochastic methods have been also applied [15]. At present, the analyzers and disambiguation tools for the dependency-based syntactic annotation are being developed [16], [9]. In all cases, the correct data (the manually disambiguated data) is used both to validate the grammars and disambiguation tools as well as to apply methods of machine learning [15].

3. Syntactic annotation

Syntactic annotation means adding syntactic information to a text using special markers which provide information about the syntactic structures of sentences; e.g. labelled bracketings or symbols indicating dependency relations between words.

Annotation schemes usually differ in the labels used and in some cases the nodes composing the trees have different functions. However, most schemes provide a similar constituency-based representation of relations among the syntactic components (see [17]). In contrast, dependency schemes (e.g., [18]; [19]; [20]) do not provide a constituent-based analysis but rather specify explicitly the grammatical relations among the components of a sentence.

The debate whether a constituency-based or a dependency-based formalism should be used when developing a Treebank is still open. In fact, some researchers have adopted a middle-ground position, as in [21], where they use the dependency-

based approach only to combine the basic components of the sentence (noun phrases, prepositional phrases and the verb).

The above described formalisms may be suitable in general. However, the success and influence they may exert on applications highly depend on the language under consideration. After considering a number of trials presented in [22], [23], and [24], we have decided to follow the dependency-based procedure to deal with the free word-order structure displayed by the Basque syntax. The dependency-based formalism describes the relations between components (i.e. word-forms). This way, for each sentence in the corpus we explicitly determine the syntactic dependencies between the head and its dependants. This is the formalism used in the Prague Dependency TreeBank (PDT) [25], which is considered as the first consistently annotated Treebank based on dependencies.

The dependency scheme we have adopted is based on [26] and we defined the hierarchy relations shown in figure 1, following that scheme. This hierarchy consists of several general levels, which are further specified in subsequent levels. Structurally case-marked complements, thematic roles (*arg_mod*)², modifiers, auxiliaries and conjunctions belong to the general level. In addition, structurally case-marked complements are divided into noun phrases and clauses. Each continuous gradation achieves further specification by taking into account their grammatical function (e.g. *ncsubj*, *ncobj*, and *nczobj*). Below we present the representation of the grammatical relations regarding the mentioned dependency-tags, which are structurally case-marked non-clausal (nc) complements:

ncsubj (Case, Head, Head of NC, Case-marked element within NC, subj)

ncobj (Case, Head, Head of NC, Case-marked element within NC, obj)

nczobj³ (Case, Head, Head of NC, Case-marked element within NC, ind.obj)

² Although this field is previewed, it is not filled yet. We have planned to complete this task in future steps, when treating semantics.

³ *nczobj* would be equivalent to the English *nciobj* (non-clausal indirect object).

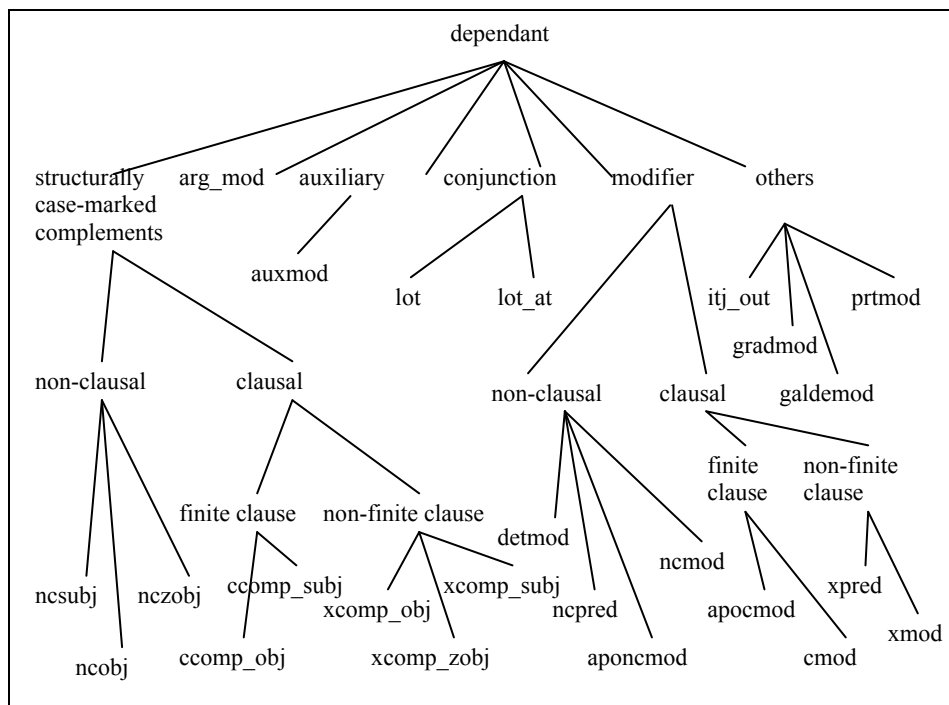


Figure 1: Hierarchy of grammatical relations.

For example, the sentence “Aitak haurrari sagarra eman dio” (‘Father has given an apple to the child’) is annotated using the three mentioned tags for the *aitak* (‘father’), *haurrari* (‘to the child’) and *sagarra* (‘an apple’) dependants, typed in italic:

ncsubj (erg, eman, *aitak*, aitak, subj)
 nczobj (dat, eman, *haurrari*, haurrari, ind.obj)
 ncobj (abs, eman, *sagarra*, sagarra, obj)

Dependency relations can also be represented by a tree structure, as in Figure 2. Head is shown as node at the upper end of branches and dependants are shown at the lower end of branches. Thus, "eman" is analyzed as the head of "aitak", "haurrari", "sagarra" and "dio".

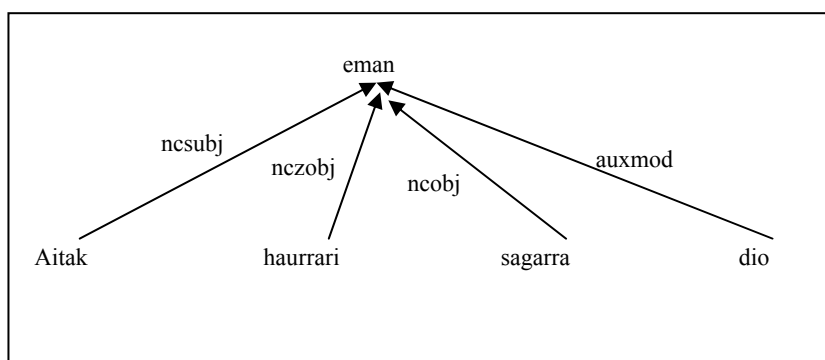


Figure 2: The tree structure of the sentence “Aitak haurrari sagarra eman dio”.

The main features of the syntactic annotation are the following ones:

- Only explicit elements are annotated; that is, neither dropped elements (such as pro, PRO or other types of ellipsis), control structures nor co-references are marked.
- The order in the annotation is not relevant. The dependency-based formalism has been chosen actually to allow free word-order representation, which is appropriate for free order languages such as Basque as well as to represent discontinuous multiword expressions⁴.
- The adopted formalism does not belong to any concrete theory; it is thought to be a neutral formalism.

4. Evaluation

In this section we explain the methodology used for the evaluation as well as the two types of evaluation we have carried out.

4.1. Methodology

In order to evaluate the dependency-based annotation, 50 sentences including the most common verb have been selected at random and annotated by two taggers. This verb is “izan” (‘to be’), which usually appears next to other verbs which are also analyzed.

The purpose of the evaluation has been twofold. On the one hand, we wanted quantitative results: some statistics concerning the agreement-rate between the two annotators. On the other hand, we wanted to analyze in which cases the annotators disagreed in the tagging process to identify both the phenomena to be improved for future annotations and the cases which are intrinsic to the language’s ambiguity and complexity.

⁴ It does not mean that the order of the words in the sentence is not relevant from other point of view such as semantics, but from a pure phrasal and functional point of view (and taking into account that in Basque the functions are not changed depending on the order of the phrases) the dependency-based formalism does not require annotators to maintain the exact order, and then they can tag the word in the preferred order.

For the quantitative evaluation, we have applied the Kappa agreement index based on [2]. In addition, we present the tags used by the annotators with their absolute number (see table 2). Taking into account that in the annotation process some sentences have been excluded for several reasons (section 4.2), we have distinguished two disagreement types: disagreements in the annotation and disagreements when excluding sentences. The Kappa index is applied in the annotated data.

For the qualitative evaluation, we have taken into account the linguistic phenomena of each example in which the annotators disagreed as well as the possible reasons which caused those disagreements.

4.2. Quantitative evaluation

The annotators can exclude sentences which are syntactically incorrect or extremely long. Consequently, it is necessary to distinguish two kinds of disagreements: disagreement when annotating the same sentence and disagreement when one annotator annotates a sentence and the other one excludes it. The Kappa index is then applied only in the first case: the comparison is made in sentences that both annotators have annotated. For this reason, the statistics we present in this paper will be divided into these two cases.

The general percentages of the two disagreement types are presented in Table 1.

Table 1. General percentages of the two disagreement types.

	No. of disagr.	%
Disagreement when annotating	30	88,23
Disagreement when excluding sentences	4	11,76

4.2.1 Disagreements when tagging a sentence

Before explaining the Kappa results, in table 2 we show the results of the annotation, specifying the number of labels used by each annotator and the match between them.

Table 2. Matching between the annotators (excluded sentences are not included).

	Labels	Annot.1	Annot.2	Agr.	Disagr.	%
1	ncmod	125	118	100	43	27,92
2	ncsubj	43	45	36	16	10,38
3	lot	40	41	27	27	17,53
4	ncpred	35	34	23	23	14,93
5	detmod	20	21	19	3	1,94
6	auxmod	18	19	18	1	0,64
7	entios	9	15	8	8	5,19
8	ncobj	10	10	8	4	2,59
9	cmod	9	9	7	4	2,59
10	lotat	6	6	5	2	1,29
11	postos	3	2	2	1	0,64
12	xcomp_obj	2	3	2	1	0,64
13	xmod	4	3	2	3	1,94
14	aponcmod	6	5	1	9	5,84
15	ccomp_obj	2	3	1	3	1,94
16	gradmod	1	1	1	0	0
17	menos	1	1	1	0	0
18	nczobj	1	1	1	0	0
19	haos	1	4	0	5	3,24
20	ccomp_subj	1	0	0	1	0,64
	TOTAL	337	341	262	154	99,99

In total there have been 30 discrepancies between annotators, that is, different options when annotating the same word (see section 4.3 Qualitative evaluation, for further explanations), and they have caused 154 disagreements. Annotator 1 has used 75 labels; annotator 2 has used 79. Sometimes the disagreement is not in the label but inside the label (see example 4 in section 4.3.1). One option can often imply disagreements in more than one label (example 3 in section 4.3.1). That is way 30 discrepancies of the annotators carry out 154 disagreements in labels.

We have based on Landis & Koch [2] to get the Kappa measures (which is a more robust method than the simple percentage of agreements), and Cohen [27] for the coefficients of the agreement-rate (table 3).

Table 3. Coefficients for the agreement-rate based on [27].

Kappa Statistic	Strength of agreement
<0.00	Poor
0.0-0.20	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost perfect

The results obtained from the Kappa inter-annotator agreement-rate are shown in table 4. Two Kappa measures are provided: one belongs to annotator 1, taking annotator 2 as gold standard, and vice versa.

Table 4. Kappa results.

	Used tags	Agreement	Percentage of the agreement Pr(a)	Kappa	Possible agr. at random Pr(e)
Annot. 1	337	262	0.7774	0.73	0.1825
Annot. 2	341	262	0.7683	0.72	0.1782

The agreement-rate between the two annotators (0.7683, 0.7774) is in fact considerable; there is almost one point difference between them. This is because annotator 1 has used 337 tags while annotator 2 has used 341. Annotators' agreement is substantially bigger than possible agreement at random (0.1782).

Although the results are acceptable, we think they could be improved if we chose other verb instead of "izan" ('to be'). In fact, ambiguity in the tags ncsbj and ncpred is probably bigger in this verb than in others. We have observed it in the qualitative evaluation (see example 13 in section 4.3.1).

4.2.2 Disagreements when excluding sentences

The agreement-rate between the two annotators when tagging and excluding sentences is 92 %. Therefore, in 8 % of the cases there was disagreement (table 5).

Table 5. Agreement-rate between the two annotators when tagging or excluding a sentence.

	Annot. 1	Annot. 2	Agr.	Disagr.	TOTAL
Analysed clauses	43	43	41	4	86
Excluded clauses	7	7	5	4	14

4.3. Qualitative evaluation

After showing the results of the quantitative evaluation, in this section we present in which cases and why annotators did not agree. We first focus on the linguistic phenomena in which the annotators disagreed, and then on the possible reasons for disagreements.

The linguistic phenomena in which the annotators disagreed are:

- 1- **Different head**: the annotators selected a different head.
- 2- **Different label**: the annotators selected a different dependency-tag.
- 3- **Different case**: the annotators selected a different case in the dependency-tag.
- 4- **Incomplete tree**: the annotator forgot a tag and the tree is not complete.
- 5- **Excluded sentence**: the annotator excluded the sentence.

As concerns the reasons for disagreements, we have distinguished three types: ambiguity, erroneous annotation and gaps; the second and third types consist of some subtypes.

1- Ambiguity: The annotators analyze the sentence in different ways but both analyses are possible and correct.

2- Erroneous annotation:

- The annotator does not realize s/he has made an error: the annotator does not doubt the analysis of a word when it is in fact wrong and not even mentioned in the manual.
- The annotator does not follow the manual.
- The annotator does not use auxiliary labels when needed. The correct treatment of multiword expressions require some auxiliary tags to join the elements that form the expression when the provided syntactic analysis has not treated multiword expressions as such. The annotator may not use those auxiliary tags to join the multiword expressions.

3- Gaps:

- Gaps in the manual: the manual does not cover all the possible existing linguistic phenomena.
- Gaps in the previous modules, such as the analysis of complex postpositions, multiword lexical units, name-entities or morphosyntax.

In the following tables, we show the data concerning the disagreements between the two annotators when tagging a sentence (table 6) and when excluding a sentence (table 7).

Table 6. Disagreements when annotating sentences.

	Anbig.	Errors			Gaps		Total
		Annotation	Following manual	Using Auxiliary labels	Manual	Other modules	
Head	3	1	1	0	4	0	9
Label	0	2	2	6	4	0	14
Case	0	1	2	1	1	1	6
Tree	0	1	0	0	0	0	1
Total	3	5	5	7	9	1	30
Total	3	17			10		30
%	10 %	56.66 %			33.3 %		100 %

Table 7. Disagreements when excluding sentences.

	Annotator's error	Manual gap
Sentence limit	1	0
Length	0	1
Grammatical error	1	0
Total	2	1
Total	3	
%	9 %	

The most frequent disagreement is due to annotators' errors and the most frequent disagreement subtype lies in the erroneous annotation group (56.66 %). Gaps in general represent the 3.3 %. Most of the gaps are caused because the annotation manual is still being built. 10 % of the cases occurred because of language's ambiguity. Fewer disagreements come from specific errors.

4.3.1 Some representative examples

In this section we show some examples which represent the most common types of disagreements and the main reasons for them. However, in [8], more of them are deeply explained.

--- Different head

- Neither all the sentence connectors nor conjunctions are listed in the manual, but only general considerations are remarked and illustrated with some examples. Therefore, one of the annotators has sometimes considered an element a conjunction and the other one a sentence connector or an adverb. Consequently, the head also results different. That is the case of the elements "bestek beste" ('between others') and "ondorioz" ('in consequence') in the examples bellow.

[1] Granadan, Jaime Mayor Oreja Barne ministroa, Juan Cotino Espainiako Poliziako zuzendaria, PPko Teofila Martinez eta Jose Moratalla Granadako alkatea izan ziren, **besteak beste**.

In Granada, there were the Home Secretary Jaime Mayor Oreja, the head of the Spanish Police Juan Cotino, Teofila Martinez from PP and the mayor of Granada Jose Moratalla, **among others**.

In example [1], one annotator interpreted “besteak beste” as a coordinating conjunction and the other one as an explicative sentence connector.

[2] Madrilén arabera, laguntza horiek "orokorrak" dira, eta **ondorioz**, ez dira legez kanpoko Estatu laguntzak.

According to Madrid, those subsidies are “general” ones, and **as a consequence**, they are not illegal state subsidies.

In example [2], one annotator interpreted “ondorioz” as an adverb and the other one annotated it as a conjunction.

--- Different label

▪ Multiword expressions constitute another controversial issue in the tagging task. There is a wide range of multiwords in Basque. Besides, there are new expressions continuously being created. As a consequence, it is quite common to find units in the corpus that in previous syntactic analysis have not been treated as such, that have not been correctly detected by the automatic tool. Because the range of possibilities is high, annotators do not agree when jointing the multiword expressions. In example [3], for instance, they do not agree when delimiting the multiword entity: one annotator considers “*Justizia eta Lan sailburua*” (‘Justice and Job member’) a whole entity and the other one joins the two words “*Justizia eta Lan*” (‘Justice and Job’) with the coordinating conjunction “eta” (‘and’).

[3] Eusko Jaurlaritzaren izenean Sabin Intxaurrega **Justizia eta Lan sailburua** izan zen hiletetan.

On behalf of the Basque Government, the **Justice and Job member** Sabin Intxaurrega was in the funeral.

--- Different case

▪ Sometimes different case might be assigned because of a gap in the manual. When explaining how to treat the punctuation marks that work as conjunctions, there are no specifications regarding which kind of relations should be added. One annotator decides to annotate them always as “emen” (coordinating conjunction) and the other annotator chooses it depending on the sentence. For instance, in example [4]

annotator 1 uses the disjunctive relation “haut” and annotator 2 the coordinating conjunction “emen”.

[4] Nik uste dut aldi honetan indarkeria ez dela bakarrik Gaza, Zisjordania **edo** Jerusalem Ekialdeko palestinarren kontra.

I think that this time violence is not only against the Palestinian of Gaza, Zisjordania **or** East Jerusalem.

--- Excluded sentences

▪ In the manual the possibility of excluding sentences is previewed. One of the reasons for excluding a sentence is its length. Nevertheless, this criterion is not exactly defined and each annotator is free for excluding sentences based on his/her subjectivity. Although they agree in many cases, there are some differences. Sometimes, the sentence is long but not difficult, and the tagging results easy. In this case, one annotator decides to tag it, and the other one does not. Maybe, we should specify a number of words to take into account to exclude long sentences.

▪ Annotator's error for not following the manual. In the manual sentences' boundaries are clearly defined. However, an annotator may not take it into account and s/he may exclude a sentence for being “wrong-limited sentence” when it is actually well delimited. In example [5], the colon is not considered a boundary by one annotator (although it is defined as such in the manual) and s/he excludes it.

[5] Bi dira ezaugarri garrantzitsuenak Zenarruzabeitiaren arabera:

There are two main characteristics, according to Zenarruzabeitia:

--- Ambiguity

There are some ambiguous sentences that can have more than one interpretation. Ambiguous sentences can be, therefore, differently annotated.

In example [6], ambiguity lies in “*these last days*” since it can be “*it is obvious these last days*” or “*these last days has attacked*”.

[6] Begi bistakoa da **azken egunotan** Israelek palestinarren kontra egindako eraso oldeak esanahi argia duela guretzat .

It is obvious **these last days** Israel has harshly attacked against Palestinian and this attack has a clear meaning for us.

In example [7], the word “*albistea*” (‘new’) can be considered either the subject (*ncsubj*) or the predicate (*ncpred*) of the verb “*izan*” (‘to be’). There is not any clear criterion in the Basque grammar to disambiguate this kind of attributive relations with

the verb “*izan*” (‘to be’). Neither the manual has a concrete rule for that. Therefore, annotators may not agree when tagging this kind of sentences⁵.

[7] Nazioarteko laburretako lehen **albistea** da argazkia duen bakarra.

The first brief international **new** is the only one having a photo.

5. Conclusions

In this paper we have presented the results obtained from the evaluation of the syntactic annotation of the EPEC corpus. The results have been satisfactory: there has been a substantial agreement-rate between the two annotators, following [27] coefficients. However, we think the result would have been better if we had chosen another verb different to “*izan*” (‘to be’), since this verb shows ambiguity in the tags *ncsubj* and *ncpred*. This will be an interesting issue to analyze in the future. On the other hand, it should be necessary to evaluate a bigger set of sentence to get more reliable data and confirm or refuse either the statements or the evaluation data we present here. This is another planned task for the near future.

The evaluation process has made explicit some gaps to be improved in the annotation manual. Therefore, if we consider the disagreements caused because the phenomena was not clearly specified in the manual and we clarify those issues, the results would improve in 33 % (table 6). In fact, the development of an exhaustive annotation manual is an ongoing work: it is almost impossible to cover all linguistic phenomena a priori, although we think we have got a good base.

One objective way to improve the results would be to exclude too long sentences, sentences consisting of more than a concrete number of words (i.e. more than 40 words). Furthermore, the state of mind of the annotator is not always the same, which is also influential in the annotation task.

To get a 100 % of agreement-rate between the two annotators is in fact a utopian goal, since language is intrinsically ambiguous and open to different but correct analysis.

References

1. Aduriz I., Aranzabe M.J., Arriola J.M., Atutxa A., Díaz de Ilarraza A., Ezeiza N., Gojenola K., Oronoz M., Soroa A., Urizar R. 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. and Archer D. editors, *Corpus Linguistics Around the World*, 1-15. Rodopi (Netherland).
2. Landis, JR and Koch, GG. 1977. The measurement of Observer Agreement for Categorical Data. *Biometrics*. Vol. 33, pp. 159-174.
3. Bengoetxea K. and Gojenola K. 2007. Desarrollo de un analizador sintáctico estadístico basado en dependencias para el euskera. *Procesamiento del Lenguaje Natural*, SEPLN’07. Universidad de Sevilla.

⁵ When the annotator realizes that the sentence is in fact ambiguous, s/he chooses one analysis and then marks the sentence as ambiguous.

4. Alegria I. 1995. Euskal morfologiaren tratamendu automatikorako tresnak. Doktoretza-tesia, Euskal Herriko Unibertsitatea (UPV/EHU).
5. Aldezabal I., Ceberio K., Esparza I., Estarrona A., Etxeberria J., Iruskietta M., Izagirre E., Uria L. 2007a. EPEC (Euskararen Prozesamendurako Erreferentzia Corpora) segmentazio-mailan etiketatzeko eskuliburua, UPV/EHU / LSI / TR 11-2007.
6. Aduriz I., Díaz de Ilarraza A. 2003. Morphosyntactic disambiguation and shallow parsing in Computational Processing of Basque. In Oyharzabal B. editor, *Inquiries into the lexicon-syntax relations in Basque*. ASJUren gehigarria. Euskal Herriko Unibertsitatea (UPV/EHU), Bilbo.
7. Aduriz I., Aranzabe M.J., Arriola J.M., Díaz de Ilarraza A. 2006b. Sintaxi partziala. In Fernández B. and Laka I. editors, *Andolin gogoan: Essays in Honour of Professor Eguzkitza*, UPV/EHUko Argitarapen Zerbitzua, Bilbo.
8. Aldezabal I., Aranzabe M.J., Arriola J.M., Díaz de Ilarraza A., Estarrona A., Fernandez K., Iruskietta M., Uria L. 2007b. EPEC (Euskararen Prozesamendurako Erreferentzia Corpora) dependentzietekin etiketatzeko eskuliburua. UPV/EHU / LSI / TR 12-2007.
9. Aranzabe M.J. 2008. Dependentsia-ereduan oinarritutako baliabide sintaktikoak: zuhaitz-bankua eta gramatika konputazionala. Doktoretza-tesia. Euskal Herriko Unibertsitatea (UPV/EHU).
10. Agirre E., Aldezabal I., Estarrona A., Pociello E. 2008. A methodology for the joint development of the Basque WordNet and Semcor. Dutch SemCor workshop. Amsterdam.
11. Pociello E. 2008. Euskararen ezagutza-base lexikala: Euskal WordNet. Doktoretza-tesia, Euskal Filologia Saila (UPV/EHU). Leioa.
12. Aduriz I., Aldezabal I., Alegria I., Artola X., Ezeiza N., Urizar R. 1996. EUSLEM: A Lemmatiser / Tagger for Basque Proc. EURALEX'96, Part 1, 17-26. Göteborg (Sweden). ISBN 91-87850-14-1.
13. Karlsson F., Voutilainen A., Heikkilä J., Anttila A. 1995. *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin.
14. Tapanainen P. and Voutilainen A. 1994. Tagging Accurately – Don't guess if you know. Proceedings of the 4th Conference on Applied Natural Language Processing, ANLP'94.
15. Ezeiza N. 2003. Corpusak ustiatzeko tresna linguistikoak. Euskararen etiketazaille morfosintaktiko sendo eta malgua. Doktoretza-tesia, Euskal Herriko Unibertsitatea (UPV/EHU).
16. Aduriz I. 2000. EUSMG: morfologiatik sintaxira murriztapen gramatika erabiliz. Euskararen desanbiguzio morfologikoaren tratamendua eta azterketa sintaktikoaren lehen urratsak. Doktoretza-tesia, Euskal Herriko Unibertsitatea (UPV/EHU).
17. Abeillé A. 2003. *Treebanks: Building and Using Parsed Corpora*. Kluwer Academic Publisher, Dordrecht, The Netherlands.
18. Sleator D. and Temperley D. 1993. Parsing English with a link grammar. Third International Workshop on Parsing Technologies.
19. Järvinen T. and Tapanainen P. 1997. A Dependency Parser for English. Technical Report, n° TR-1, Department of General Linguistics. University of Helsinki.
20. Bunt H., Carroll J., Satta G. 2004. *New Developments in Parsing Technology*. Kluwer Academic Publishers. Text, speech and language technology, volume 23.
21. Montemagni S., Barsotti F., Battista M., Calzolari N., Corazzari o., Lenci A., Zampolli A., Fanciulli F., Massetani M., Raffaelli R., Basili R., Paziienza M., Saracino D., Zanzotto F., Mana N., Pianesi F., Delmonte R. 2003. Building the Italian Syntactic-Semantic Treebank. In Abeillé A. editor, *Building and Using Parsed Corpora*, 189-210. Kluwer Academic Publisher, the Netherlands.
22. Skut W., Krenn B., Brants T., Uszkoreit H. 1997. An Annotation Scheme for Free Word Order Languages. Proceedings of the Fifth Conference on Applied Natural Language Processing, 88-95. Washington, D.C., USA.

23. Järvinen T. and Tapanainen P. 1998. Towards an Implementable Dependency Grammar. Proceedings of the Workshop on Processing of Dependency-Based Grammars, COLING-ACL'98, Montreal.
24. Oflazer K., Zeynep D., Tür H., Tür G. 1999. Design for a Turkish Treebank. Proceedings of Workshop on Linguistically Interpreted Corpora. Bergen.
25. Böhomová, A.; Hajic, J.; Hajicova, E. and Hladka B. 2003. The PDT: a 3-level annotation scenario. In Abeillé, editor, Treebanks: Building and Using Parsed Corpora. Kluwer Academic Publishers, Dordrecht, The Netherlands.
26. Carroll J., Briscoe T., Sanfilippo A. 1998. Parser evaluation : a survey and a new proposal. Proceedings of the First International Conference on Language Resources and Evaluation, 447-454. Granada, Spain.
27. Cohen J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement, pages 37-46.