THEORETICAL AND METHODOLOGICAL ISSUES OF TAGGING NOUN PHRASE STRUCTURES FOLLOWING DEPENDENCY GRAMMAR FORMALISM

M. J. Aranzabe, J. M. Arriola and A. Diaz de Ilarraza
IXA Group & UPV/EHU

1. Introduction

A Treebank is a text corpus in which each sentence has been annotated with its syntactic structure. Although the construction of a treebank is an expensive task, we believe that it is indispensable for the development of real applications in the field of Natural Language Processing (NLP) and also for the development of the Information Society. At a purely linguistic level, the Treebank is an essential database for the study of a language given that it provides analyzed/annotated examples of real language. The linguistic study directly results in an improvement in the quality of several applications, such as Part-Of-Speech (POS) taggers and parsers (Collins 1997, 2000; Charniak 2000), because it provides common training and testing material allowing different algorithms to be compared and improved.

In the last few years, treebank corpora such as the Penn Treebank (Marcus et al., 1993) and the Prague Dependency Treebank (Böhmová et al. 2003) have become a crucial resource for building and evaluating natural language processing tools and applications. As Abeillé (2003) sets out, there are efforts underway for Czech, German, French, Japanese, Polish, Spanish and Turkish, to name just a few. In Kakkonen (2005) we can find the state of the art of dependency-based treebanks.

The Basque Dependency Treebank (BDT) is actually the Reference Corpus for the Processing of Basque (EPEC) annotated at syntactic level. The EPEC is a 300,000 word corpus of standard written texts which aims to be a training corpus for the development and improvement of several NLP tools. It has been manually tagged at different levels: morphology, lemmatization and surface syntax (Aduriz et al. 2006). The next level of tagging —annotation of dependency relations— is currently being carried out in BDT.

In this paper, we describe the annotation of noun phrase (henceforth, NP) constructions in detail following the Dependency Grammar theory (Tesnière 1959). For a better understanding of our work it should be noted that for us, NP is a purely descriptive term. We are not concerned with understanding the internal structure of NPs. The syntactic description of Basque NPs has been mainly developed within the generative framework by Goenaga (1980), Eguzkitza (1993), Laka (1993), Artiagoitia (2002),

[ASJU Geh 51, 2008, 57-69]
http://www.ehu.es/ojs/index.php/asju

Trask (2003), and other attempts have been made in applied linguistics (Odriozola & Zabala 1992).

Mention should be made of Goenaga (1980), as being the first work in which the Basque NP structure is analyzed in detail using generative theory. He characterizes the syntactic structure in terms of hierarchically embedded constituents or phrases: more precisely, they are all derived from the same abstract linguistic approach: phrase-structure theory. Similarly, we present in this paper our work on syntactic annotation based on the dependency model. Specifically, this marks the first formalization for the annotation of Basque NP tagging.

Phrase-structure theory and dependency theory are two different methods of conceptualizing the linguistic structure of sentences. Focusing on the second of these, we should stress that in grammars constructed within the dependency theory (e.g., Hudson 1990, Mel'cuk 1988), syntax is handled in terms of grammatical relations between pairs of individual words, such as the relation between the subject and the predicate or between a modifier and a common noun. Grammatical relations are seen as subtypes of a general, asymmetrical dependency relation: one of the words (the head) determines the syntactic and semantic features of the combination. In addition, the head also controls the characteristics and placement of the other word (the dependent). The syntactic structure of a sentence as a whole is built up from such dependency relations between individual pairs of words.

In mathematical terms (Nugues 2006), the dependency relation imposes a hierarchical structure on the words of a sentence that has the characteristics of a directed tree. A directed tree is a completely connected, two-dimensional, directed acyclic graph with a single root. Each node of the tree represents a word, and directional arches between the nodes represent the dependency relation, leading from head to dependent. The tree is headed by the highest word in the sentence, the root, which is the word that does not possess a head of its own.

We opted for annotating syntax following the dependency annotation rather than phrase-structure. We give more detailed reasons for our choice of dependency annotation in Section 2.3. The rest of Section 2 sets out the basic ideas of our annotation scheme and the annotation hierarchy. In Section 3, we describe some noun phrase constructions in detail. We propose the annotation procedure for coordination in Section 4, and we conclude with a discussion of future work in Section 5.

2. Framework for the syntactic annotation of the corpus

Syntactic annotation is the practice of adding syntactic information to a text by incorporating markers that give information on the syntactic structure of the sentences: e.g. labelled bracketing, or symbols indicating dependency relations between words. Although they differ in the labels and, in some cases, the function of various nodes in the tree, most annotation schemes provide a similar constituency-based representation of relations among syntactic components (see Abeillé 2003). In contrast, dependency schemes (e.g., Sleator & Temperley 1993, Tapanainen & Järvinen 1997, Bunt et al. 2004) do not provide a constituency analysis but rather specify grammatical relations among elements explicitly.

2.1 Constituency-based formalism

In this type of formalism, every single constituent that makes up a syntactic constituent is tagged, including the syntactic category itself; thus, the final result derives from defining the emerging constituents and their categories (noun phrases, sentences, etc.).

The most complete and most widely-used English corpus, namely the *Penn Treebank* (Marcus et al. 1993), employs this sort of tagging. The following is an illustration of how a sentence would be represented in this corpus:

(1) *John tried to open the window*[1]

```
(S (NP (N1 (N John_NP1)))
 (VP (V tried_VVD)
  (VP (V to_TO)
   (VP (V open_VVO)
    (NP (DT the_AT)
     (N1 (N window_NN1) ) ) ) ) ) )
```

This method has three outstanding properties:

1. It is based on linear word order; that is to say, the order of syntactic components reflects the order in which they appear in the sentence.
2. Hierarchical information is made explicit.
3. The information function wich is implicit, is irrelevant.

2.2. Dependency-based formalism

Unlike the constituency-based approach, dependency-based formalism (Järvinen & Tapanainen 1997) describes the relations between the components. This tagging formalism has been used for German (*NEGRA*) (Brants et al. 2003) and Czech (*PDT)* corpora,[2] among others. In this formalism, the representation for (1) above would be as follows:

The properties of this method include:

1. The relevance of word order is minimized.
2. It is a method strongly based on hierarchical relations.
3. The functional information is extremely important.

2.3. Constituency-based vs. dependency-based formalism

There is still an ongoing debate as to whether a constituency-based or a dependency-based formalism should be employed in completing the Treebank. Some

---

[1] Example taken from Carroll et al. (1998).

[2] *http://ufal.mff.cuni.cz/pcedt/doc/PCEDT_main.html*

researchers have taken the middle-ground between these two options, as in Montemagni et al. (2003), who employ the dependency-based approach only to combine the basic components of the sentence (noun phrases, prepositional phrases and the verb), without reaching the word-level for dependency purposes.

The formalisms described above may be generally suitable, but the success and influence may exert on applications highly depends to a great extent on the language under consideration.

Based on a number of tests set out in Skut et al. (1997), Tapanainen & Järvinen, (1998) and Oflazer et al. (1999), to deal with the free word-order displayed by Basque syntax, we have decided to follow the dependency-based procedure. The following issues also had a critical influence on our decision:

— Dependency-based formalism provides a way of expressing semantic relations that will constitute a good base tackling the next steps in the analysis-chain such as verb valence and thematic role studies (Agirre et al. 2006).
— We consider that the computational tools developed thus far in Natural Language Processing for Basque will serve to achieve dependency relations. Besides, the rich information involved would allow transformation from trees to other ways of representation.
— From our viewpoint, it is more straightforward to evaluate the relation between the elements that make up a sentence than the relation between elements included in parentheses, since the latter involves the additional task of determining where the parenthesis start and end.
— In our opinion, dependency-based formalism is a more accurate method for annotating empty elements, such as *pro*,[3] long–distance dependencies and discontinuous constructions.

2.4. Theoretical and methodological basis

Taking into account the literature on tagging corpora in different languages, we decided to adress certain parameters for determining the theoretical and methodological basis that are needed to build the Treebank. The basic decisions include the following:

2.4.1. Which elements will be tagged?

Our object of study is the sentence; i.e., the text enclosed between two full stops (and also some other punctuation marks such as the exclamation marks, question marks and so on). Furthermore, as well as the explicit elements making up the sentence, we have also considered certain elided elements such as the "*pro*". Empty elements, such as *pro*, long–distance dependencies and discontinuous constructions can be intuitively annotated. In theoretical terms, we could annotate the empty elements following the dependency model. We also consider multiwords, entities and com-

[3] *pro*: elided syntactic arguments that typically arise when the predicate displays agreement with the elided argument *pro* itself.

plex postpositions as analysis units. The sentences below exemplify different types of analysis units, shown in bold:

(2) *Proposamenarekin **bat egin** zuen Espilondok.*
(Espilondo joins the proposal.)
(3) ***Henriette Airek** olerki unibertsalari buruzko bere gogoetak azalduko ditu.*
(Henriette Aire will explain her thoughts about universal poetry.)
(4) *Leihoko **kartelen artetik** begiratzen du.*
(He/she looks through the posters affixed to the window.)

2.4.2. Do we follow any theory?

An annotation scheme usually has to be *theory-independent* in order to allow different interpretations of the tagged texts in different linguistic frameworks. The advantage of assuming a particular theory is that it may solve many problems. The disadvantage, however, is that theories are unable to predict many aspects contained in a corpus. In general, however, there is always a way to overcome the theory's occasional shortfalls when it comes to handling real texts. Consecuently, considering as we do that the advantage outweighs the disadvantage in absolut terms, we have to some extent follow the generative approach in certain aspects, for instance, when analysing empty categories (such as *pro*).

2.4.3. Definition of the annotation scheme employed

In order to define the tagging system we have assumed the hierarchy proposed in Carroll et al. (1998). They propose an annotation scheme in which each sentence in the corpus is marked up with a set of grammatical relations, specifying the syntactic dependency which holds between each head and its dependent(s). Following this line of work, we have developed a tag set based on hierarchies of grammatical relations (see Figure 1). In this paper we will focus on those related to NP.

The dependency grammatical relations corresponding to NP can be described from two perspectives: i) NP head non-clausal relations (explained in detail in Section 3): ncsubj, ncobj, nczobj, ncmod, ncpred and itj_out (see Table 1) and ii) the non-clausal modifiers of NP heads: detmod, ncmod, aponcmod and gradmod (see Table 2).

3. Noun Phrase structure: noun heads with their dependents

As we have already stated we are not concerned with understanding the internal structure of noun phrases. NP stands for a dependency relation headed by a noun although, as Artiagoitia (2002) points out, the definition fails fully to explain the structure of NPs.

Our approach is intended to provide consistent argument labelling that will facilitate automatic extraction of relational data, without attempting to justify any theory.

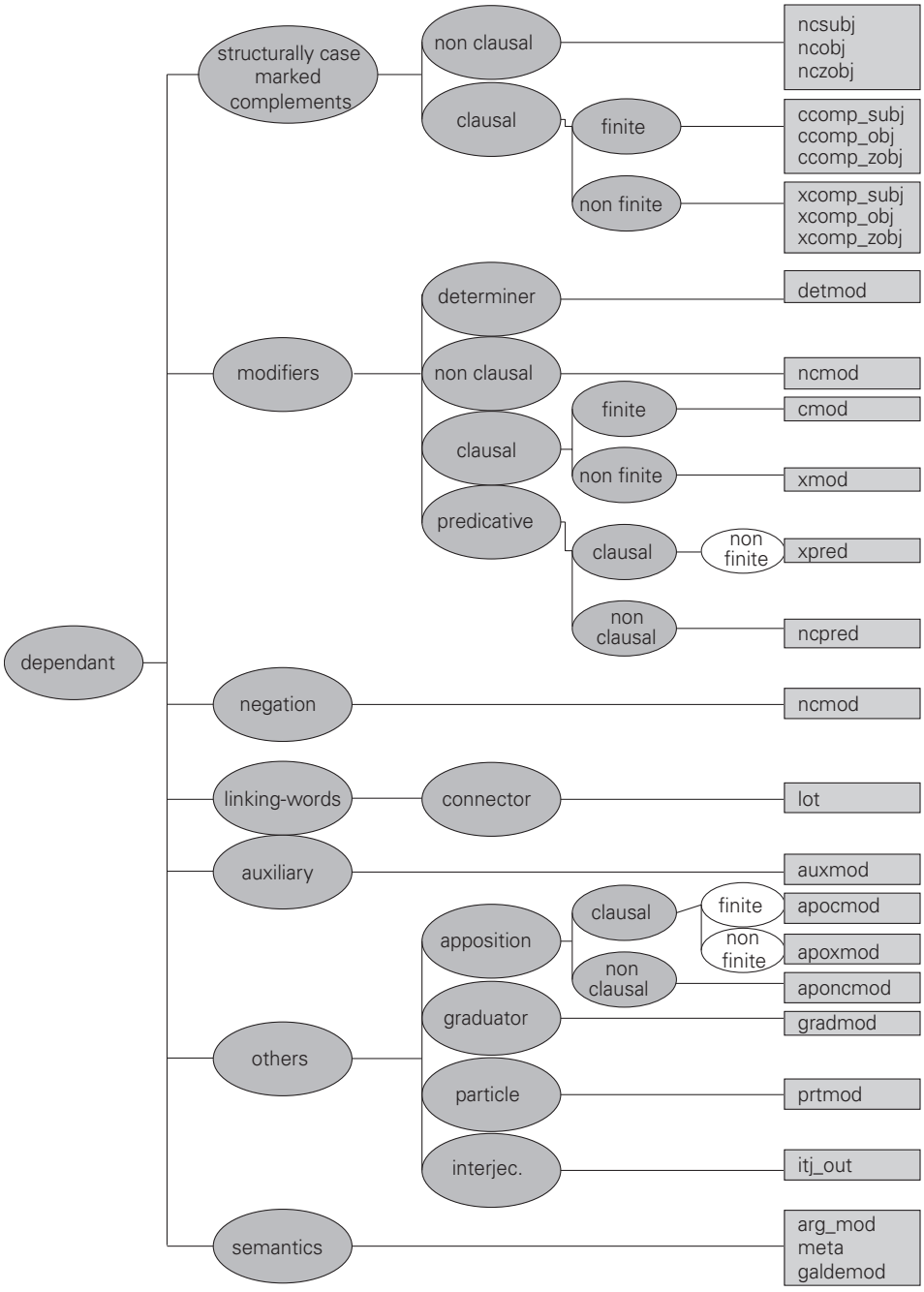

Figure 1

Hierarchy of grammatical relations

3.1. Which component of the NP will be the Head?

Basque is what is know as a 'head-final' language, since heads tend to be placed at the right hand end of phrases. If we look at the structure of phrases in Basque, we can see that this morphological marker is placed in the last component of the phrase that carries it, regardless of the POS. Thus, the case marker can be attached to the head as in (5) or to a modifier of the head (such as an adjective) as in (6) and sometimes to the determiner, as in (7)

(5) *Zenbait zalantzak ezusteko bidetik lortu zuten argia.*
(Some doubts were solved in an unexpected way.)
(6) *Edozein mutil altuk (ergative case marker) egiten du.*
(Any tall boy does it.)
(7) *Zalantza horiek ezusteko bidetik lortu zuten argia.*
(Those doubts were solved in an unexpected way.)

In order to maintain coherence in each relation when the element carrying the case/determiner and the noun head are not coincident, we decide to include both elements[4] together explicitly. We consequently use a list of tuples to represent head/modifier relations in the dependency tree. For example, a structurally case-marked complement in which the complement is nc (non-clausal) has the following format:

— Case: the case-marker by means of which the relation is established between the head and the head of the NP.
— Head: the governor of NP.
— Head dependent.
— Case-marker: the component of the NP that carries the case.
— Syntactic function: the syntactic label assigned to the relationship

The analyses of the NP included in the following sentences exemplify this formalization. In the NP "zenbait zalantzak" in Example (5), "zalantzak" is the element that carries the case marker and, at the same time, constitutes the head of NP, so, the subject relation looks like the ncsubj dependency shown below.

detmod (-, zalantzak, zenbait)
ncsubj (erg, lortu, zalantzak, zalantzak, subj)

In Example (7), the phrase "zalantza horiek", "zalantza" is the head of the NP, and then we would add the component that carries the case marker, namely "horiek". Some of the relations associated to the NP follow:

ncsubj (erg, lortu, zalantza, horiek, subj)
detmod (-, zalantza, horiek)

---

[4] The decision, however, is not specific to Basque: more generally, it arises in the word-based Constraint Grammar analyzer (Karlsson et al. 1995). Our manual tagging seems to be as compatible as possible with output obtained by the parser, for evaluation purposes. The easiest way to achieve this involved adapting the original tag-set as proposed by Carroll et al. (1998), including, in some cases, an additional slot. Note that we do not change the dependency initial philosophy; we merely accommodate it to our needs

3.2. The NPs annotated in the corpus

In this section we will use examples to explain the two different perspectives used to tag NPs: i) the relations established between the noun and the verb (ncsubj, ncobj, nczobj, ncmod, ncpred and itj_out see Table 1) and ii) the modifiers of NP heads (detmod, ncmod, aponcmod and gradmod, see Table 2).

Let us begin with the first group. Following the classification presented in Figure 1, the relations presented in the Table 1 can be grouped as typical case-marker complements (ncsubj, ncobj, nczobj), modifiers (ncmod), predicative modifier (ncpred) and others (itj_out).

In order to better understand the examples we represent the heads of the NP in bold and their governors underlined. If we look at the underlined elements we can see that these dependency relations are established with respect to the main element of the sentence. For this reason in all cases the governors are the verbs. Brackets are also used to represent phrases.

Table 1

Examples with relations headed by verbs

Examples	Dependency tag
1. *[Orduan] [**Francine**] [gizonaren begiez] arduratu zen.* (Then Francine took care of the man's eyes.) 2. *[Nekez] ahaztuko dituzte [**askok**] [egun haiek].* (Many people will not forget those days easily.)	ncsubj
3. *[Nhamdi-k] [**ukabilak**] estutu zituen.* (Nhamdi clenched his fists.)	ncobj
4. *[Astero astero] esan zaie bertaratu diren [**talde** guzti-guztiei]* (It has been said every week to all the groups that have come round.)	nczobj
5. *[**Seminariora**] zihoan [berriro].* (He was going to the seminar again.) 6. *[Zuk] galdua zenion [beldurra] [itsasoari] [**txiki-txikitatik**].* (You have lost your fear of the sea since your childhood.)	ncmod
7. *[Iritzi hau] [naturaren behaketa zuzenaren **fruitu**] zen.* (This opinion was fruit of a direct observation of nature.)	ncpred
8. *[Euriak] ez zaitu bustitzen, [**Valentine**].* (The rain is not wetting you, Valentine.)	itj-out

A characteristic of all the examples except the fourth one is that the element of the phrase linked to the verb contains the case marker. In 4, the noun "talde" is linked to the verb by means of a "nczobj" dependency relation although the case marker is included in the determiner that modifies the noun. In all those phrases when the noun is elided, the determinant (Example 2) or the adjective (6) are considered as heads. In this first approach we make no distinction between the noun predicative and verb

predicative; this is why in Example (7) the noun "fruitu" is linked to the verb rather than to the noun "iritzi". We will shortly refine this analysis.

The last example illustrates the "itj_out" relation. This relation differs from the others in so far as it does not represent a function in the sentence structure but, because it relates a noun "Valentine" and a verb "bustitzen", it has been included in this group.

Table 2 shows internal relations of NP, that is, the dependents of NP head. As we did in Table 1, we represent the heads of the NP underlined and their modifiers in bold. Some types of NP structures have been included in order to show their internal dependency relations. Examples 1 to 3 are examples of "ncmod"; all of them are linked to the noun by means of the same relation although the dependents belongs to different categories: "atmosferikoari" is an adjective (Example 1), "Arrasateko", in 2, is a noun modifier and "nekazari" is part of a compound noun. In Example 4 the demonstrative "hori" appears to the right of the noun while in 5 the ordinal "bigarren" precede the noun. Both elements are linked by the "detmod" dependency relation. In example 6 we have the apposition structure classified like others in Figure 1. It represents the relation between a noun and the head of the preceding NP. In that case it is the relation between the heads of two phrases. In the modifier relation expressed by "aponcmod" the modifier is "idazle" and the head "Axularrek".

Finally, in 7, we show an example of a "gradmod" relation wich like the others, has the idea of being a relation between a noun head and a modifier that is a graduator ("oso").

Table 2

Examples with internal NP relations

Example	Dependency tag
1. *[Nola] deitzen zaio [zirkulazio **atmosferikoari**]?* (What is atmospheric circulation called?) 2. *[**Arrasateko** zenbait familiak] [bigarren tarifa hau] kontratatu zuen.* (Some families from Arrasate hired this second rate.) 3. *[Astelehenean] [**nekazari** manifestaldi bat] izan zen.* (There was a farmers demonstration on Monday.)	ncmod
4. *[Zertara] zetorren [erretolika **hori**]?* (Why did that argument come up?) 5. *[**Bigarren** kanpamentu **hartatik**] [sarjentu] atera zen.* (From that second camp, he emerged a sergeant.)	detmod
6. *[Axularrek], [gure **idazle** handiak], idatzi zuen [liburu hori].* (Axular, our great writer, wrote that book.)	aponcmod
7. *[Azken biak] [**oso** itsusiak] ziren.* (The last two were really ugly.)	gradmod

The tables above have been written from a purely dependency relation perspective, so that the different elements that constitute the NP are grouped in terms of dependency tags. With a view to giving a general view of the structure of a sentence

following the Dependency formalism, Figure 2 shows the example 2 of Table 2 with all the dependency relations displayed.

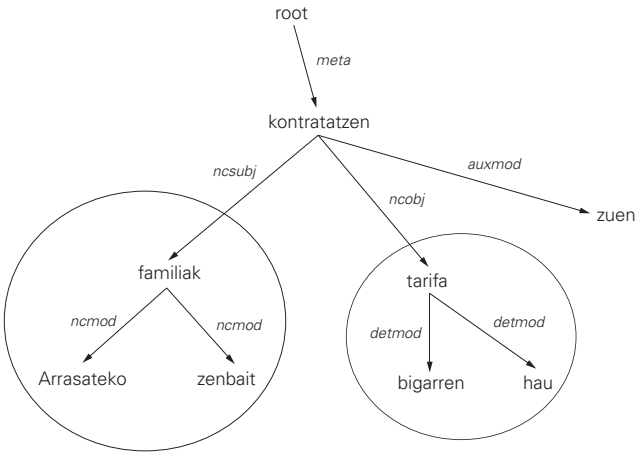

Figure 2

Example of the dependency tree of a sentence

4. Analysis of coordinate noun phrases

Coordination is as problematic for the Dependency Grammar formalism as for other theoretical traditions. In order to capture the idea that the constituents that are coordinated are at the same level, we have considered two options extensively explained in the literature (Böhmová et al. 2003, Järvinen et al. 1997): i) to presume one of the elements coordinated depends on the other and ii) to add a new imaginary node maintaining the coordinated elements at the same level. In our case, for computational reasons, we opt for the second one which it is expressed by considering the coordination element as a head of the coordinate phrase.

Figure 3 shows an example of coordination at the level of the noun phrase that illustrates our choice.

(8) *Horixe zen magoak eta nik genuen sekretua.*
(That was the secret the illusionist and I had.)

In (8), the coordinated elements "magoak" and "nik" are represented at the same level and they have as their governor the connective "eta", which takes the dependency relation with respect to the verb, in this case "ncsubj".

The dependencies associated to this phenomenon in the example are the following:

lot (emen, eta, magoak)
lot (emen, eta, nik)
ncsubj (erg, genuen, eta, nik, subj)

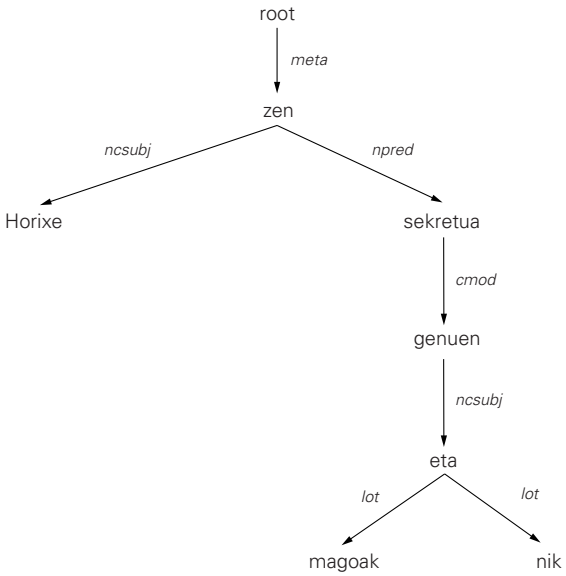

Figure 3

Example of the dependency tree of NP coordination

The first slot in "lot" relation expresses the type of coordination. So we use "emen" for copulative coordination, "aurk" for adversative, "haut" disjunctive, "espl" for explicative and so on.

The explanation given above could be extended to the coordination of more than two elements.

5. Conclusions

This paper has described the noun phrase structure by means of the Dependency Grammar theory. It represents the first formalization for the annotation of Basque NP as part of a more general work, the aim of which is to describe all the syntactic phenomena and which will form the basis for the development of NLP applications.

We have started by setting out the reasons for creating BDT Treebank, i.e., a syntactically tagged corpus. After considering and analyzing the main existing possibilities, we have decided to follow the formalism based on dependency relations for basically two reasons: first, because it is known to be more suitable for languages with a free word-order like Basque, and second, because, apart form being intuitive and easy, its flexibility allows the introduction of new types of tags such as those corresponding to thematic roles. The latter is an important aspect for any research we will conduct in the future.

We have taken the step of analyzing the syntactic structures by explicity expressing the relation between the head and the dependent. Additionally, we have found solutions to problems that have emerged in performing this analysis (such as coordi-

nation). During the annotation process we are currently carrying out (we now have 100,000 words tagged and plan to analyze 300,000 words), new refinements and proposals will be needed. To conclude, we would like to stress the urgent need for a syntactically tagged corpus, which would serve to evaluate and improve the *parser* for Basque that we are developing in the group. And it will also be a key ingredient for syntactic studies from a theoretical point of view. The *treebank* can be used to verify our linguistic intuitions.

References

Abeillé, A., 2003, *Treebanks: Building and Using Parsed Corpora*, Kluwer Academic Publishers, Dordrecht.

Aduriz, I., Aranzabe, M. J., Arriola, J. M., Atutxa, A., Díaz de Illarraza, A., Ezeiza, N., Gojenola, K., Oronoz, M., Soroa, A., Urizar, R., 2006, "Methodology and steps towards the construction of EPEC, a corpus of written Basque tagged at morphological and syntactic levels for automatic processing", in A. Wilson, P. Rayson, and D. Archer (eds.), *Corpus Linguistics Around the World,* Rodopi, 1-15.

Agirre, E., Aldezabal, I., Etxeberria, J. and Pociello, E., 2006, "A Preliminary Study for Building the Basque PropBank", in *Proceedings of the 5th International Conference on Language Resources and Evaluations (LREC)*.

Artiagoitia, X., 2002, "The functional structure of the Basque noun phrase", in Artiagoitia, Lakarra & Goenaga (eds.), *Erramu Boneta: Festschrift for Rudolf P. G. de Rijk*, Supplements of *ASJU*-UPV/EHU, Bilbao, 73-90.

Böhomovà, A., Hajic, J., Hajicova, E. and Hladka B., 2003, "The PDT: a 3-level annotation scenario". In Abeillé (ed.), *Treebanks: Building and Using Parsed Corpora*, Kluwer Academic Publisher, Dordrecht.

Brants, T., Skut, W., Krenn, B. and Uszkoreit, H., 2003, "Syntactic annotation of a German newspaper corpus", in Abeillé (ed.).

Bunt, H., Carroll, J. and Satta, G., 2004, *New Developments in Parsing Technology,* Kluwer Academic Publisher, Dordrecht.

Carroll, J., Briscoe, T. and Sanfilippo, A., 1998, "Parser evaluation: a survey and a new proposal", in *Proceedings of the International Conference on Language Resources and Evaluation*, Granada, Spain, 447-454.

Charniak, E., 2000, "A Maximum-Entropy-Inspired Parser", *Proceedings ANLP-NAACL'2000*, Seattle.

Collins, M., 1997, "Three generative lexicalised models for statistical parsing", *Proceedings ACL'97*, Madrid.

—, 2000, "Discriminative Reranking for Natural Language Parsing", *Proceedings ICML-2000*, Stanford.

Eguzkitza, A., 1993, "Adnominals in the Grammar of Basque", in J. I. Hualde and J. Ortiz de Urbina (eds.), *Generative Studies in Basque Linguistics*, John Benjamins, Amsterdam, 163-187.

Goenaga, P., 1980, *Gramatika bideetan*, 2. argitalpen zuzendua, Erein, Donostia.

Hudson, R., 1990, *Word Grammar*, Basil Blackwell, Oxford.

Järvinen, T. and Tapanainen, P., 1997, "A Dependency Parser for English", Technical Report, No. TR-1, Department of General Linguistics. University of Helsinki.

Kakkonen, T., 2005, "Dependency Treebanks: Methods, Annotation Schemes and Tools", in *Proceedings of the 15th Nordic Conference of Computational Linguistics*, Finland.

Karlsson, F., Voutilainen, A., Heikkilä, J and Anttila, A., 1995, *Constraint Grammar*, Mouton de Gruyter, Berlin.

Laka, I., 1993, "Unergatives that Assign Ergative, Unaccusatives that Assign Accusative", in J. Bobaljik and C. Phillips (eds.), *Papers on Case & Agreement 1*, MIT Working Papers in Linguistics, Cambridge MA, 149-172.

Marcus, M., Santorini, B. and Marcinkiewicz, M. A., 1993, "Building a large annotated corpus of English: The Penn Treebank", *Computational Linguistics* 19, 313–330.

Mel'cuk, I. A., 1988, *Dependency syntax: theory and practice*, State University of New York Press, Albany.

Montemagni, S., Barsotti, F., Battista, M., Calzolari, N., Corazzari, O., Lenci, A., Zampolli, A., Fanciulli, F., Massetani, M., Raffaelli, R., Basili, R., Pazienza, M. T.; Saracino, D., Zanzotto, F., Nana, N., Pianesi, F. and Delmonte, R., 2003, "Building the Italian Syntactic- Semantic Treebank", in Abeillé (ed.) 189–210.

Nugues, P. M., 2006, *An introduction to Language Processing with Perl and Prolog.* Springer-Verlag. Berlin Heidelberg.

Odriozola, J. C. and Zabala, I., 1993, *Izen-sintagma. Idazkera teknikoa II,* EHUko argitalpen zerbitzua, Bilbo.

Oflazer, K., Zynep, D. and Tür, G., 1999, "Design for a Turkish Treebank", *Proceedings of Workshop on Linguistically Interpreted Corpora*, at EACL'99, Bergen.

Skut, W., Krenn, B., Brants, T. and Uszkoreit, H., 1997, "An Annotation Scheme for Free Word Order Languages", *Fifth Conference on Applied Natural Language Processing (ANLP'97)*, Washington, DC, USA, 88-95.

Sleator, D. and Temperley, D., 1993, "Parsing English with a link grammar", *Third International Workshop on Parsing Technologies*.

Tapanainen, P. and Järvinen, T., 1997, "A non-projective dependency parser", *Proceedings of the 5th Conference on Applied Natural Language Processing* (ANLP'97), 64-71.

— and —, 1998, "Dependency concordances", *International Journal of Lexicography* 11: 3, 187-203.

Tesnière, L., 1959, *Eléments de Syntaxe Structurale*, Librairie Klincksieck, Paris.

Trask, R. L., 2003, "The Noun Phrase: nouns, determiners and modifiers; pronouns and names", in J. I. Hualde & J. Ortiz de Urbina (eds.), *A Grammar of Basque*, Mouton de Gruyter, Berlin-New York.