

An LFG parser for Basque (I)*

JOSEBA ABAITUA
(University of Manchester.
Institute of Science and Technology)

0. *Abstract*

In recent years, research in the realm of Artificial Intelligence has become increasingly involved with linguistic analysis, with the aim of designing computers capable of handling natural language. This is an obvious consequence of the determining role that computers of this kind are bound to play in areas such as machine translation, expert systems, access to data banks, intelligent planning systems, etc. The initial motivation for the work we present here arose from the idea that Basque too should be a language that benefits from practical applications in these areas. What is presented in the following pages is, however, a small scale project in this sense.

The aim of the work is twofold. On the one hand, some of the most controversial aspects of Basque syntax are investigated, and eventually a formal description of a grammar for a subset of the language is proposed. On the other hand, it is shown how a grammar so defined can be implemented as computer software. A parser is taken to be that part of a natural processing system which determines whether a string of words is well formed, according to a particular grammar, and describes the structures the string can have.

The scope of the grammar we have implemented here is limited to very simple constructions, and it should therefore be understood that no other application than that of serving as an illustrative example could be derived from it in its present state. However, although there is still little practical application to be obtained from the following pages, it is believed that the range of topics covered in a more theoretical level is quite significant.

Computer implementations impose an important constraint on linguistic analysis, that of being well formalized. Setting out from this prerequisite, some new grammatical frameworks have been evolved in recent years with the idea of offering strictly formalized mechanism for linguistics analyses. The idea of capturing the developments of contemporary linguistic theories and the model chosen in our project is one of the best known new frameworks of this trend, the Lexical Functional Grammar of Bresnan and Kaplan (1982). In addition to the power of its formal devices, LFG is believed to maintain its rank as a proper theory of language, in the sense of being able to provide an

(*) This paper contains the research carried out in the Centre for Computational Linguistics at the University of Manchester Institute of Science and Technology during the course 1984/85. The work was presented as a thesis and the text offered here maintains, with some minor alterations, the original version. This research was supported by a grant from the Department of Education, Universities and Research of the Basque Government.

alternative paradigm to other existing theories like Chomsky's (1981) Government and Binding, for example. Thus, it is essential under LFG as well that any proposed analysis be linguistically valid, and not only computationally feasible. This is an important dimension of the analyses offered in the following pages. Chapter one presents a theoretical justification for the project. It explains the reasons for adopting LFG as the framework in which to describe the grammar of Basque. It also indicates the principal difficulties this theory might encounter with regard to some of the controversial aspects of Basque. This includes a discussion on grammatical relations and their connection with Basque as an ergative and non-configurational language.

Chapter two outlines an initial set of phrase structure rules for Basque. The constituent order of the language, a structural position for focus and the syntactic encoding of grammatical functions are analysed.

Chapter three extends the syntactic analysis of the previous chapter. The first section is devoted to a formalization of the periphrastic conjugation of verbs. The second section concerns an analysis of long distance dependencies, also known as syntactic binding or constituent control. The analysis is related to the solution proposed for focus, and it investigates how appropriate an LFG approach is to this problem.

Finally, chapter four offers a broad description of the parser. Firstly, a brief account of the programming language employed is given, which includes an assessment of Prolog's suitability for LFG. Secondly, we provide a step-by-step account of the LFG parser for Basque, the grammar of which has already been sketched.

The parser has been written in C-Prolog 1.2 and it is run on the unix bsd 4.2 on a VAX 750/11 at UMIST. Some examples of the output of the parser are given in Appendix 1.

1. Grammatical relations. Basque and lexical-functional grammar

In this chapter we explain reasons for having adopted a particular grammar framework, namely, LFG. It will be shown that this has relevant consequences, not only for the practical purposes of the project, but also from a theoretical point of view. I present a brief discussion of grammatical relations, the connection of this discussion with Basque as an ergative language and why LFG promises a priori to be a suitable explanatory model.

1.1. *Aim of the project and reasons for choosing LFG*

The main aim of the project is to write a parser for a subset of Basque grammar which is as broad as possible. Lexical-Functional Grammar (LFG) has been chosen as a framework in which to describe the grammar of Basque.

LFG appears to be a feasible model for computational implementation. Besides the system under development by the Centre for Computational Linguistics at UMIST, several other systems have been implemented before, Frey and Reyle's (1983) and Yasukawa's (1984) examples are the most representative, besides Kanplan's own system (in Lisp).

We must bear in mind that LFG grew out of and brought together ideas from Transformational Linguistics (i.e. Bresnan 1976) and Com-

putational Linguistics (i.e. Kaplan's ATN and General Syntactic Processor, (1975)). A formalism like LFG can be seen as a large class of definite clauses. The atomic grammar symbols of context-free grammars are replaced by more complex symbols, that can eventually be matched by unification¹. LFG, in other words, can be translated into definite clause grammar (DCG) of Pereira and Warren (1980) which makes it a particularly suitable formalism for being implemented in Prolog, as Prolog is in fact a definite-clause proof procedure. On the other hand, it is hoped that the development of a formal system like LFG will help to represent syntactic knowledge in a more appropriate way than simply representing it by arbitrary Prolog programs embedded in definite-clause grammar rules.

Basically, LFG has two stage models of syntactic description which are efficiently implementable: i) A context free analysis of the surface structure of a sentence, 'c-structure'. And ii) a computation of the grammatical relations of the sentence, 'f-structure', which is achieved by the solution of functional equations associated with the nodes in c-structure. The structure of the sentence is built by application of lexical rules in addition to the syntactic rules. The fact that the lexicon is a primary knowledge base makes LFG an even more versatile model for computational purposes. There are also important linguistic reasons for choosing LFG, as will be shown below.

1.2. *Status of Grammatical Relations in linguistic theory and their role in LFG*

Discussion of grammatical relations has occurred fairly frequently in recent linguistic research (see for example Hoekstra, 1984 and Williams, 1984)². The notions 'subject' and 'object' as primitives were posed within the Relational theory developed by Perlmutter and Postal (1974). This grammar was presented as an alternative to the current Transformational paradigm, in which subject and object were derived from structural primitives, such as precedence and dominance. Besides the advantage of recapturing the notions of the Traditional Grammar, the Relational Grammar (RG), unlike the Transformational, allowed direct account of cross-linguistic generalizations, passivization for example, in a very straightforward way.

The functional theory LFG also considers subject and object as

(1) The unification procedure for matching complex symbols can be ascribed to the work by Robinson (1965), Colmerauer (1978), Pereira and Warren (1980), and Kay (1983).

(2) Williams' (1984) argumentation against grammatical relations as primitives seems fairly damning with regards to LFG. Given the redundancy of notions like «object» and «subject» as primitive elements for a theory of grammar, it appears as if the sole argument to use in defence of LFG's functional primitives were their useful properties for grammatical descriptive purposes, similar to the usefulness of concepts like *noun* or *verb*, etc. Thus, it could be stated, in short, that LFG deals with grammatical relations as primitives just because these are practical elements for the description of a language, as it was, probably, the reason that made the Traditional Grammar to develop them. In Basque these grammatical functions cannot be so clearly defined in Williams' terms.

universal primitives, and calls them grammatical functions. Grammatical functions are, in fact, a central notion in LFG, a theory that operates on grammatical functions, like SUBJ, OBJ, OBJ2, etc., as well as on structural categories, like NP, V, PP, etc.

Grammatical relations constitute the association between the predicate argument structure of a sentence and its surface structure. They are encoded in the f-structure which together with the c-structure are the two levels of syntactic description of a sentence of a language. F-structures represent grammatical relations in an invariant universal format, which is independent of language particular differences in surface form.

In LFG, lexical entries can represent semantic argument structures independently of phrase structure (PS) forms. In other words, the various functional arguments of a sentence are already encoded in lexical entries and no PS manipulation is needed to express the grammatical relations. The syntactic instantiations of the grammatical functions can be read off from functional annotations to the surface structure. The theory requires no 'normalized' PS representation to express predicate argument relations, the structural component of the grammar is simplified, and the entire transformation derivation is replaced by a single level of PS. In this sense, LFG is what Baker (1983) calls a 'monostratal' theory.

The mapping between thematic role structure and surface forms is not effected by syntactic transformations (or structural computation), it is effected by correlating the grammatical functions that are assigned to the lexical predicate argument structure of the lexical form with the grammatical functions in the surface structure. The guiding idea of LFG is that only lexical rules can alter these functional argument correspondences.

Coming back to the postulation of grammatical functions as universal primitives, it must be pointed out that a universal definition of the notion subject is a rather polemic point, particularly when it concerns ergative languages such as Basque.

1.3. *Nature of Ergativity*

An attempt to provide a definition of the notion subject as a universal was that of Keenan (1976). (See also Perlmutter, 1982, for a more recent approach). Keenan sought to abstract a set of properties which were characteristics of subject noun phrases (NPs), and then determine which combination of these would be jointly necessary to fix the subject of an arbitrary sentence in an arbitrary language. However easy though this seemed to be, the assumption of the universality of subject remained contentious (cfr. Schachter's work for Phillipine languages, 1976, 1977).

Plank (1979) regarded ergativity as a serious challenge to the syntactic theory and also a new ground for testing grammatical relations such as subject and object. This particular task was undertaken by Comrie (1978) and Dixon (1979). Dixon proposed to define subject as a universal notion, grouping the subject of a transitive verb, or agent (A), and the subject of an intransitive verb (S). The object of a transitive verb is defined in semantic terms as the patient (P). For an ergative language, Dixon says, (S) and (P) would be marked by the same absolutive (ABS) case, whilst (A) will get a different one, the ergative (ERG) case. (In contrast with accusative languages, in which (S) and (A) get the nominative case and (P) the accusative).

Semantic roles have played an important part in the identification of grammatical functions, and this has been even more obvious in the case of ergative languages (see for example more recent papers like Comrie, 1983, or Cooreman, 1984). Dowty (1982) has argued against such strategy. There is no need to maintain notions like 'agent' and 'patient', he says, if it is assumed a referential theory of semantics and Montague's compositional principles for connecting syntax with meaning³.

This discussion on semantic roles aside, a syntactic manifestation has also been regarded in addition to the morphological distinction of ergativity. An ergative syntax would be characterized by rules such as subordination, deletion and so on, treating (P) and (S) in the same way and (A) differently. Thus, one could say that when (P) and (S) are

(3) Dowty's approach for defining grammatical functions in Montague's grammar is done by means of a syntactic rule of the form of an ordered triple: The first member names a syntactic operation (F), the second member is the sequence of categories that are the inputs to the rule, the third member is the category of the output of the rule. For Basque, thus, the subject-predicate rule would be as in (1), and the verb-object rule as in (2), given the ordered triples (3) and (4):

- (1)
$$\begin{array}{c} \text{[Jon dator]}_t \\ \swarrow \quad \searrow \\ \text{Jon}_T \quad \text{dator}_{IV} \\ \text{'John-ABS comes'} \end{array}$$
- (2)
$$\begin{array}{c} \text{[Jonek goxokia dakar]}_t \\ \swarrow \quad \searrow \\ \text{Jonek}_T \quad \text{[goxokia dakar]}_{IV} \\ \text{goxokia}_T \quad \text{dakar}_{IV} \\ \text{'John-ERG the sweet-ABS brings'} \end{array}$$
- (3) Subject-Predicate rule (F1):
S1: $\langle F1, \langle IV, T \rangle \rangle_t$
- (4) Verb-Predicate rule (F2):
S2: $\langle F2, \langle TV, T \rangle \rangle_{IV}$

treated alike in contrast to (A) we refer to ergative syntax. It is known, however, that most of ergative languages have non-ergative syntax (cf. S. Anderson, 1976; Comrie, 1978; and Dixon, 1979).

Following Dowty, we may say that there exist three ideal types of ergative languages: deep ergative languages (in which not only case marking but also all other syntactic properties that distinguish subjects from objects follow an ergative pattern), surface ergative languages (in which relatively few characteristics follow an ergative pattern), and split ergative languages in which ergative patterning appears consistently in some tenses but never in others.

1.4. *Basque as an ergative language, and the need for subject function*

The controversy on the nature of subject has involved studies of the Basque language, a language known for being fully ergative morphologically. Most of the papers above have references to Basque and even are entirely devoted to it, as the one by Brettschneider (1979), for example. Anyhow, the controversy among vascolegists is not new.

There has been a widespread disagreement as to what, if anything, is the subject of a Basque sentence. Trask (1984) recounts the diversity of opinions regarding the notion subject.

There is, first, the position of those who propose (P) as the subject of a transitive verb, instead of (A). Adherents to the theory that Basque sentences are passive will support this position, best presented by Gavel (1930). A similar opinion is shared by Entwistle (1936), and lately and although they don't consider transitive verbs as passives, Naert (1956) and Tchekhoff (1978).

A second position posits the non-existence of subject in Basque sentences, both transitive and intransitive. Martinet (1958, 1962) and Wilbur (1970, 1979) support this idea advocating some version of Case Grammar. See also Van Valin (1977) for a similar approach, and Azkarate *et al* (1981) for a different one. Recently Bossong (1985) has also supported this position which has been contested by Abaitua and Trask (forthcoming).

These two positions about the notion of subject base their conclusions on the nominal and verbal morphology of Basque, rather than on syntactic arguments. (See Trask, 1984, for a further discussion).

A third position, which as a matter of fact will also agree with the criteria proposed by Dixon (1979), is that of considering (A) the subject of a transitive sentence. This is the case for Laffite (1931), Michelena (1977, 1978), Rotaetxe (1978, 1980), Heath (1977), S. Anderson (1976), and Trask (1981, 1984). It is also the case with reservations for Rebuschi (1978), J. Anderson (1977) and Brettschneider (1979). The case grammar interpretation of Rebuschi and Anderson does not attempt to dispense completely with the grammatical function subject.

Trask (1984) undertakes in his approach the tests for subjecthood developed by Keenan (1976) (also Comrie, 1978, is taken into consideration). Trask's main effort is to present a substantial body of syntactic evidence. According to this analysis, and in connection with the active theory for transitive verbs, ABS NPs in intransitive sentences (S) and ERG NPs in transitive sentences (A) exhibit properties of syntactic unity, which makes Dixon's universal proposals suitable for Basque to be morphologically ergative, but also obviously syntactically accusative type (cf. Comrie, 1983). Trask shows how Equi-rules apply to all grammatical subjects (either in ABS or ERG) whatever semantic role they have. Similarly, subject NPs appear unable to be reflexive or reciprocal and suffer demotion in causativization as typically exhibited by subjects in other languages⁴.

(4) Two observations are due at this point. The first regards the definition of transitive and intransitive verbs. Following the 'unaccusative' hypothesis (first discussed by Perlmutter 1978 and latterly by Burzio 1981), Beth Levin (1983) proposes the allocation of Basque intransitive verbs into two verbal subclasses: unaccusatives and unergatives. Unaccusative verbs would be universally characterized (in Government and Binding theory) by their inability to assign a thematic role to the subject. These would be identified in Basque by those intransitive verbs that select the auxiliary *izan* and have their argument marked with the absolutive case. In contrast, unergatives, like transitive verbs, assign a thematic role to the subject and would be identified in Basque by those intransitive verbs that select the auxiliary *ukan* and have their argument marked ergative:

(1) unergative:

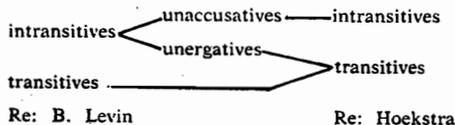
gizonak korritu du
man-ERG run AUX-ukan
'the man run'

(2) unaccusative:

gizona etorri da
man-ABS come AUX-izan
'the man came'

If this classification of intransitive verbs were maintained it would add further complexity to the present discussion of subjecthood in Basque. However, if we adopt Hoekstra's (1984:227) redefinition of transitivity, this problem could be avoided. Hoekstra dismisses the traditional classification of verbs where transitives are defined by those that combine with an NP, or that subcategorize for an NP. He proposes that «a more sensible classification of verbs could be made in terms of the property of selecting thematic subject». So, since the inability of selecting a thematic subject is what distinguishes unaccusative verbs from the rest, we could disregard Levin's classification and maintain the traditional idea that intransitives in Basque are solely those which select the auxiliary *izan*, i.e., those that do not assign a thematic role to the subject (figure 3).

(3)



The second observation to be made here is connected to this concept of the inability of intransitive verbs to assign a thematic role to the subject. In government and binding terminology this means that verbs with this property lack an argument subject position, both in the lexicon and at D-structure; or, in other words, that their sole argument is not really a subject but an object. In languages such as English or Italian, with an accusative system of case marking, a verb that fails to assign a thematic role to the subject, also fails to assign case to the object position (Burzio's generalization), but in order not to violate the Case Filter, a deep-object to be moved to the subject position at S-structure to inherit nominative case. Basque would differ from accusative systems of case marking in this aspect, B. Levin explains, in that intransitive [unaccusative] verbs can assign absolutive case to their object directly. What would be at stake now would not be the identification of ergative noun phrases as subjects any more, which is the main motivation for Trask's (1984) argumentation, but rather the subjecthood of an intransitive verb's absolutive noun phrase. In the government and binding theory we might be forced to agree with B. Levin and accept that this argument is in fact an object. A solution to this problem in the framework of LFG has been investigated in Abaitua (in prep.). Anyhow, for the purpose of the present paper, we maintain the classical notion that the (S) argument, marked absolutive, of an intransitive verb is its subject. We also assume, for the time being, that Trask's evidence

Basque is, therefore, a surface ergative language. Dowty's (1982) proposal for the identification of the function subject in such languages also agrees with the facts of Basque: «If the IV consists of a TV plus object, the subject will be marked with ERG case, but if IV is intransitive, the subject will be marked with ABS case. The object of a TV is uniformly marked with ABS case». (Dowty, 1982:112). In a later article, Trask (forthcoming) recalls again that NPs with the function subject cannot be said to be categorized by the verb in Basque and, therefore, that an analysis for Basque should provide an efficient mechanism to identify the function. Under the GPSG framework, which is the one that Trask follows, a configurational analysis is needed in order to identify grammatical functions, in similar terms to Dowty's proposal. In other words, in GPSG grammatical functions are represented by a semantically interpreted constituent structure. This would be a good reason to reject a non-configurational analysis for Basque, in Trask's opinion⁵.

The distribution of languages among configurational and non-configurational is a topic of debate. For example, there are suggestions (Horrocks, 1983) that a typically configurational language such as English may have 'flat' structure in sentences in which a gap is required. For Basque the claim that all sentences have a flat structure is quite well accepted; Trask appears to be an exception. Among those who consider Basque non-configurational are Azkarate *et al* (1981), Ortiz de Urbina (1983), Geggus (1983), Jelinek (1984), and the strongest argumentation is put forward by Rebuschi (1984). See also de Rijk (1978: 85), Goenaga (1980: 204) and Michelena (1981: 79). Rebuschi shows the problems that arise if we define subject and object in terms of configurations, mainly due to the clumsiness of the VP node when it has to handle the free constituent order allowed in Basque. Still, Trask argues that the syntactic properties of Basque appear to be more similar to those of familiar configurational languages than to those of well established W(*) languages⁶. The existence of subject

for the syntactic unity of the notion subject is acceptable. Disagreement with Trask will come from his proposal of a configurational structure to identify this notion of subject in Basque (see below).

(5) GPSG claims that subjects are not subcategorized by the verb. One of the arguments against a non-configurational analysis for major constituents of Basque follows from this theory's internal prediction.

(6) In close connection to the syntactical argumentation for an subject/object asymmetry is for Trask (1984:251—) a syntactic process of agreement, a notion that he introduces. As is known, agreement in Basque is due to verbal morphology. The verb may agree in person and number with an NP which is subject, object or second object, either in ergative, absolutive or dative case (the so-called grammatical cases by Euskaltzaindia, 1979). A verb may show agreement with from none to three NPs, and there may never be more than one morph from each group in a single verb form. Now, in addition to this type of agreement Trask proposes a further one, a «syntactic» agreement:

Subject NPs are not governed by a verb, that is, that the presence of a verb in a sentence does not determine the case of the subject. For example, the majority of Basque verbs can be used either transitively or intransitively, and therefore the subject will stand in the appropriate case, ergative or absolutive. Trask hence proposes that a subject in Basque might carry a morpheme indicating the transitivity of its verb, this is to say, that it would agree in transitivity with its verb.

Trask accepts the peculiarity of the above statement, as it involves the agreement of a NP with a verb, something that, as he is aware of, is not allowed by current theories of agreement. Within the GPSG framework (Gazdar, 1985), agreement is unidirectional, so that it X which can

is seemingly well established, he argues. It is true, in fact, that a successful non-configurational analysis has not yet been proposed, and when it has been attempted, no convincing account for grammatical functions has been achieved.

An LFG approach seems to be promising. Remember that in LFG grammatical functions are not reducible to canonical PS configurations; on the contrary, the PS categories themselves appear reducible to function primitives (cf. Jackendoff 1977). The relation between structural configurations and grammatical functions is in LFG many to many, as the correspondence between semantic predicate argument structure and grammatical functions.

What this all implies is that the proposed configurational analysis of Basque for maintaining the function subject is not a condition for LFG. In other words, the usefulness of subject in Basque grammar does not prevent an analysis of it as non-configurational.

In the next chapter we will present a non-configurational identification of grammatical functions in Basque. A flat structure at S level will be proposed. This will not contradict the fact that Basque is a partially configurational language, in which only maximal projections are order free.

2. An initial set of phrase structure rules for Basque

This chapter will examine the clause structure of Basque. One of the basic features of lexicalist grammar is that of having relational primitives. It is claimed that these relational primitives are a useful tool to account for various phenomena in a language, and therefore also useful for our description of Basque.

Basque happens to be one of those languages in which the identification of traditional concepts such as «subject of», «object of», etc., have been widely debated. Here, however, we are not going to relapse into this debate. Given the usefulness of dealing with syntactic notions such as subject and objects (calledas, we know, grammatical functions in LFG), the next step to be resolved would be their identification. In the case of those languages referred to as having fixed word order, grammatical functions can be easily identified from word order. But, if

agree with Y, but not then Y with X. In other words, a verb is allowed to agree with its NP arguments, but NPs are not supposed to be able to agree with verbs. Trask suggests finally that we consider transitivity as a feature not to be attached to verbs, but to the entire sentence, a feature which typically manifests itself in the verb, but which can also turn up elsewhere (as Australian Lardil object NPs agree with their verbs in tense, for example).

We suppose the LFG analysis of agreement is unidirectional as well, in that a verb's lexical entry can contain a statement like (\uparrow SUBJ NUM) = SING, rather than a noun's lexical entry saying something about the number of the verb. An NP could be given an equation like (\uparrow PRED TRANS) = +, showing that the PRED of the S was transitive. One minor problem would be ensuring this equation only appeared on the NP, not on the head N (where it would be saying that the NP was transitive, not the verb).

However, as will be show below, none of these equations is necessary. It is sufficient to state that a transitive verb (which subcategorizes for first objects) requires its subcategorized subject to be marked ergative. To this extent, agreement will be understood to be just directional.

we assume the convenience of building a grammar which, besides being able to characterize all possible surface structures (by means in LFG of a context free type formalism), allows no deletion or movement operation, it will become clear that a configurational approach to account for grammatical functions is not of much use in Basque. That is to say, in Basque the order of constituents in the sentence does not provide any relevant information as to what grammatical functions are involved. Not unless an underlying word-order is considered, together with a set of transformational rules to be applied on that underlying structure¹. This however, does not fulfill our requirement of not having derivation rules.

Firstly, we will be looking at some suggestions that have been put forward to determine word-order in Basque. Secondly, the grounds for postulating a structural position in Basque will be studied. Finally, a new mechanism for encoding grammatical functions in Basque will be presented.

2.1. *Word order phenomena in Basque*

Recent research on text analysis techniques have put forward evidence that the final order of constituents of a sentence (as well as the choice of some constructions instead of others, e.g. passives and actives) is determined very much by stylistic and pragmatic factors². It can be said, consequently, that the scope of the syntactic notion of sentence 'S' may only partially predict the order of its elements. And this will vary according to the degree of configurationality of a language. The freer word order a language has, the more influenced it will be by pragmatic factors. On the other hand, because of the generative role a theory of grammar is committed to, the provision of mechanisms that allow us to generate all acceptable constructions has come to be rather more important than the mere description of the language itself. In other words, when formalizing the rules that will generate sequences of a particular language, we are concerned with providing the necessary device that will enable our formalism to generate all possible constructions, regardless of their being more or less common in the language.

For Basque, one would expect and one finds a significantly fluid word order, given the rich case marking system. All permutations of maximal projections of lexical categories are, in fact, possible on the whole, with a few restrictions in some embedded sentences. Therefore,

(1) Following Slobin (1982), Pinker (1982) says: «for case inflected languages [like Basque], children will utter sentences in the dominant word order (if there is one) and will use the dominant word order as a cue in comprehending sentences, before they have mastered their language's case system».

(2) An interesting work with regards to this point for Spanish is John Green's (1976). See also Heles Contreras (1976).

when supplying our grammar with generative rules, one has to endow the grammar with a mechanism that will make possible all kinds of constituent order. In terms of frequency, S(ubject), O(bject), V(erb) is the most common order (de Rijk, 1969, Villasante, 1980). But again, for our purpose what matters is to provide a mechanism that will account for all other possible orders, including the least common.

Thus, given a sentence like (1),

- (1.a) neskak panpina dakar
 the girl-ERG the doll-ABS brings
 'the girl brings the doll'

all permutations of its elements are possible:

- (1.b) neskak dakar panpina
 (1.c) badakar neskak panpina
 (1.d) badakar panpina neskak
 (1.e) panpina neskak dakar³
 (1.f) panpina dakar neskak

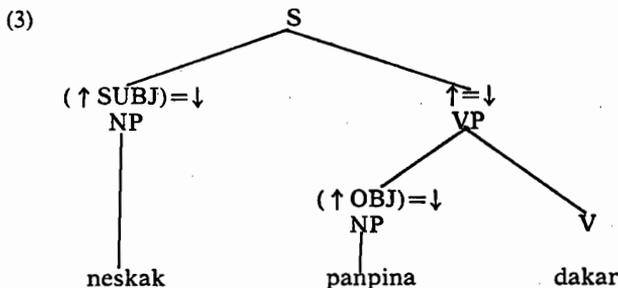
We set out with a classical bipartition rule such as (2), which will enable a configurational encoding of grammatical functions.

$$(2.a) \quad S \rightarrow \quad NP \quad VP \\
 (\uparrow \text{SUBJ})=\downarrow \quad \uparrow=\downarrow$$

$$(2.b) \quad VP \rightarrow \quad NP \quad V \\
 (\uparrow \text{OBJ})=\downarrow$$

The equations here convey that the NP to the right of the arrow in (2.a.) is the subject of its mother node S, and that all the information about the VP is also about the S in the f-structure of the sentence. Thus, the NP to the right of the arrow in (2.b.) is also the object of S.

With the rules in (2) no problems will arise in generating the common sentence (1.a.):



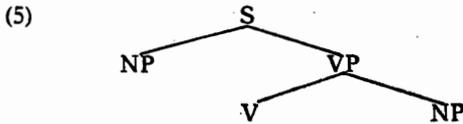
However, we know from (1.b.) that the constituent order in the VP node could well be the other way round. We could escape from this dilemma by proposing a transformation (or, scrambling rule) in our

(3) The order OSV has been suggested to be ungrammatical, unless subject is focused (see below). This is the opinion of Goenaga (1980), Trask (1984), and it has also been implicitly stated by Michelena (1981) (see note 5).

PS rule, similar to the one proposed by Andrews (1982:430) in his analysis for Icelandic and Mohanan (1982:511) for Malayalam, as in ⁴.

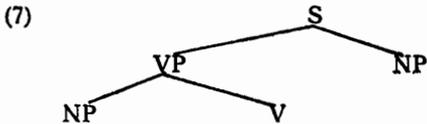
(4) NP V → V NP

Thus, a c-structure like (5) would also be possible:

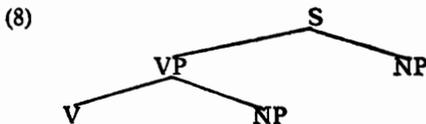


To account for (1.f.), a new transformation on our (2.a.) rule could be proposed for (7).

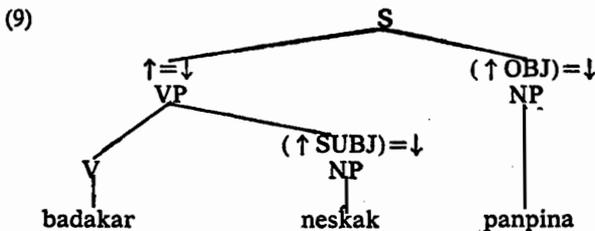
(6) NP VP → VP NP



And making use of both (4) and (6), the following structure for (1.d.) would be possible as well:



But these transformations on PS rules fit awkwardly into LFG framework, besides the fact that no transformation could be proposed anyway to account for (9), in (1.c.),



which is a grammatical sentence of Basque (but obviously not its tree structure).

No alternative remains, therefore, but to propose a non-configurational initial rule of the type of (10).

(10) S → NP* V NP*
 ↑ = ↓

(4) It is not clear that the idea of «unmarked» word order really means anything in GPSG. Their ID/LP format for grammars does provide a way of handling free word-order language, which would have fewer LP rules.

So far, the annotations (\uparrow SUBJ) = \downarrow and (\uparrow OBJ) = \downarrow that were added to [NP,S] and [NP,VP] respectively in (2.a-b) are now omitted in (10). This is because NPs can stand for subject as well as for object function. The dominance principle of (2) plays now no role, since all NPs are daughters of the same node S. A new way for encoding grammatical functions has to be found (see below).

As has been said, by means of a rule like (10) we do not capture the fact that the most frequent constituent order in Basque is SOV. In a survey carried out by de Rijk (1969) it was shown that although all six constituent orders are possible, SOV was the most statistically predominant order (over 50%) as well as the preferred order by most native speakers. Basque also conforms with Greenberg's list of linguistic universals concerned with languages of dominant SOV order. All the same, when de Rijk looks for language internal reasons in relative clauses to account for this, he concludes: «it is far from obvious that the structure of the relative clause in Basque supports the verb final theory. All we can say is that it does not contradict it. More generally, the same can be said for all the facts adduced in this paper. They are consistent with an underlying SOV order, but they do not require it». The underlying SOV order has again been proposed to be taken as the unmarked word order in Basque by Trask (1984). He says: «All other observed word orders of Basque sentences could be accounted for with simple and general rules».

We are going to present next some ideas as regards marked and unmarked sentences in Basque, which might further shed some light on this question.

2.2. *The Focus Position*

We commented before that in the production of sentences in a text, the notion of treating these strings of sentences as a set is relevant to constituent order. This cross-clause determinant, based on pragmatic principles rather than on syntactic ones, has been somehow vindicated by Michelena (1981:63). The order of constituents in the sentence of Basque is underdetermined as regards syntactic criteria, he says, and only pragmatic or stylistic criteria can be adduced to account for it⁵. It is important to emphasize that statements like do not contradict the fact that eventually the predominant word order of Basque is SOV. Yet Michelena (1978:221) also recalls the order SOV as being fully common.

(5) What Michelena (1981: 62-63) actually states is (in a rough translation from Basque): «The mentioned rules [Altube's rules for constructing sentences and focus positioning (see below)] do not apply when one is just thinking, or just talking to someone. As Goenaga says, when a language output is in the form of a narration, the order of constituents becomes free with regards to Altube's rules and the focus constraint is not obeyed (for example SOV, OVS, VSO, VOS, etc.). This is because there is no need for Altube's focus constraint to be obeyed, I claim. One could hesitate whether one order is preferred to another in terms of stylistic suitability, but never in terms of grammaticality».

There is one factor, however, that has been postulated in the literature as being of syntactic relevance in connection with word order in Basque. It is the sometimes so called «distinguished structural position for focus»⁶. A rule for positioning a focused constituent in a fixed place in the sentence was first thought up by Altube (1920, 1929). After him several other authors have pointed out that in order to construct a correct Basque sentence, one has to bear in mind the focused element ('galdegaia' or 'elemento inquirido') and place it immediately before the verb (Zamarripa, 1928; Lafitte, 1944; Umandi, 1976).

The relevance of the focus position has been broadly treated in recent literature. After reviewing some of the most interesting ideas that have been postulated, we will adopt here essentially the same position that regards focus as a structural position and VP the node immediately dominating both the focus position and the verb.

2.2.1. *Previous studies*

One of the first serious attempts to provide a comprehensive description of the behaviour of the focused constituents is that of Donzeaud (1972). Donzeaud seeks to provide a focus assignment rule based on syntactic factors. The rule she proposes will then mark the position of semantic reading in the sentence which corresponds to a focused constituent in the syntactic surface structure; this rule will assign the value of focus to any constituent in the preverbal position. This proposal is inspired by Jackendoff's claim (1972:237) that focus should be assigned at two levels of derivation: At the level of semantic representation of a sentence and during the syntactic derivation at surface structure. Also Chomsky (1970:202) gives evidence for a surface structure determination of focus. Donzeaud argues that in Basque focused constituents must occur in a definite position. The assignment of the focus marker at the level of deep structure, he says, seems problematic because Basque is a scrambling language. The order of constituents cannot be determined by any PS rule. The focus assignment would be given by a surface interpretation rule.

Besides Donzeaud, de Rijk (1969, 1978) also makes some interesting remarks on the subject. De Rijk defines focus as that constituent of an utterance that corresponds to the question-word in the associated question. Here de Rijk is consciously narrowing the scope on which the assignment of focus should apply. To this extent, is emphasising what he had already said in his first paper: «there may be no focus at all in a sentence». The idea of a semantic interpretative rule to apply in surface structure, by means of which focus could only be assigned to a constituent in focus position (i.e. immediately preceding or following the verb, depending on its being affirmative or negative), was also

(6) To this extent, Michelena will agree with de Rijk's opinion (see below).

pointed out by de Rijk before Donzeaud had developed it. In his second paper, de Rijk introduces an original idea, a transformational rule of verb phrase formation⁷. In underlying structure, he argues, an NP can be marked by a syntactic feature, Focus. At some stage of the derivation, a movement rule, called Focus Positioning, would apply, placing all NPs marked +Focus in front of the verb. Therefore, a resulting constituent X, next higher node above V, dominating the focused constituent and the verb itself, could hardly be called anything else but VP. This perceptive analysis is essentially the approach which will be defended here. It has also been adopted by Michelena (1981) and Goenaga (1980). Goenaga's formulation of the two rules for Focus introduction and Focus positioning are as in (11).

(11.a.) Focus Introduction:⁸

$$\left. \begin{array}{c} \{ [+Focus], X \} \\ 1 \quad 2 \\ 0 \quad 2 \\ 1 \end{array} \right\} \rightarrow$$

(11.b.) Focus Positioning:

$$\begin{array}{c} X - NP - Y - V - Z \\ [+Focus] \\ 1 \quad 2 \quad 3 \quad 4 \quad 5 \rightarrow \\ 1 \quad 0 \quad 3 \quad 2+4 \quad 5 \end{array}$$

A slightly different assumption has been made by Azkarate *et al* (1981) and latterly by Ortiz de Urbina (1983). The approach is similar to de Rijk's approach by virtue of what they call «the nuclear verb phrase», namely, that sequence of elements beginning with the focus position and the verb itself. The main difference is that assuming VP to be generated at deep structure, by means of a base rule like (12)⁹.

$$(12) \quad \bar{V} \rightarrow \text{FOC } \bar{V}$$

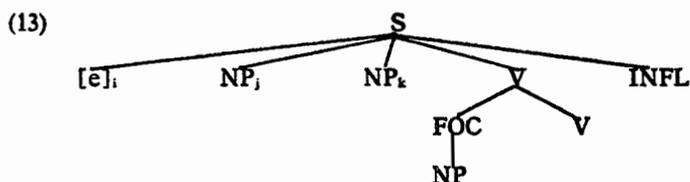
Ortiz de Urbina defends the proposal of focus as a structural position to be generated by base-rules (if a parameter of Universal Grammar, UG, has been so fixed) as an alternative to Phonetic Form (PF) stylistic rules or constraints, or Logical Form (LF) interpretation rules, as normally proposed by the Government and Binding framework (GB). He calls it the focus parameter, i.e. to assume the existence of a single bar projection of V incorporating both V and the focus

(7) A VP node created by a transformational rule is hardly the kind of things one wants transformational rules to do.

(8) Goenaga's first rule is not very appropriate. The feature «Focus» does not define a category, and hence it must be attached to something like NP, PP, or just X.

(9) We do not present Azkarate *et al*'s analysis in all its detail. There are some points in it that were later corrected by Ortiz de Urbina, whose analysis is presented here in full. This base rule is one of the things corrected. In Azkarate *et al*'s rule $V \rightarrow e V$, *e* wrongly stands for an empty structural position for focus.

position. In this proposal, focus position, like the COMP position in other languages, is considered a θ -position (or rather a A-position), to which arguments may be moved without violating the θ -criterion. Unlike COMP, however, it does not c-command the remaining elements in S. This, as the author recognizes, will create some problems for the identification of the empty category left by the application of move α , i.e. movement of a constituent into the focus position. In a configuration like (13), after one of the NPs, whether wh-word or focused element, has been moved to the empty position FOC, Ortiz de Urbina identifies the gap $[e]_i$ as an empty non-anaphoric pronominal, *pro*¹⁰. It is governed by INFL and as such it is not subject to the Empty Category Principle. Thus, even when $[e]_i$ is the empty category left by moved wh-word it will not be regarded as variable, he says, since it is not \bar{A} -bound. The co-superscripted NP does not c-command it and therefore, cannot bind it. (See below).



Ortiz de Urbina, similarly to Azkarate *et al*, presents the evidence of the focus node, involving cases of long wh-movement, when several intervening clauses separate the S-structure position of the wh-word from its D-structure as in (14) and (15).

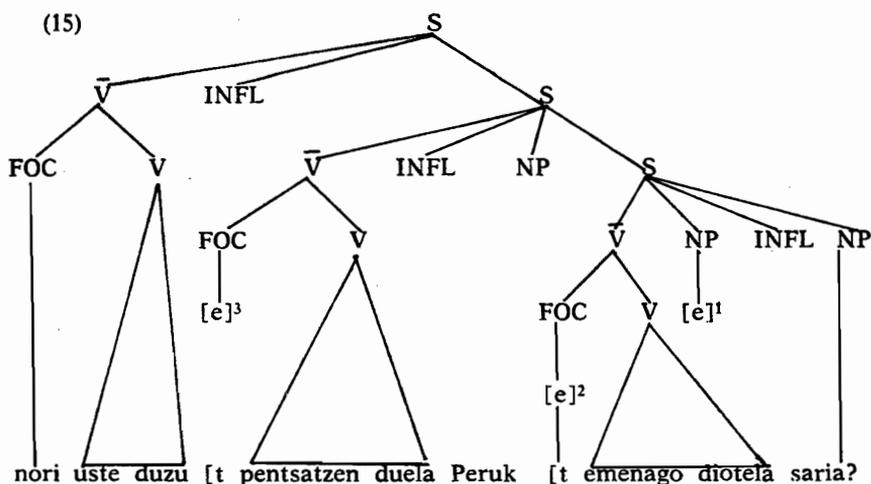
- (14) Nori uste duzu [pentsatzen duela Peruk [emango diotela saria?
 Who-DAT think Aux-you [believe Aux-he Peter-ERG [give Aux-they-
 him-it-COMP prize-ABS?
 'To whom do you think (that) Peter believes (that) they will
 give the prize?'

Regarding focus as a structural position, the hypothesis of a cyclical focus to focus position application of move α can be proposed, as in (15). Under this hypothesis, move α would involve one single bounding node S and, hence, there will be no subadjacency violation.

On the other hand, if wh-movement were direct, it would be predicted that movement of wh-word will occur only between adjacent cycles, since movement to a position two cycles up would involve a violation of subadjacency, under the natural assumption that S is a bounding node in Basque.

Besides $[e]^1$, Ortiz de Urbina accepts that the status of $[e]^2$ and $[e]^3$ present a problem as well. If contextual determination of empty category is as usually assumed, he says, $[e]^2$ and $[e]^3$ are not variables,

(10) As in Chomsky (1982), *pro* vs. *PRO*.



since they are not \bar{A} -bound. $[e]^1$ is in an argument position, and the element coindexed with them, *nori*, does not c-command them and, therefore, cannot bind them. However, $[e]^2$ and $[e]^3$ would be variables following the criterion that they be coindexed with element *sin* \bar{A} -position. Certainly, the claim that the traces of wh-movement are not variables, as semantically they presumably are, is an embarrassment for Urbina's GB approach to Basque.

The focus to focus hypothesis also predicts, Ortiz de Urbina says, that there can be no intervening focused elements between the original D-structure position of the element and its landing site. This will be so, if it is assumed that there can be no doubly filled focus position. A trace will be left after a cyclic application of move from one focus to another. This way, intervening clauses cannot have focused constituents. Ortiz de Urbina, as well as Azkarate *et al*, want to present this as an explanation for the word order restriction in clauses (those in between an extracted wh-word and its gap), in which a verb initial structure is required. But this analysis seems to imply that Basque is underlying verb-initial, otherwise non-focused NPs could occur before the verb.

The conclusion that the base rules of Basque generate a structural position governed by the verb that serves as an escape hatch for some elements, does not seem very convincing in GB terms: focus appears to be structurally in the wrong position to act as COMP-like escape hatch, according to GB's predictions.

Finally, I am going to present a different approach to the treatment of the focus position, based on Kiss' (1981) analysis for Hungarian and developed by Rebuschi (1984). Rebuschi proposes defining the focus position in terms of S, rather than in terms of V.

$$(16) \quad S' \rightarrow (Q) (F) S$$

'F' stands for focus position and 'Q' stands the landing site of auxiliary verbs in negative sentences, as well as for wh-words in general.

(17) Q → (Q) (AUX)

Quite surprisingly, for the first time in the literature, wh-words and focused constituents are treated differently. He proposes, in short, to analyse the COMP, T(topic), Q and F(ocus) slots to the left of V as positions not dominated by S, but rather as \bar{A} -positions dominated by some projection of S'. He argues that Q and F positions are dominated by S', whilst COMP and T belong to a higher projection of S, S'' (18).

(18) S'' → (T) (COMP) S'
 S' → (Q) (F) S
 S → V AUX NP* PP*

NPs and PPs will be assigned case, not by configurational means, but from their verbs and according to the inherent semantic properties of the head of their phrases. They will not be ordered at D-structure, but will be at the right of V or outside S. This proposal, Rebuschi claims, makes it possible to analyse all sentences in (1.a-f).

In contrast with Azkarate *et al* and Ortiz de Urbina's proposal, although topicalized elements can become the Topic of a higher clause, Rebuschi says, extraction of focused constituents is not allowed (19).

(19) *[F oinez] oroitzen naiz [etorri zinela
 walking remember Aux-I come Aux-you-COMP
 'walking I remember that you came'

This constraint is endorsed by the subjacency condition, i.e. no constituent can be moved out from more than one bounding category:

(20) [F] oroitzen naiz [s'' [s' [F oinez] [s etorri zinela

Sentence (19) is however perfectly grammatical and comprehensible (for my informants). After all, Rebuschi's proposal looks a bit too *ad hoc*. It fails, for example, to assign the structure of the common case when neither of the elements preceding the verb are focused, as we know that not all sentences in Basque contain focused constituents. (And as far as we know, Rebuschi has not proposed any scrambling rule to derive NPs at the left of V)¹¹. Moreover, with a rule like 'S → V AUX NP PP' the impression of Basque being a verb initial language is given,

(11) Rebuschi (1985) has reviewed this proposal: «De toute maniere, il faut renoncer au systeme plus complexe que j'ai propose dans Rebuschi (1984), selon lequel F et P [i.e. S] se c-comandent reciproquement».

Rebuschi's new proposed PSRs are as follows:

- (1) S' → (COMP) (T) S
 S → (F) V INFL NP* PP*

with the following move_α rules:

- (2) a. For non-marked sentences, V goes under F.
 b. Also, for non-marked sentences, NP*, or PP*, or X are attached to V, to create [_v X+V].
 c. For marked sentences, instead of (2) a., an NP or a PP goes under F.

which strongly contradicts the general agreement that, if anything, Basque is a verb final language.

2.2.2. Focus in terms of V

Azkarate *et al* and Ortiz de Urbina's approach is the most comprehensive. Their main disagreement with de Rijk's approach is that of not allowing a doubly filled focus position (based on their analysis of focus position as an escape hatch in Basque). De Rijk explicitly states that «the VP node dominates one or more NPs in focused position».

Again, the analysis that will be adopted here is essentially the same to that which defines focus position in terms of the verb. Because our approach deals with no transformation rule, a node for focus should be given with the general set of PS rules. The main feature that we want to capture in our analysis is the option for the focus position to be filled or not, regardless of the order in which constituents appear in the sentence. In other words, we will aim to provide a syntactic device that accounts for the fact that any constituent can be placed in a position that has been fixed as position for those elements that under certain conditions turn out to be focused.

We shall therefore expand (10) further as follows (21):

$$(21) \quad S \rightarrow NP^* VP NP^* \\ \uparrow = \downarrow$$

Where VP is rewritten as in (22):

$$(22.a.) \quad VP \rightarrow (NP) V$$

An annotation (\uparrow FOCUS) = \downarrow will be added to the NP, so that only focused NPs can fill it (22.b.):

$$(22.b.) \quad VP \rightarrow \begin{array}{c} NP \\ (\uparrow \text{ FOCUS}) = \downarrow \end{array} V$$

The NP (\uparrow FOCUS) = \downarrow will be the position for wh-words as well. This captures the fact of focused elements being defined in terms of their correspondence to a wh-word.

The problem of focus position being an escape hatch in Basque, as well as other aspects involved with our treatment of the focus position, will be discussed more fully later on (see section 3.2.).

We are now going to address the problem, yet to be resolved, of the syntactic encoding of the grammatical functions.

2.3. Syntactic encoding of grammatical functions in Basque

In a previous section we have seen that a definition of the functions subject and object in terms of syntactic configurations is not a very suitable approach for Basque, unless we assume that some syntactic

operations will apply at some stage in the derivation of surface sentences. This extreme, however, is what we would like to avoid here.

The necessity of having a VP node in the structure of Basque sentences to account for the identification of grammatical functions has been a topic of debate in the literature for the last few years. With a keen sense of intuition, de Rijk (1978) made the following comment: «In spite of the ergative morphology, we are forced to recognize the syntactic relevance of the classical notion of subject in Basque. This, however, does not imply that we also accept a base rule of the form $S \rightarrow NP + VP$. This traditional subject predicate dichotomy is much too coarse to accommodate the subtle workings of Basque syntax». This comment was later echoed by Michelena (1981) and Goenaga (1980).

Eversince, a flat structure analysis has been proposed by Azkarate *et al* (1981), Ortiz de Urbina (1983), Rebuschi (1984) and Geggus (1983). And all these proposals have been strongly criticized by Trask (forthcoming). For Trask the evidence of the syntactic relevance of the notion subject and object is an important reason to maintain the VP node.

2.3.1. Previous Analyses

Within the GB framework, Azkarate *et al* claim that an analysis (for languages such as Italian or Spanish) inspired in the pro-drop parameter would require (for Basque) that all three categories which correspond to subject, direct object and indirect object in accusative languages, were un Governed position at S-structure. This must be so because Basque allows null subjects as well as null objects, due to verb agreement, and because inversion of all these categories is possible. The R-rule hypothesis (which brings INFL down under \bar{V} node) cannot account for the extended pro-drop phenomena of Basque, they claim; it follows that INFL must be in \bar{V} and at the same time that none of the mayor categories is in \bar{V} .

The proposed set of PS rules would be:

$$(23) \quad \begin{array}{l} S \rightarrow X^{n*} \bar{V} \\ \bar{V} \rightarrow e \quad \bar{V} \\ \bar{V} \rightarrow V \quad \text{INFL} \end{array}$$

X^{n*} being a set of categories, including S, NP, PP with inherent case which must also be un Governed. INFL includes tense and inflectional categories and the thematic positions, ABS, DAT, and ERG, which are governed by V. These θ -positions are co-superscripted at D-structure with the categories in X in accordance with the binding conditions. This will allow the pro-drop parameter to be independent of the inversion phenomena or free word order. In conclusion, all arguments are un Governed but in binding relations with their anaphors (INFL) which are

governed by the verb. In this scheme, they claim, configurations play no role and, therefore, the asymmetry between subject and object restricting syntactic phenomena is not recognized in Basque.

By means of a similar strategy, Ortiz de Urbina reviews analysis. In contrast with Azkarate *et al*, he proposes INFL to govern both subject and object NPs. The same relationship that allows determination of subject-pro in pro-drop languages would be extended to all arguments with which Basque verbs agree, i.e. ERG, ABS and DAT nominals. To make this possible, Basque should be considered a non-configurational language, in other words, there will be no maximal projection of V including NPs. Under this assumption, however, some problems with θ -role and case assignment arise, he accepts. Also move α will be far too constrained, as the only movement possible will be to a focus position, the only position properly governed. Other NP movement, as ABS NP raising in some adjectival phrases (AP), will be ECP violations. Ortiz de Urbina concludes that after all, configurationality might have to be included (not only at LF level but) in the general analysis of Basque. Anyhow, the identification of grammatical functions, he claims, will not be characterized in terms of dependency, but in terms of case.

Rebuschi's (1984) main point of considering Basque a non-configurational language falls into the difficulty of proposing a configurational structure at D-level (to identify grammatical functions) when there may be no trace of it at S-structure level. Transformations will encounter difficult problems according to Chomsky's (1981) constraints on transformation.

Given the base rule:

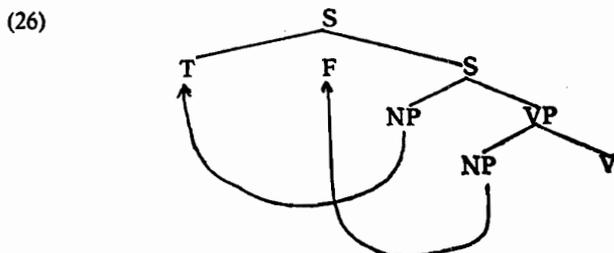
$$(24) \quad \begin{array}{l} S \rightarrow NP VP \\ VP \rightarrow NP V \end{array}$$

to generate the underlying constituent order at D-structure, a surface sentence, displaying object-subject-verb order, as in (1.e.), will require some sort of movement rule. It will not be possible to claim a simple permutation of constituent, if they are understood to be A-positions, because of a violation of the θ -criterion. Alternatively, it could be posited, that at S-structure the moved constituents do not occupy A-positions. But again, this would imply, Rebuschi says, that the positions the constituents occupy at S-structure never correspond to their positions at D-structure, which violates the projection principle.

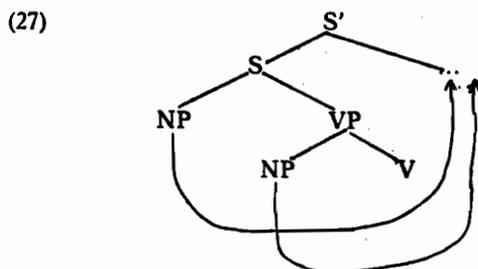
A second possibility, more in line with what generative grammar understands by \bar{A} -positions, would be to propose two A-positions, T and F (standing for Topic and Focus) outside S. This would expand (24) as in (25):

$$(25) \quad S' \rightarrow T F S$$

In this case, he says, a double operation of movement would be needed to account for a sentence with subject-object-verb order:



and consequently, no kernel sentence (Chomsky, 1957) will be ever possible, since two transformations have to apply as soon as S contains two arguments¹². On the other hand, further transformations would be needed to account for extrapositions out of S in verb initial sentences:



In conclusion, he says that to propose a configurational structure at D-level is not sensible if no trace of it is left at surface level. Therefore, to account for grammatical functions in terms of configurations is not desirable, and, in this sense, it would be better to consider Basque as a non-configurational language.

The claim that Basque is a configurational language has been put forward by Trask (forthcoming) in the form of two arguments. Verbs subcategorize for objects, but not for subjects. Moreover, subjects are well distinguished from objects by their inability to be neither reflexive nor reciprocal. They also control a wide range of Equi-deletions. Anyhow, Trask does not go into a detailed explanation of how all this will work in a formalized system. There is no explanation yet of how such a system will identify the function subject and at the same time allow free order of maximal projections.

It is our impression that a configurational analysis could be well argued on the basis of most recent GB research, particularly Williams

(12) The 1957 model used a concept of «kernel sentence», that is, a sentence that was derived using only obligatory transformations, and no optional ones. So the «kernel» consisted of simple, active, declaratives (no passives, negatives, questions, etc.). Since 1965, however, this division of obligatory/optional rules has altered, and the concept of the kernel no longer has any relevance. So this point of Rebuschi's really does not have any force.

(1977-1984). A base universal rule such as $S \rightarrow NP VP$ could be proposed at D- and S-structure level. To account for the different constituent orders it would be just enough to resolve some operations by means of scrambling rules, which would generate a non-configurational structure at PF. VP would be required to be transparent to scrambling, as proposed for non-configurational languages in general (1984:640-650). S-structure would always have a NP VP split, and alternative orders will occur at PF after scrambling rules apply.

2.3.2. *LFG's approach*

When we assume a LFG type model we gain, however, an artifact that provides us with an autonomous level of representing grammatical functions distinct from the configurational approach. In LFG grammatical functions would be accounted for not in the c-structure of the sentence, but in an independent f-structure. We regard grammatical functions as syntactic primitives, which are not universally definable in terms of syntactic configurations, case features, semantic roles or independently motivated logical operations (Dowty, 1982), cf. Mohanan (1983).

We take the syntactic encoding of the grammatical functions to be a part of the mapping between predicates and arguments; in the case of Basque, marking the arguments with case inflections and, redundantly, by agreement features in the predicate. As is known, Basque verbs can agree with their three main arguments, subject, object and indirect object. Therefore, we postulate, the encoding of grammatical functions not to be carried out in terms of syntactic configurations, but in terms of devices such as case and agreement.

In this way, we avoid all uses of transformational derivation. To give a quick account of why this is so convenient, we present a brief summary of some of the reasons that have been put forward by Bresnan *et al* (1982) to avoid transformations.

In transformational theories, the mapping between a predicate argument structure and its surface form is performed via PS computation. An ordered set of PS computation must be instantiated in the process of decoding the grammatical relations of a predicate argument structure. In a process of parsing this computation will apply first on the representation of the surface form (not S-structure, but rather PF in GB) of that predicate argument structure. Remember that grammatical relations are encoded by means of the so-called 'feeding relations' (those relations in which the output provided by each operation creates the necessary input for the next operation in a 'cascade'). An initial difficulty can be understood when we consider the complexity of the decoding. The complexity of the decoding process is a direct function, it is claimed, of the length of the transformational derivation (a hypothesis known as the 'derivational theory of complexity', which is a rather

debatable hypothesis). Moreover, if we consider Basque to be non-configurational at PF, the mapping to its S-structure would further involve operations of scrambling. If all scrambling rules applied simultaneously, there would be no true feeding relations, which, as said, encode the grammatical relations (and grammatical functions among them).

Therefore, avoiding transformational derivations, we do not have to provide any ordered sequence of scrambling rule application at PF level. It is not clear at all how the application of scrambling rules could be arranged, anyway.

Following Kaplan and Bresnan (1982:297-299), the basic principle of non-configurational encoding is to associate pairs of functions-assigning and feature assigning rules of the form:

$$(28) \quad \begin{array}{l} (\downarrow F) = v \\ (\uparrow G) = \downarrow \end{array}$$

Where G is the function selected by the value v of the feature F .

For example, taking F to be CASE and v to be ERG, $(\downarrow \text{CASE}) = \text{ERG}$ could be associated with $(\uparrow \text{SUBJ}) = \downarrow$. These pairs of schemata are arbitrarily associated with categories in c-structure rules, as will be shown below.

In addition to (8), a schema $\uparrow = \downarrow$ is also associated arbitrarily with categories in c-structure rules. (The association with $\uparrow = \downarrow$ can also depend on some inflectional features). The head of the sentence is defined to be any major category which is annotated with $\uparrow = \downarrow$ and which has a PRED. It follows from the consistency condition that the head is unique. Every phrase has a unique head, and the features of a phrase are identified with those of its head. However, in non-configurational encoding, functions need not be assigned to maximal projections; instead they may be assigned to submaximal projections, or even to single lexical categories (this may be very useful for Basque, since cases go at the end of a phrase and do not have to be attached to its head). The condition of wellformedness of f-structures eliminates 'incorrect' assignments, admitting only those that yield consistent (or unique), coherent and completed f-structures. These three conditions on functional well formedness can be defined as follows:

- i) Uniqueness: A grammatical relation's particular attribute may have at most one value in a given f-structure.
- ii) Completeness: A f-structure must contain all the governable grammatical functions governed by its predicate.
- iii) Coherence: All the governable grammatical functions that a f-structure contains must be governed by its predicates.

Configurational encoding of grammatical functions in Basque would be possible if the functions were identified by the category and by the order of maximal constituents within the immediately dominating phrase.

However, non-configurational encoding seems more natural, as in Basque functions are identified, we claim, by their case and agreement features of possible submaximal constituents belonging to an unordered set of maximal categories. (Only the case marking of constituents may appear to be subcategorized for, and not their PS configurations).

For Basque, the principles of interpreting NPs as subject, object and indirect objects can be given as follows:

- i) Ergative case as the subject of transitive verb (SUBJ).
- ii) Absolutive case as the subject of intransitive verb (SUBJ).
- iii) Absolutive case as the object of transitive verb (OBJ).
- iv) Dative case as the indirect object (OBJ2).

These principles can be formally encoded in terms of the following functional equations (29):

- (29) i) $(\uparrow \text{SUBJ}) = \downarrow$
 $(\downarrow \text{CASE}) = \text{ERG}$
 ii) $(\uparrow \text{SUBJ}) = \downarrow$
 $(\downarrow \text{CASE}) = \text{ABS}$
 iii) $(\uparrow \text{OBJ}) = \downarrow$
 $(\downarrow \text{CASE}) = \text{ABS}$
 iv) $(\uparrow \text{OBJ2}) = \downarrow$
 $(\downarrow \text{CASE}) = \text{DAT}$

Any of these sets can be assigned to any of the NP* constituents dominated by S in (21).

We can also give a set of redundancy rules to be added to the V constituent to account for the agreement phenomena of Basque (30).

- (30) $(\uparrow \text{SUBJ CASE}) = \text{ERG}$
 $(\uparrow \text{SUBJ CASE}) = \text{ABS}$
 $(\uparrow \text{OBJ CASE}) = \text{ABS}$
 $(\uparrow \text{OBJ2 CASE}) = \text{DAT}$

These equations will be added in the following way: If a verb's lexical form contains (OBJ), the equations $(\uparrow \text{SUBJ CASE}) = \text{ERG}$ and $(\uparrow \text{OBJ CASE}) = \text{ABS}$ will be added. If no (OBJ) is contained, then $(\uparrow \text{SUBJ CASE}) = \text{ABS}$ will be added instead. Whenever (OBJ2) appears, an equation like $(\uparrow \text{OBJ2 CASE}) = \text{DAT}$ will also be added.

The addition of these equations to the V node in the c-structure guarantees that the subject case of a transitive verb cannot be anything other than ERG, for if the subject case bearing element has any inflectional case other than ERG, the promotion of this feature will result in a clash of case values and an ill-formed f-structure. This has the advantage of allowing constituents bearing grammatical functions (SUBJ, OBJ, OBJ2) to be freely absent, as is often the case in Basque (remember the pro-drop parameter proposed in the literature). When an optional constituent bearing a nuclear grammatical function is omitted from c-structure but required by a lexical form, the grammatical function is interpreted as an anaphor (see section 3.1.).

The uniqueness, completeness and coherence conditions prevent other f-structure solutions. It is worth noticing that a different order of constituents in (33) would not have yielded a different f-structure in (34) since the syntactic encoding of grammatical functions has been achieved by means of case and agreement features, and not in terms of syntactic configurations like precedence or dominance.

REFERENCES

- ABAITUA, J. (in preparation), *Lexical Rules for Basque*.
 ———, & TRASK, R. L. (forthcoming), «Accusativity in Basque».
 AKMAJIAN, A., STEELE, S. & WASOW, T. (1979). «The Category AUX in Universal Grammar», *Linguistic Inquiry* 10, 1-64.
 ALTUBE, S. (1929; reprinted 1975). *Erderismos*. Bilbao: Euskaltzaindia.
 ANDERSON, J. M. (1977). *On case grammar: prolegomena to theory of grammatical relations*. London: Croom Helm.
 ANDERSON, S. (1976). «On the notion of subject in Ergative languages», in C.N. Li (ed.), 1-26.
 ANDREWS, A. D. (1982). «Case in Modern Icelandic». In Bresnan (ed.), (1982), 427-503.
 AZKARATE, M., FARWELL, D., ORTIZ DE URBINA, J. & SALTARELLI, M. (1981). «Word order and Wh-movement Basque». In *Proceedings of the 12th Annual Meeting of the North Eastern Linguistics Society*. Cambridge Mass.
 BAKER, M. (1983). «Object, Thems, and Lexical Rules in Italian». In L. Levin, M. Rappaport and A. Zaenen (eds.), (1983), 1-45, *Papers in Lexical Functional Grammar*. Indiana, Bloomington: Indiana University Linguistics Club.
 BOSSONG, G. (1985). «Ergativity in Basque». *Linguistics* 22, 341-392.
 BRESNAN, J. (1976). «The form and function of transformations». *Linguistic Inquiry* 7, 3-40.
 ———, (ed.) (1982). *The Mental Representation of Grammatical Relations*. MIT Press.
 BRETTSCHNEIDER, G. (1979). «Typological Characteristics of Basque». In F. Plank (ed.) (1979), 371-384.
 BURZIO, L. (1981). *Intransitive Verbs and Italian Auxiliaries*, Ph. D. dissertation, MIT.
 CHOMSKY, N. (1957). *Syntactic Structures*. The Hague, Mouton.
 ———, (1970). «Deep Structure, Surface Structure, and Semantic Interpretation». In D. D. Steinberg & L. A. Jakobovits. (1971), 183-216. *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*. Cambridge: University Press.
 ———, (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
 ———, (1982). *Some concepts and consequences of the theory of Government and Binding*. MIT Press.
 CONTRERAS, H. (1976). *A Theory of Word Order*. Amsterdam: North-Holland.
 COMRIE, B. (1978). «Ergativity». In W. P. Lehman (ed.), (1978), 329-394, *Syntactic Typology*. Sussex: The Harvest Press.
 ———, (1983). «Universals: form and functions». *Linguistics* 21.
 COOREMAN, A., FOX, B., & GIVON, T. (1984). «The Discourse Definition of Ergativity». *Studies in Language* 8.
 DIXON, R. M. W. (1979). «Ergativity». *Language* 55, 59-138.
 DONZEAUD, F. (1972). «The Expression of Focus in Basque». *ASJU* VI, 35-45.

- DOWTY, D. (1982). «Grammatical Relations and Montague Grammar». In P. Jacobson and G. K. Pullum (eds.), 79-103.
- ENTWISTLE, W. J. (1936). «On the Passivity of the Basque Verb.» *Medium Aevum* 5, 104-114.
- EUSKALTZAINDIA. (1973). «Aditz Laguntzaile Batua». (Separata). *Euskera* XVII. Bilbao.
- , (1979). «Bergarako biltzar ondoko erabakiak». *Euskera*. Bilbao.
- FALK, Y. N. (1983). «Subjects and Long-Distance Dependencies». *Linguistic Inquiry* 12, 245-270.
- , (1984). «The English Auxiliary System. A Lexical Functional Analysis». *Language* 60, 483-509.
- FREY, W. & REYLE, U. (1983). «A Prolog implementation of LFG as a base for a natural language processing system». *Proceedings of the first Conference of the European Chapter of the Association for Computational Linguistics* 52-57.
- GAZDAR, G. (1982). «Phrase Structure Grammar». In P. Jacobson & G. K. Pullum (eds.), 131-186.
- GAVEL, H. (1930). «Quelques observations sur la passivité du verbe basque». *RIEV* 21, 1-14.
- GEGGUS, J. (1983). «Multiple Agreement: The Case of Basque». Manuscript.
- GOENAGA, P. (1980). *Gramatika Bideetan?*. San Sebastián: Erein.
- , (1984). *Euskal Sintaxia: Konplementazioa eta Nominalizazioa*. Ph. D. Dissertation. Vitoria-Gasteiz: Euskal Herriko Unibertsitatea.
- GREEN, C. (1969). «Application of Theorem Proving to Problem Solving». *IJCAI* 1.
- , (1976). «How free is word order in Spanish». In M. Harris (ed.). *Romance Syntax: Synchronic and Diachronic Perspectives*. Salford: University.
- HEATH, J. (1977). «Remarks on Basque Verbal Morphology». In W. A. DOUGLAS, R. W. ETULAIN & W. H. JACOBSEN (eds.), *Anglo-American Contributions to Basque Studies. Essays in Honor of Jon Bilbao*. Desert Research: Institute Publications in Social Sciences 13, 193-201.
- HOEKSTRA, T. (1984). *Transitivity: Grammatical Relations in Government and Binding Theory*. Dordrecht: Foris.
- HORROCKS, G. (1983). «The Order of Constituents in Modern Greek». In G. Gazdar, E. Klein & G. K. Pullum (eds.), *Order, Concord and Constituency*. 95-112. Dordrecht: Foris.
- JACKENDOFF, R. S. (1972). *Semantic Interpretation in Generative Grammar*. MIT Press.
- , (1977). *X Syntax: A Study of Phrase Structure (Syntax)* Linguistic Inquiry Monograph 2.
- JACOBSON, P. & PULLUM (eds.), (1982), *The Nature of Syntactic Representations*. Dordrecht.
- JELINEK, E. (1984). «Empty Categories, Case and Configurationality». *NLLT* 2, 39-76.
- KAPLAN, R. M. (1975). «On process models for sentence comprehension». In D. Norman & D. Rumelhart, (eds.). *Explorations in Cognition*. San Francisco: Freeman.
- , & BRESNAN, J. (1982). «Lexical Functional Grammar: A Formal System of Grammatical Representation». In J. Bresnan (ed.), 174-281.
- KEENAN, E. L. (1976). «Towards a Universal Definition of 'Subject'». In C. N. Li (ed.), 303-334.
- KISS, K. (1981). «Structural Relations in Hungarian, a Free Word Order Language». *Linguistic Inquiry* 12, 185-213.
- KOWALSKY, R. A. & KUEHNER, D. (1971). «Linear Resolution with selection function». *Artificial Intelligence* 2, 227-260.

- LAFITTE, P. (1931). «Pour ou contre la passivité du verbe basque?». *Gure Herria*, mai-juin, 263.
- , ([1944], 1979). *Grammaire basque: Dialecte navarro-labourdin littéraire*. San Sebastian: Elkar.
- LANGACKER, R. W. (1969). «On Pronominalization and the Chain of Command». In D. A. Reibel & Schane (eds.), 160-186. *Moder Studies in English*. Prentice Hall, Englewood Cliffs, N. J.
- LEVIN, B. (1983). «Unaccusative Verbs in Basque». Paper presented at NELS, 13.
- LEVIN, L., RAPPAPORT, M. & ZAENEN, A. (eds.), (1983). *Papers in Lexical Functional Grammar*. Indiana Bloomington: Indiana University Linguistic Club.
- LI, C. N. (ed.), 1976. *Subject and Topic*, New York: Academy Press.
- MARTINET, A. (1958). «La construction ergative». *Journal de Psychologie Normal et Pathologique* (juillet-septembre).
- , (1962). «Le sujet comme fonction linguistique et l'analyse syntaxique du basque». *Bulletin de la Societe Linguistique* 57, 73-82. Paris.
- MICHELENA, L. (1977). «Notas sobre compuestos verbales vascos». *Revista de Dialectologia y Tradiciones Populares XXXIII*, 245-271.
- , (1978). «Miscelánea filológica: relato y orden de palabras». *FLV* 29, 204-235.
- , (1981). «Galdegaia eta Mintzagaia euskaraz». In *Euskal Linguistika eta Literatura: Bide Berriak*. 57-81. Bilbao: Universidad de Deusto.
- MOHANAN, K. (1982). «Grammatical Relations and Clause Structure». In Bresnan (ed.), (1982), 504-589.
- , (1983). «Move NP or Lexical Rules? Evidence from Malayalam Causativization». In L. Levin et al, (eds.), (1983), 47-111.
- NAERT, P. (1956). «Le verbe basque est il passif?». *Studia Linguistica* 10, 45-49.
- ORTIZ DE URBINA, J. M. (1983). «Empty Categories and Focus in Basque». *Studies in Linguistic Science* 13, 1.
- PEREIRA, F. C. N. (1981). «Extrapolation Grammars». *American Journal of Computational Linguistics* 7, 4, 243-256.
- , & WARREN, D. H. D. (1980). «Definite Clause Grammar for Language Analysis: A Survey of the Formalism and a Comparison with Augmented Transition Networks». *Artificial Intelligence* 13, 213-278.
- , (1983). «Parsing as a Deduction». *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, 137-144. Boston, Massachusetts.
- PERLMUTTER, D. M. & POSTAL, P. M. (1974). *Lectures on Relational Grammar*. Summer Linguistic Institute of Linguistic Society of America. Amherst: University of Massachusetts.
- PERLMUTTER, D. M. (1982). «Syntactic Representation, Syntactic Levels, and the notion of Subject». In Jacobson et al, (eds.), (1982), 283-340.
- PINKER, S. (1982). «Lexical Interpretative Grammars». In Bresnan (ed.), (1982), 655-726.
- PLANK, F. (ed.), (1979). *Ergativity: Towards a Theory of Grammatical Relations*. New York: Academic Press.
- REBUSCHI, G. (1978). «Cas et fonction sujet en basque». *Verbum: Revue de linguistique publiee par l'Universite de Nancy* II, 1, 69-98.
- , (1983). «A note on focalization in Basque». *Journal of Basque Studies* 4, 2, 29-42. Indiana: University of Pennsylvania and Society of Basque Studies in America.
- , (1984). «Positions, configurations et classes syntaxiques. Aspects de la construction de la phrase basque». *Communication-Congres «Arturo Campion»*. Pamplona-Irunea: Euskaltzaindia.
- , (1985). «Theorie du liage et langues nonconfigurationnelles: quelques donnees du basque navarro-labourdin». Université de Nancy.
- RIJK, R. DE (1969). «Is Basque an S.O.V. Language?». *FLV* 3, 319-351.
- , (1978). «Topic Fronting, Focus Positioning, and the Nature of the Verb

- Phrase in Basque». In F. Jansen (ed.), *Studies on Fronting*, 81-112. Lisse: Peter de Ridder Press.
- ROBINSON, J. A. (1965). «A machine-oriented logic based on the resolution principle». *Journal of the Association of Computing Machinery* 12, 23-41.
- ROTAETXE, R. (1978). «Lingüística-lógica: la construcción ergativa vasca». *Revista Española de Lingüística* 8, 431-445.
- , (1980). «Ergatiboaren sujetotasunaz». *Euskera* XXV, 417-424.
- SCHACHTER, P. (1976). «The Subject in Philippine Languages: Topic, Actor, Actor-Topic, or none of the Above?». In C. N. Li (ed.).
- , (1977). «Reference-Related Properties of Subjects». In Cole and Sadock (eds.). *Grammatical Relations*. New York: Academic Press.
- SHIEBER S. M. (1985). «An Introduction to Unification-Based Approaches to Grammar». Paper presented as a Tutorial Session at the 23rd Annual Meeting of the Association for Computational Linguistics. University of Chicago, Illinois.
- SLOBIN, D. I. (1982). «Universal and Particular in the Acquisition of Language». In L. Gleitman & E. Wanner (eds.). *Language Acquisition: The State of the Art*. Cambridge: University Press.
- STEELE, S. (1978). «The Category AUX as a Language Universal». In J. H. Greenberg (ed.), 7-45. *Universals of Human Language, Vol. 3: Word Structure*. Stanford: University Press.
- , (ed.) (1981). *An Encyclopedia of AUX*. Cambridge: MIT Press.
- TCHEKHOFF, C. (1978). *Aux fondements de la syntax: l'ergatif*. Paris: Presses Universitaires de France.
- TRASK, R. L. (1981). «Basque Verbal Morphology». In *Euskalarien Nazioarteko Jardunaldiak. Iker 1*. 285-304. Euskaltzaindia (ed.). Bilbao.
- , (1984). *Synchronic and Diachronic Studies in the Grammar of Basque*. Ph. D. Thesis. University of London.
- , (forthcoming). «On the Representation of NP gaps».
- UMANDI (pseudonym of A. URRESTARAZU). (1976). *Gramática Vasca*. Tolosa: Kardaberaz.
- VAN VALIN, R. D. (1977). «Ergativity and the Universality of Subjects». *Papers from the Regional Meeting of the Chicago Linguistic Society* 13, 689-705.
- VILLASANTE, L. (1980). *Sintaxis de la oración simple*. Oñate: Aránzazu.
- WILBUR, T. H. (1970). «Ergative and Pseudo-Ergative in Basque». *FLV* 4, 57-66.
- , (1979). *Prolegomena to a Grammar of Basque*. Amsterdam: John Benjamins B.V.
- WILLIAMS, E. (1984). «Grammatical Relations». *Linguistic Inquiry* 15, 4, 639-673.
- WHITELOCK, P., Johson, R. & BENNETT, P. (1984). *The UMIST Translation System. CCL Report*. University of Manchester Institute of Science and Technology.
- YASUKAWA, H. (1984). «LFG system in Prolog». *Proceedings of the 22nd Annual Meeting of the Association for Computational Linguistics*, 358-361. Stanford University, California.
- ZAENEN, A. (1983). «On Syntactic Binding». *Linguistic Inquiry* 14, 3, 469-504.
- ZAMARRIPA, P. (1928). *Gramática Vasca*. Bermeo: Gaubeka.