Supplements of the *Anuario de Filología Vasca «Julio de Urquijo»*, n.º XLVI

INQUIRIES INTO THE LEXICON-SYNTAX RELATIONS IN BASQUE

Bernard Oyharçabal
(ed.)

University of Basque Country Gipuzkoako Foru Aldundia
Bilbao Donostia-San Sebastián
2003

ASJU-REN GEHIGARRIAK
ANEJOS *ASJU*
SUPPLEMENTS OF ASJU

V. NILS M. HOLMER, *El idioma vasco hablado. Un estudio de dialectología vasca,* 1964, 1991. 12 €.

VI. LUIS VILLASANTE, *Fr. Pedro A. de Añíbarro. Gramática vascongada,* 1970. 8 €.

VII. CÁNDIDO IZAGUIRRE, *El vocabulario vasco de Aránzazu-Oñate y zonas colindantes,* 1971. Segunda edición con índice inverso de Gerardo Markuleta, 1994. 12 €.

VIII. *Papers from the Basque Linguistics Seminar. University of Nevada. Summer* 1972, 1974. 10 €.

IX. JULIEN VINSON, *Essai d'une bibliographie de la langue basque. Con las anotaciones del ejemplar de Julio de Urquijo,* 1984. 20 €.

X. LUIS MICHELENA, *Sobre historia de la lengua vasca,* 1988. Out of print.

XI. LUIS MICHELENA - IBON SARASOLA, *Textos arcaicos vascos. Contribución al estudio y edición de textos antiguos vascos,* 1989. 12 €.

XII. HUGO SCHUCHARDT, *Introducción a las obras de Leizarraga. Sobre el modo de disponer la reimpresión, en particular sobre las erratas y variantes en el texto de Leizarraga,* 1989. 8 €.

XIII. MANUEL AGUD - † ANTONIO TOVAR, *Diccionario etimológico vasco,* I. *A-Ardui,* 1989, 1993. Out of print.

XIV. JOSEBA A. LAKARRA (ed.), *Memoriae L. Mitxelena magistri sacrum,* 1991. 36 €.

XV. RICARDO GÓMEZ - JOSEBA A. LAKARRA (arg.), *Euskalaritzaren historiaz I: XVI-XIX. mendeak,* 1992. 27 €.

XVI. BEÑAT OYHARÇABAL, *La pastorale souletine: édition critique de «Charlemagne»,* 1990. 27 €.

XVII. RICARDO GÓMEZ - JOSEBA A. LAKARRA (arg.), *Euskalaritzaren historiaz II: XIX-XX. mendeak.* In preparation.

XVIII. JOSEBA A. LAKARRA, *Harrieten Gramatikako hiztegiak* (1741), 1994. 8 €.

XIX. MANUEL AGUD - † ANTONIO TOVAR, *Diccionario etimológico vasco, II. Ardun-Beuden,* 1990. Out of print.

XX. LUIS MICHELENA, *Lenguas y protolenguas,* 1963, 1986, 1990. 8 €.

XXI. ARÈNE GARAMENDI, *El teatro popular vasco. (Semiótica de la representación),* 1991. 12 €.

XXII. LASZLÓ K. MARÁCZ, *Asymmetries in Hungarian.* 1991. 15 €.

XXIII. PETER BAKKER, GIDOR BILBAO, NICOLAAS G. H. DEEN, JOSÉ I. HUALDE, *Basque pidgins in Iceland and Canada,* 1991. 10 €.

XXIV. MANUEL AGUD - † ANTONIO TOVAR, *Diccionario etimológico vasco,* III. *Beule-Egileor, Babarraso-Bazur,* 1991. Out of print.

Inquiries into the Lexicon-Syntax Relations in Basque

Supplements of the *Anuario de Filología Vasca «Julio de Urquijo»*, n.º XLVI

INQUIRIES INTO THE LEXICON-SYNTAX RELATIONS IN BASQUE

Bernard Oyharçabal
(ed.)

eman ta zabal zazu

Universidad del País Vasco
servicio editorial

Euskal Herriko Unibertsitatea
argitalpen zerbitzua

Gipuzkoako Foru Aldundia
Diputación Foral de Gipuzkoa

The studies in this book originates from an International Scientific Cooperation Program called *Lexical Syntax in Basque* (1999-2001), and directed by M. Azkarate (EHU-University of the Basque Country) and B. Oyharçabal (IKER-UMR 5478). This program was funded by the CNRS in France and the Research Department of the Basque Autonomous Government's Ministry of Education and Research.

ISBN: 84-8373-580-6
Depósito legal/Lege gordailua: BI - 176-04

Fotocomposición/Fotokonposizioa: Ipar, S. Coop.
Particular de Zurbaran, 2-4 - 48007 Bilbao

INDEX

FOREWORD

The studies in this book originated from an international program called *Lexical Syntax in Basque* (1999-2001), which was funded by the Department of Humanities and Social Sciences of the CNRS (UMR 5478) and the Basque Autonomous Community's Ministry of Education and Research. As its tittle indicates, this program was oriented toward the study of relations between lexicon and syntax in Basque. The program assembled approximately about fifteen Basque researchers, both computer scientists and linguists, most of whom were associated with the University of the Basque Country. Half the papers consists of studies conducted in computational linguistics, and half result from more classical morphosyntactic studies.

The five articles written by computational researchers reflect a collective undertaking which has been developed in recent years by members of the IXA group in San Sebastian. One of this research group's primary goals is to build automatic analyzers of texts. The papers presented in this book, which are mainly devoted to verbal subcategorization, provide an insight into the manner in which this significant aspect of lexical syntax is dealt with in the field of corpus linguistics.

Two syntactic analyzers have been created for the parsing of Basque corpora. One (called PATRIXA) uses a unification-based formalism, and it is dealt here by Aldezabal, Aranzabe, Atutxa, Gojenola and Sarasola, in the context of the automatic analyzis of verbs' subcategorization; the other one is based on a Constraint Grammar (Karlsson 1995). The foundation of the latter is presented in this book by Aduriz & Diaz de Ilarraza in a paper devoted to the way morphosyntactic ambiguities in texts may be analyzed.

Furthermore, two other applications are presented which pertain to the extraction of information about verbal subcategorization. Arriola, Artola and Soroa show how dictionaries can be used as a starting point to guide lexical acquisition from corpora, while Aldezabal, Aranzabe, Atutxa, Gojenola and Sarasola examine how lexical information based on the argument *vs* adjunct distinction can be extracted from raw corpora, after the application of a partial parser and statistical filters. In the fifth paper, Aldezabal and Goenaga discuss several general problems regarding the way verbal subcategorization can be dealt with in computational linguistics, and they present the results obtained in their analysis of subcategorization patterns for one hundred verbs.

As indicated before, the other papers have a more classical flavour. The paper by Alberdi consists of a careful examination of the use of ergative case in Basque borrowed verbs, especially with non transitive verbs. He shows that Spanish unergative verbs

which are formally reflexives (Mendikoetxea 1999) are treated as absolutive intransitive verbs in Basque borrowings, and he demonstrates that in fact, the tendency to use an ergative case marking with intransitive borrowed verbs (Sarasola 1979) is restricted to non-reflexive (most of the time agentive) Romance verbs. The remainder of the inquiries deal with verb derivation. Artiagoitia's paper analyzes the case of *-garri,* a deverbal suffix which has been traditionally considered as having two opposite values (an active one and a passive one). Artiagoitia offers a unified account of the suffix, and dismisses the notion that *-garri* masks two homophones. Because the so-called active value of *-garri* appears with object-experiencer psych-verbs, he proposes an analysis which frames these verbs as lacking an external argument and having a derived subject (Belletti & Rizzi 1989). Such an analysis is not compatible with the one offered by Oyharçabal in his paper devoted to verbs having the causative alternation. In Oyharçabal's view, alternating causative verbs (including psych-causatives of the *frighten / preoccupare* type) project a *Cause* head in syntax (Pylkkänen 2002). Based on lexical decomposition, his papers examines the different types of lexical causatives in Basque, and concludes they can be characterized as verbs that express a change in the physical constitution, the location or the psychological state of the theme, patient or experiencer. Zabala offers a broader analysis of causation in her work. She analyzed this notion along with other semantic concepts such as agentivity, control and intentionality. Starting from data obtained in technical texts, she suggests checking predictions that concern semantics and morphosyntax of certain kinds of verbs, including deverbal derived nouns and adjectives. The study of Odriozola is devoted to a systematic classification of derived verbs: denominal, deadjectival and depostpositional (or deadverbial) verbs are analyzed, mainly using semantic criteria in each case.

In 2001, as the *Lexical Syntax in Basque* program was finishing, Professor Miren Azkarate, who was the head of the group of researchers from the University of Basque Country in this project, was called upon to assume new responsabilities and was no longer in position to contribute to this volume. Although she was unable to participate in this edition, my colleagues and myself would like to thank her for her commitment to in the completion of this project. We would also like to thank Joseba Lakarra for the help he has provided to us in the preparation of this volume.

In closing, it was during the correction of the proofs of the papers that we learned that Rudolf de Rijk had passed away. We knew him to be gravely ill, and like other Basque linguists, we were greatly saddened at this news. Basque syntactic studies owe so much to his work. We dedicate this book to his memory.

B. Oyharçabal

MORPHOSYNTACTIC DISAMBIGUATION AND SHALLOW PARSING IN COMPUTATIONAL PROCESSING OF BASQUE

Itziar Aduriz and Arantza Díaz de Ilarraza

Abstract

Our goal in this article is to show the improvements in the computational treatment of Basque, and more specifically, in the areas of morphosyntactic disambiguation and shallow parsing.[1] *The improvements presented in this paper include the following: analyses of previously identified ambiguities in morphosyntax and in syntactic functions, their disambiguation, and finally, an outline of possible steps in terms of shallow parsing based on the results provided by the disambiguation process. The work is part of the current research within the field of Natural Language Processing (NLP) in Basque, and more specifically, part of the work that is being done within the IXA*[2] *group.*

1. Introduction

Morphosyntactic ambiguity is, as is well-known, one of the most difficult problems in NLP. Ambiguity arises from previously done morphological analyses, and hence, it is closely dependent on decisions made at the morphological level. This paper includes a presentation of the systematic analysis of ambiguity in Basque. Also, we briefly describe the morphological analyzer in order to understand how ambiguity arises in Basque, namely from the lexical Basque Database (i.e., Euskararen Datu-base lexikala (EDBL)), which is the basis for all information.

Analyses of morphosyntactic ambiguity are followed by the process of disambiguation of such analyses, which is a crucial step in NLP as mentioned above. For this task, we have created a grammar composed of over one thousand rules by following linguistic criteria. We also present the formalism chosen to carry out this step, namely, Constraint Grammar.

When treating language computationally, syntactic analyses necessarily follow morphological analyses and their disambiguation. Linguistic research has shown that any thorough syntactic analysis on languages encounters difficulties, even when adopting a specific theoretical framework. Matters get much worse in a computational

---

[1] Shallow Parsing is also known as Partial Syntax. The aim of this type of analysis is to analyze all the structures in a corpus, albeit not in depth.

[2] http://ixa.si.ehu.es

approach to language, and as a solution, several pre-syntactic analyses (or steps parallel to the syntactic analyses) have been proposed. This is the case of the so-called *partial analyses* and the *analysis of parts of sentences.*

2. Analyses of sintagmatic components: disambiguating-grammars and syntactic analyzers

The main difference between morphological and syntactic analyses relates to a change in viewpoint. Whereas a morphological analysis involves a paradigmatic viewpoint, those analyses that come after the morphological level (i.e., those directed to syntax and syntactic analyses themselves) involve a sintagmatic perspective. The former are limited to the word-level, and in contrast, the later target bigger units formed by words: chunks, phrases, clauses, etc. Thus, the onset of the sintagmatic analysis directly follows the morphological analysis. In fact, it is when we morphologically disambiguate a word that we start to consider its context, and even more clearly, when we reach the step of the syntactic analysis, which analyzes the relations between sintagmatic components.

There are various approaches to morphosyntactic disambiguation and to syntactic analysis, and variations depend on the criterion that is selected in each case. These criteria include the type of information we take as basis, or the result that we may want to obtain. In short, differences among various approaches depend on the viewpoint that is selected for approaching syntax. For our purposes, and in order to understand the formalism we have selected for the case of Basque, we will only mention the most prominent trends and their features.[3]

There are three prominent tendencies in disambiguating-grammars[4] and in syntactic analyzers: those based on linguistic descriptions, those based on statistical techniques, and finally, hybrid methods, which combine both.

2.1. Linguistic descriptions as basis

A. In disambiguating-grammars: these grammars are based on rules that are created by using linguistic knowledge, and are called knowledge-driven taggers. It is commonplace to create the rules manually, rendering them both very precise and costly. The predecessor for this tagging system is the tagger called TAGGIT created to tag the Brown Corpus in the 70's (Greene & Rubin 1971). The most widespread successor of this system, namely Karlsson's *Constraint Grammar* (henceforth CG), is the grammar we selected for morphosyntactic disambiguation in Basque. Based on this formalism, the group directed by Karlsson created the EngCG, namely the disambiguating-grammar for English. A more detailed presentation of this grammar is included later in this paper.

B. In syntactic analyzers: these are based on theories of grammar, such as *Lexical Functional Grammar* (LFG), *Generalized Structure Grammar* (GPSG), *Head Phrase Structure Grammar* (HPSG), *Government and Binding* (GB), etc. They mainly focus on

---

[3] See (Ezeiza 2002), a dissertation including further information on this topic.

[4] Disambiguating-grammars are also sometimes called *Taggers.*

sentences that are interesting from a linguistic viewpoint, rather than on real texts. However, parsers based on such descriptions typically fail when faced with texts in newspapers and technical texts. A further arising problem, due to the ambiguity, is that, for those sentences that they recognize, they provide several alternative interpretations and do not decide on the correct one. This is the reason why they are known to be of limited use in NLP. However, there are some applications that are based on theories of grammar that have targeted real texts. One example is the *Xerox Linguistic Environment* (XLE) (Kaplan & Newman 1997), which eases the creation of wide-coverage grammars based on LFG and obtains lexical and morphological information from external sources.

2.2. Probability-based techniques

During the last decade, approaches based on statistics have become increasingly common in taggers as well as in analyzers. They are based on empirical evidence that is retrieved by automatically analyzing large corpora.[5] However, it is not the case that they do not involve any basic linguistic knowledge. Rather, manual work on creating grammars is minimized to the limit, and linguistic knowledge from tagged corpora is retrieved by probabilistic means.

Most probability-based systems make shallow analyses by following this strategy. Taggers are one example, whose aim is to assign the syntactic category that fits to each word. Several statistical methods that involve various degrees of complexity have been employed to achieve this aim. The most simple ones (which use bigrams),[6] or those displaying greater difficulty, such as the decision-trees in (Màrquez 1999) (taggers employing machine learning systems), and the memory-based learning in (Daelemans et al. 1996).

Overall, the use of purely statistical methods has encountered difficulties in treating phenomena that appear outside the domain of limited texts. Thus, by looking at the results obtained, we conclude that such taggers display limited successful performance. For instance Voutilainen (1994); Brill & Wu (1998) report a %95-97 success for several languages. Such percentages are unacceptable when we consider syntax, since they would imply the existence of one error per sentence in a great number of sentences.

2.3. Approaches combining probability and linguistic knowledge

These approaches combine probabilistic methods and linguistic knowledge with the aim of gathering their advantages. In general, linguists write the rules of grammar, but the application of the rules is typically based on statistical knowledge. This statistical knowledge is extracted from large tagged corpora. The following employ this system of work: *IBM/Lancaster Approach* (Black et al. 1993) and Padró (1997).

Also, the *CG* formalism by Karlsson would fit in this group, since, although it is linguistic in nature, it employs some (although little) statistical information in the

---

[5] Also known as *treebank* or *parser bank*.

[6] The probability of a tag is calculated by considering the tag of the surrounding word.

English version. Apart for tagging purposes, CG is also used to improve shallow syntactic analyzers with great success. In fact, it has been one of the most successful system in the market for the last years.

The next section is about the CG analyzer, which is our choice for the morphosyntactic disambiguator as well as for the shallow syntactic analyzer.

3. Constraint Grammar (CG) parser

The CG parser was created in the 80's by Fred Karlsson and colleagues at the University of Helsinki. Here is the list of its most important characteristics, as stated in Karlsson et al. (1995):[7]

- Goal: the most important goal of this analyzer is to reduce ambiguity, i.e., to decide on correct/adequate analyses among the possible interpretations of a form. Another aim is to provide the shallow syntactic analysis of a text that has been previously morphologically analyzed.
- The goal of grammars aimed at *Parsers* (CG being one of them) is not to indicate the (non)grammaticality of sentences, but rather to provide a solution for every analysis by dispensing with the biggest possible number of erroneous/inadequate interpretations. A solution is sought for every element that needs analyzing in the text. In this sense, we may qualify this grammar as robust.
- It is independent from languages and from programming codes.
- Grammars and lexicons are adaptable to particular types of texts.
- The basis of the grammar is composed of constraint rules. Yet, when rules cannot provide a solution, there is room for the use of elements that contain probabilistic features; this contributes to robustness in the grammar.
- The core and basis of the grammar is the morphological analysis and the lexicon.
- The task of the grammar is threefold:
  - morphosyntactic disambiguation related to context;
  - assignment of boundaries between clauses;
  - assignment of surface syntactic functions and their disambiguation.
- CG is restrictive in the sense that its goal is to reduce morphological and syntactic ambiguity, in other words, to discard analyses that do not correspond to particular contexts.
- Since ambiguity arises at the word-level, the object-unit of analysis is the word.
- Syntactic analysis assigns a function to each word: first, it will complete words lacking a syntactic function in the database, and next, it will engage in the task of syntactic disambiguation. Thus, it will also provide information about the existing relations between words. However, the analysis is shallow and linear in the sense that it does not directly establish any tree or hierarchical relation.
- The basis of the rules is composed of the information provided by analyses of grammars and corpora.

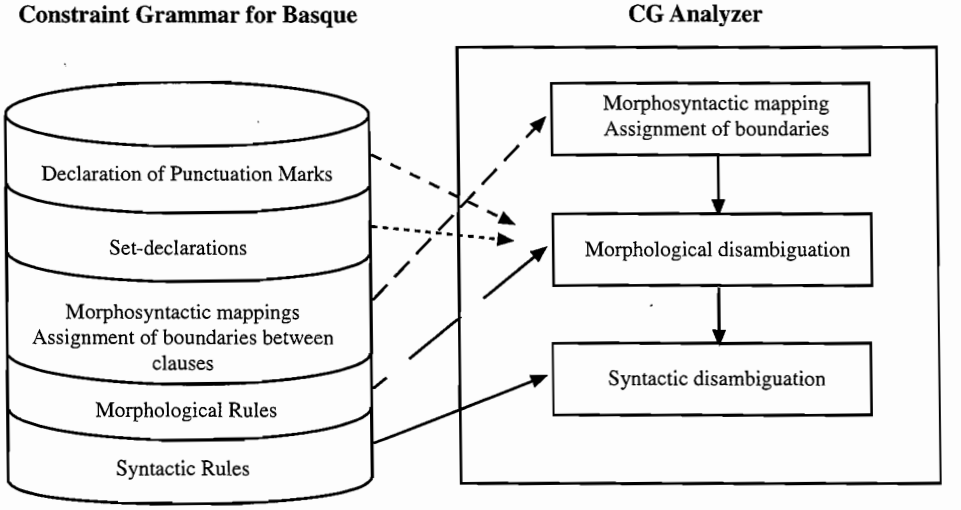
[7] The listed characteristics in the text strictly follow the scheme in the book.

— The amount of theoretical abstraction employed in the CGP is low compared to the rules in theories based on formal syntax, such as those that make use of the Government and Binding Theory or the Generalized Phrase Structural Grammar (GPSG).
— Rules are independent from each other.
— Pre-processing is important. The tasks of the pre-processor include tokenization, i.e., item-recognition,[8] assignment of boundaries to paragraphs, recognition of multi-word units, étc.
— There is room for using the standard SGML coding system in rules.

Constraint-grammars have been extensively employed in writing grammars for languages, although not all have been published. For English, we find EngCG (Karlsson et al. 1995), for Turkish Oflazer & Kuruöz (1994), for French (Chanod & Tapanainen 1995), for Finish, for Swedish, for Swahili, for Danish, for German, for Portuguese and for Spanish (Sánchez 1997), and for Basque (Aduriz 2000). The next section is a detailed presentation of the later.

4. Disambiguating-grammar for Basque: EUSMG

Considering the features described above, we chose the CG formalism for our purposes of starting to handle the syntax of Basque. The correctness of our choice has already been proven from a theoretical viewpoint. Let us next consider the application we have created for Basque. Although overall we have followed general principles, we have also been forced to make certain particular decisions.



IV.1. Architecture of the CG analyzer

[8] Tokenization involves token or item recognition. In other words, the isolation of the units that the morphological analyzer will use as input. These units may be elements such as lexical elements or marks of punctuation.

This grammar contains six parts, and the parser recognizes them all, even when they are empty of content (signaled with *NIL*). We will next present the details of the parts of the grammar, but first, consider the picture above in IV.1, which shows the general architecture of the grammar in a schematic guise.

4.1. Declaration of punctuation marks that delimit sentence boundaries

This will define periods, semicolons, question marks, etc.: DELIMITERS = "<$.>" "<$;>" "<$?>" "<$!>".

4.2. Set-declarations

Rules employ features of grammatical elements (including category, case, definiteness, etc.) in order to refer to the context of the word that is being treated. Often, it is possible to group elements that contain similar features. However, in order to use this information in rules, it is necessary to previously define the sets or groups. This section defines such sets. Consider the following examples:

[9] LIST ADLAG = "izan" "*edun" "*edin" "*ezan"

[10] LIST PERIFRASTIKOAK = ADOIN BURU EZBU GERO

[11] LIST DENB = DENB

4.3. Morphosyntactic mappings

The purpose of the mapping is to add information. The existing relations between morphological and syntactic features are typically expressed by mapping-rules. The mapping process assigns a syntactic function to each morphological interpretation. They are employed to assign syntactic functions that do not originate in the database. The grammar contains 83 mapping-rules.

Mapping-rules display the following shape:

<domain (operation), syntactic tag, the word *TARGET*, goal-interpretation, the word *IF*, contextual conditions >

— MAP (@-JADNAG) TARGET (ADI) IF (0 BURU) (1 ADL);

- Example: Basoan biziaz aberetu EGIN[12] dira

(lit. forest-loc. live-instr. beasten make have)
'By living in the forest, they have become beasts.'

---

[9] Set-declarations invariably start with the word LIST. ADLAG is the abbreviation for Auxiliary.

[10] The abbreviations for this set are the following: ADOIN "root of verb"; BURU "perfective"; EZBU "imperfective"; GERO "future tense".

[11] DENB "tensed".

[12] The word being treated will be written using capital letters.

Numbers signal the position of the element that we are working on (see more on this below, in the section that explains disambiguating rules). Numbers may either have a positive or negative value, which indicate the right or the left side respectively (for instance (1 ADL) refers to the right side). Position "0" refers to the element we are presently working on (e.g. (0 BURU)). Taking into account these clarifications, let us now consider how the mapping-rule above is paraphrased: *map the syntactic function @-JADNAG*[13] *to forms of the category verb (ADI), if the form itself is perfective* (0 BURU) *and if an Auxiliary* (1 *ADL*) *is placed to its right.*

4.4. Assignment of boundaries between clauses.

The assignment of clause boundaries is carried out by the mapping-rules mentioned above. When applicable, what is being assigned is the word MUGA. It recognizes both coordinate clauses that are attached by a conjunct as well as clauses attached by subordination. Let us consider an example:

— MAP (@MUGA) TARGET (KONP) IF (NOT 0 BAIT) (1C ADI-ADT OR ADPOSAG)[14]

- Example: Etorriko ZELA uste nuen
  Etorriko ZELA pentsatzen nuen

  (lit. 'Come-fut would-that thought I-did'),
  ('I thought that he/she would come.')

This rule states the following: *assign a boundary mark to completive subordinate clauses if they are immediately followed by a verb.* For more information on this, see Appendix C in Aduriz (2000).

4.5. Disambiguating Rules

Morphosyntactic and syntactic disambiguation is carried out by the same type of rules. These rules deal with both general and specific phenomena, and they contain the following domains:

<(domain) operation, goal-interpretation, the word *IF*, conditions of the word we are dealing with, contextual conditions >

The following example illustrates these domains:

— REMOVE (ADI) IF (0 ADJ) (NOT -2 DET)(-1 ZERO) (1 DET)

- Example: Bizitoki JAKIN bat ez zutela...

  (lit. 'home known one not had-they-that'),
  ('That they did not have a specific home.')

---

[13] -JADNAG "Non-finite main verb".

[14] Here is what the abbreviations in this rule stand for: KONP "completive"; ADI-ADT "verb and synthetic verb"; ADPOSAG "the component of periphrastic verbs built with the verb *izan*" (i.e., to be) (e.g.: *behar izan* "have to").

The rule can be paraphrased as follows:

> Delete the interpretation of the verb (ADI), if the form we are considering is also an adjective (0 ADJ), and if there is no Determiner in a two-word-distance to its left (NOT -2 DET); if, to a one-word-distance to its left there in an element with no morphemes (-1 ZERO), and to its right, to a one-word-distance, there is a determiner (1 DET).

Disambiguating rules can be classified into groups in terms for their degree of certainty in the correctness of the results that they provide. Particularly, in the grammar we have developed for Basque, we distinguish four groups of rules: the first group includes morphosyntactic rules that are most certain; the second contains morphosyntactic rules of less certainty; the third subsumes syntactic rules that are most certain, and finally, the fourth contains syntactic rules that are less certain, and more generally, idiosyncratic rules. One advantage of this classification is that it provides some order to the grammar, which would otherwise be hard to achieve in the presence of such amount of rules (Sánchez 1997).

The grammar for Basque contains 1.113 disambiguating rules: 672 in the first group; 45 in the second; 289 in the third, and finally, 107 in the fourth.

5. The bases and sources of the EUSMG

5.1. The Basque Lexical Database (i.e. *Euskararen Datu-Base Lexikala*: EDBL)

The lexicon, which is the core of the morphological processor, is organized in a database, namely in the EDBL (Alegria et al. 1997). When starting in an applied project of a real scale, it is necessary that linguistic data be structured and organized in a database. Although the EDBL was at first created to deal with morphology through a two-level formalism, it is currently the general Lexicon Database employed for treating Basque computationally. It is the basic necessary source of knowledge that is used in many aspects of NLP. This is the reason why it gathers various sorts of information: morphological and syntactic. Though it still does not include semantic information (namely, distinction among different meanings), the fact that it contains a homograph-identifier signals certain proximity to including semantic information (Agirre et al. 1994; Aduriz et al. 1998). Yet, the most important source of information is lexical information. The EDBL currently contains over ninety thousand entries, a meaningful number from our viewpoint, comparable to the amount employed for applications created for other languages.

As for the lexicographic-linguistic descriptions employed in building the EDBL, rather than following the demands of the computational applications, we have submitted to standard lexicography and to linguistic rules. The reason is that the Database is general, in the sense that it is designed for its use in several applications, and hence, we cannot restrict it according to the demands of specific applications.[15]

---

[15] In our case, ambiguity could have been reduced in the database itself. For instance, in the case of forms that are potentially ambiguous such as adjective/noun homographs, ambiguity can be reduced by assigning them a specific category in the database (as in the English version of CG, ENGCG: see Karlsson et al. (1995: 94-95).

The information corresponding to each entry is gathered in domains. Here are some: canonical form, two-level form (adapted to the model that will be employed in morphology), information regarding the morphemes that each category may take (continuation-classes), homograph-identifier, source, example(s), etc.

5.2. Corpora

The CG formalism belongs to approaches that are considered as empirical. Thus, one of its most prominent characteristics is that it is inclined to real corpora (Karlsson et al. 1995: 17) in two important respects. On the one hand, corpora, along with the grammar, are the source for linguistic information, especially in the process of creating disambiguating rules; on the other, they are the necessary locus for testing the degree of precision of applications and tools. Additionally, apart from the possible applications or uses mentioned above, corpora have become a necessary tool in statistical approaches. For instance, the tagging processes that employ the Markovian model and the model of Bayes employ corpora as recognition sources.

As for Basque, there are two corpora that are approved by the *Euskaltzaindia* or the Academy of Basque Language, and both are set within a project for lexical fixation:[16] the historical corpus called —*Orotariko Euskal Hiztegia* (OEH)— and the referential corpus that contains items of current use in Basque (named *"XX. mendeko euskararen corpus estatistikoa"*).[17] The later Urkia & Sagarna (1991; Urkia 2002) is a statistical corpus on written language, and it is balanced in terms of the types of texts.[18] This corpus is yearly updated, and currently contains 4.657.165 lemmatized words.[19]

We have employed parts of the XX century balanced corpus in our project. More specifically, we have made use of one part of the corpus in the process of rule-making and another when measuring the precision of the rules. There are few available Basque corpora (as is the case of all minority languages like Basque). In contrast, there are many resources for English, such as the Brown Corpus and the Penn Treebank. In fact, the size of the corpus is extremely relevant: in order to provide a thorough description of linguistic phenomena, it is necessary that the corpus be big. Thus, the more corpora available, the more thorough and precise are the researches based on them.

More information on corpora in Basque has been published by UZEI in the minutes from the meetings that have recently taken place, where they analyzed the current situation of corpora in the Basque Country (available from http://www.uzei.com).

---

[16] "Both corpora gather written documents, but they differ in one important respect: whereas the historical corpus collects complete works that are thoroughly stripped, the referential corpus is statistical in the sense that the focus of interest is lexical variety rather than the quality of works. It is called referential precisely because it reflects the current use of Basque" (Urkia & Sagarna 1991).

[17] The corpus of the XX century is being developed in UZEI. See *http://www.uzei.com* or *http://www.euskaracorpusa.net*

[18] "Corpora can be classified in a simple manner as being: balanced/non-balanced. In balanced corpora we find balance in the types of texts that they include, leaving aside particular characteristics pertaining to special texts. In order to achieve this, one must select many small, representative parts of texts from various sources by employing statistical techniques" (Alegria 1995).

[19] These are the available data as for the end of December 2002.

5.3. Morphological Analyzer

In section 3, where the basics of the CG formalism were discussed, we have mentioned that *the morphological analysis and the lexicon are its core and basis.* The morphological analysis of the EUSMG is provided by the analyzer, and the lexical is contained in the EDBL database. The previous section has presented the main characteristics of the EDBL database. Let us consider the features of the analyzer next.

In the beginnings of the computational treatment of Basque morphology, researchers searched for the model that would best describe the Basque features. There was more than one model available at the time (see Alegria 1995, Urkia 1997), and after several tests, we viewed the two-level morphology proposed by Koskenniemi (1983) as most suitable. Next, we will briefly mention the most relevant features of two-level models as stated in the above mentioned research by Alegria and Urkia: a) it is a general model, in the sense that it is applicable to any language, precisely because it distinguishes linguistic knowledge from algorithmic knowledge. b) it is applicable both to morphological analyses of words as well as for word generation; c) surface and deep lexical systems are clearly distinguished, which permits dispensing with the use of allomorphs; d) it employs parallel rules rather than the rewrite rules from generative phonology. This renders the system simpler both conceptually and computationally.

The core components of two-level formalisms are the lexical-system, morphotactics, and morphophonological rules. Other two characteristics of this analyzer are its robustness and its flexibility, which are the consequence of the following three improvements on the two-level morphology system by Koskenniemi. First, it left room for including the lexicon of the user; second, by employing the very same two-level model, non-standard forms corresponding to standard ones[20] were treated, with the aim of providing further robustness to the analyzer; finally, the analysis of unknown words was used for analyses of texts as well as for phonological analyses.

In fact, for each word that is recognized, the analyzer brings along all the information previously contained in the form of separate morphemes. In addition, often, it forms a set of ambiguous analyses, as in the following:

((**forma "bide"**)[21]
((**analisi 1**)
((lema "bide")((SAR bide)(KAT IZE)(AZP ARR))))
((**analisi 2**)
((lema "bide")((SAR bide)(KAT PRT)(MDL ZIU))))
((**analisi 3**)
((lema "bide")((SAR bide)(KAT IZE)(AZP ARR)))
((morf "0")((SAR 0)(KAT DEK)(KAS ABS)(MUG MG)(FS1 @OBJ)(FS2 @SUB)(FS3 @PRED))))

[20] The general description of Basque morphology constructed following the two-level model was done on standard Basque. The description of the non-standard Basque was done later.

[21] This is what the abbreviation in the analyses stands for: in the first analysis, SAR "entry"; KAT "category"; IZE "noun"; AZP "subcategory"; ARR "common". In the second analysis: PRT "particle"; MDL "mood"; ZIU "certain". In the third analysis: morf "morpheme"; DEK "declinative"; KAS "case"; ABS "absolutive"; MUG "definite"; MG "indefinite"; FS "syntactic function"; OBJ "object"; SUB "subject"; PRED "predicative".

Each possible analysis may include both morphological information (category, subcategory, case-definiteness, etc.), and syntactic information (syntactic functions). Some syntactic functions have already been defined in the database; however, others are assigned by mapping-rules, as we have just seen above when dealing with the grammar. Thus, the nature of the resulting ambiguity lies on the description of the language, in the sense that it arises as a consequence of the decisions made about this description and the criteria that have been followed. Hence, linguistic description necessarily conditions the result. This is the reason why it is worth being diligent when defining the analyzer and when building the lexicon for the database. Thus, the next step involves working on the results provided by the syntactic analyzer. This is the input-base for our work, namely the problem of ambiguity arising from analyzing texts.

6. Analyzing ambiguity

6.1. Delimiting the object of study

Ambiguity brings along disruption of communication in every aspect related to language. This is so, because more than one interpretation is available for each sentence, word, etc. Language is ambiguous in its nature.

> That ambiguity is natural to common language —to what we typically call bare language— in any of its various forms is such a well-known fact that there is no need to resort to refined dialectical and rhetorical techniques to convince the skeptical about it. (...). No doubt, ambiguity is one of the recurrent universals in natural language (...). (Michelena 1972).

The literature contains many references to this phenomenon, and the problem has been approached from various viewpoints. In fact, because of the extensive domain that it covers, it includes many types of alterations in language. Computationally, the problem of ambiguity is mostly related to *parsing* or to syntactic analysis. What must be first treated is the ambiguity arising from the information provided by morphological-morphosyntactic analyses.

Ambiguity is a persistent problem in linguistic analyses, but it becomes even a more serious and complicated trouble in computational analyses. In fact, these analyses process information (lexical and morphosyntactic) that has been gathered in the computer and provide results unlimitedly, often much unexpected ones. This is why ambiguity is one of the hardest problems in NLP, and especially in syntactic analyses.

Our object of study is, thus, the ambiguity arising in computational analyses. Before entering into details, let us mention that our study is set in shallow parsing. This conditions the domain of the object of study and the manner of dealing with ambiguity, which is far from any that has been proposed in theoretical approaches. Notice that we leave aside semantic, pragmatic and deep syntactic ambiguity (the later also called structural, as in *I saw the man with the telescope*). Rather, we study more local ambiguities, those concerning morphosyntax and syntactic functions. Recall that what are being considered are objects within blank spaces or words, and that disambiguation will be carried out by considering only their local context. Therefore, we deal with categorial disambiguation (e.g. the word *omen* may well be a particle, a common noun

or a verb in Basque), morphosyntactic ambiguity (the form *etxeak* 'the houses' may either be absolutive definite plural or ergative singular) and syntactic ambiguity or the ambiguity pertaining to syntactic functions (the word *etxeak* may either be subject, object or predicative).

It is obvious that such forms are not ambiguous in certain particular contexts. This is precisely the task of this grammar, i.e., to decide on the correct category, case or function of an object in a particular context. Keep in mind that forms are analyzed in isolation up to this point. However, in this step, where disambiguation is necessary, context becomes relevant.

6.2. Types of ambiguities

After presenting the types of ambiguities we are considering, we suggest distinguishing four types of morphosyntactic ambiguities: categorial ambiguity, ambiguity related to declension affixes, ambiguity in subordinating suffixes, and finally, ambiguity in aspect and mood-tense. We separate these types from the ambiguity arising in syntactic functions, which will be presented in a separate section below. With the purpose of getting information about the four types of ambiguities mentioned above, a first overall analysis of ambiguity was made by taking into account the most relevant linguistic phenomena. For this purpose, we used the morphological analyzer and the EDBL as sources, and based on these two we withdrew the first groups of ambiguity. Next, we made research on the corpus with the aim of accounting for the appearance of those groups of ambiguities. As a result, we evaluated the size of the problem and its actual relevance in each case. This contributed to robustness in the analysis.

Let us next consider in some detail the types of ambiguities, some illustrations of the problem, and the characteristics that they display:

1. *Categorial ambiguity:* categorial ambiguity is the most complicated and outstanding problem that we encounter after overcoming the level of the word. This has been widely attested in the literature that has been concerned with syntactic analyses and disambiguation (Karlsson et al. 1995, Padró 1997, Màrquez 1999). Within the typology of ambiguities in Basque, categorial ambiguity requires special consideration for the following reason: a word with a base category easily changes its function according to the context in which it appears. In these instances, it is necessary to resort to context in order to determine its category (or function)[22] (Euskaltzaindia 1993: 134). For present purposes, we will only mention the most clear cases, and for complete details see (Aduriz 2000). One clear case is the ambiguity between adjectives and adverbs. (e.g. *azkar* 'quicle, quiclely', *luze* 'long, lengthy'); others include the ambiguity arising in roots of verbs, adjectives and adverbs (e.g. *bizkor* 'hurry up, stimulate, vigorous, vigorously's'), or the one between auxiliaries and synthetic verbs (e.g. *da* 'be', *du* 'have').

[22] The fact that the terms category and function have recurrently been confused in Basque as well as in other languages has created many problems in creating a categorial system. On this problem, see further details in Huddleston & Pullum (2002), Zabala & Odriozola (1994), Aduriz (2000).

2. *Ambiguity in declension-affixes:* this group is also very productive in Basque. In fact, by virtue of being an agglutinative language, Basque employs morphemes on words where other languages employ syntactic structures. More specifically, Basque makes use of case-declension morphemes as well subordinating suffixes. Thus, ambiguity emerging from bound morphemes is extremely relevant in Basque. Among others, we find the following ambiguities in this group: absolutive definite plural and ergative definite singular markers are ambiguous; suffix *-ko* is ambiguous between locative-genitive, attributive and distributive.
3. *Ambiguity in subordinating suffix/prefixes:* this group includes ambiguities such as the one appearing in suffix *-(e)la* or in prefix *bait-*. More specifically, *-(e)la* may have the value of a complementizer or that of an adjunct (modal or temporal). As for prefix *bait-*, it can either express relative or causative meaning, it can form subordinate complements, or it can appear in consecutive clauses. It is important to note that the type of ambiguity in this section is closely related to verb subcategorization. In this sense, the more elaborate the subcategorization, the less ambiguity problems we will encounter.
4. *Ambiguity related to verbal aspect and mood-tense:* in this group, we find ambiguity arising in certain verbs between the verb-root and the participial form (e.g. *egon* 'be, stoy', *joan* 'go', etc.). Apart from the instances just illustrated, a big percentage of ambiguity in this group is due to ambiguity between synthetic verbs in the past (e.g. *nindoan* 'I was going', *zekarren* 'he was bringing') and the subordinating particle *-(e)n*.

6.3. Measuring Ambiguity

Next, we present the data about the percentages of appearance of the various types of morphosyntactic ambiguities that were described above. The results have been organized into two groups: first, we will present the data corresponding to categorial ambiguity, and next, we will include the rest of the information provided by the morphosyntactic analyzer in order to determine the results, i.e., we will include the three types of ambiguity that have been defined in the description (ambiguity related to free morphemes, subordinating elements and those related to verbs). In order to reach our goal, we have taken as basis a corpus that contains 14,000 text words. Because the corpus is real, apart from standard forms, we also find unknown words in the texts (those that are not included in the EDBL),[23] as well as variants to standard forms (dialectal forms, for instance) (Alegria 1995, Ezeiza 1997). The analyzer has provided a potential solution for every form through the analysis of standard forms, the variants and through the analysis of unknown words.

[23] "Half of the words that remain unanalyzed are not recognized because they are not identified in the corresponding lexicon (...). Although the reasons for the non-appearance of a word in the lexicon are diverse, (...) it is often impossible or extremely difficult to find all those lemmas in the general lexicon, for they are often the result of context or of the particular use of the author" (Alegria 1995).

	Categorial ambiguity	Complete morphosyntactic ambiguity
Standard forms	%46,34	%80,09
Variants	%32,89	%81,25
Unknown forms	%57,95	%95,88
Average	%37,80	%65,75

VI.1. Measurements of ambiguity

As the chart above shows, the number of possible interpretations of non-standard forms is larger than the one found in standard forms.

Categorial ambiguity includes 20 tags: 17 general tags and other three tags that serve to tag special cases such as ellipsis. This is the reason why the %37,80 of words in a text are ambiguous. In other words, rather than obtaining a single reading for each word, we find a bit more than one and a half per word. Moreover, ambiguity percentages double when we consider the overall morphosyntactic ambiguity, which reaches %65,75. In other terms, each form contains 2,81 interpretations. This is not surprising considering that all the information of the analyzer is located here: case, number, definiteness, mood and tense in the case of verbs, etc.

This ambiguity percentage differences between categorial ambiguity and the complete morphosyntactic ambiguity are not exclusively attested in Basque. Overall, for logical reasons, it is frequent in agglutinative languages: since these languages contain a rich morphology, this type of ambiguity that is not related to categories is much more persistent compared to languages such as English or Spanish. That morphosyntactic ambiguity is a serious problem is more obviously exemplified by Basque compared to the data in other languages (Karlsson et al. 1995: 23). For example, in Finnish, it reaches %11,2. In Swedish, as in Hebrew it is larger, %60 in both. Spanish displays around %43, and English %35.[24]

7. Morphosyntactic disambiguating rules

In everyday spoken language, the speaker (and the hearer) has available resources such as accent and context to solve ambiguity problems. Morphosyntactic disambiguating rules are to replace the resources that we mechanically employ in spoken language.[25] What do we understand by the term *disambiguation*? The answer is the following: to choose one out of various possible ways of understanding a form in a given context.

---

[24] However, it is difficult to make ambiguity comparisons between Basque and other languages for two reasons: first, because the basic tags employed vary, and second, because the base-texts are also different in nature. In order to obtain comparable results, both base-texts and the tagging-system should ideally be similar (Màrquez 1999).

[25] The previous section has restricted the domain of ambiguity we have treated and the types of ambiguities we have considered.

> In Constraint Grammar, disambiguation does not mean "bring out all alternatives" but rather "pick the appropriate alternative(s) by discarding one or more inappropriate ones". The Constraint Grammar notion of morphological disambiguation is functionally similar to the notion "homograph separation" (...) (Karlsson et al. 1995).

Disambiguating rules came into existence with the aim of fixing ambiguity problems that had previously been detected. A 14.000 word corpus was used for creating the rules. The rules are tested once and again, the errors are fixed, and information added; in other words, rules are refined until texts are correctly disambiguated. Since constraint-rules are a consequence of grammar-rules, in each case of ambiguity, we will derive principles similar to the grammar-rules that are derived from sets of rules.

The next section includes an illustration of this, namely, an example of rules and principles pertaining to morphosyntactic ambiguity. In fact, we will not thoroughly explain the grammar itself, since we have already explained in detail the design of the grammar where disambiguating rules are organized as well as the details of the syntax of the rules.[26]

7.1. Rules and Principles

The grammar contains 1.113 rules, and we have written the theoretical principles based on these rules. Mostly, rules are general, in the sense that they may refer to a whole group of ambiguities. However, in some cases, rules must be particular and they may only refer to specific words in a group of ambiguities. We will present disambiguating rules with an example: we will first explain a general principle, and next, we will show one of the rules that corresponds to this theoretical principle. We will provide the example[27] together with the rule.

Let us consider one of the examples of ambiguity that was mentioned earlier, in section 6.2: the form *bizkor* may be either the root of the verb-form *bizkortu*, an adjective, or an adverb. The following is the theoretical principle that deals with this issue:

> In periphrastic forms, roots of verbs take auxiliaries of the form **edin* or **ezan*[28] (ADL1) either to their left or right depending on the sentence-type. Participles, instead, are accompanied by either the auxiliary izan or **edun.*

One of the rules corresponding to this principle is the following:

— SELECT (ADOIN) IF (0 ADJ-ADB) (1C ADL1);[29]

---

[26] A detailed explanation of the grammar is available in the report (Aduriz et al. 2000, 2003).

[27] The examples of the rules that are provided in the text are mostly taken from real corpora.

[28] **edin* and **ezan* are roots of auxiliaries that express possibility, subjunctivity and imperative meanings. Instead, in the indicative, we use *izan* and **edun.*

[29] The abbreviations used in the rules: ADOIN "root of verb"; ADJ "adjective"; ADB "adverb"; ADL1 "set that includes **edin* and **ezan*".

- Example: azkar BIZKOR zaitezen

  Lit. 'Fast hurry-up you-subj.'
  'That you may hurry-up quickly.'

The application of this rule on ambiguous forms like *bizkor* gives as a result the choice corresponding to the root of the verb and discards the adjectival and adverbial interpretive options. However, a single rule does not always provide the correct conclusive analysis. Complete disambiguation is achieved through the intersection of the set of rules that consider context. There is the possibility that a certain context is not defined or that some casuistry has not been taken into account. If this is the case, the word will remain ambiguous. This may provoke an error in choosing the correct option among the possibilities by ruling out the correct reading. Let us consider how we have measured all this casuistry.

7.2. The results

Chart VII.1 shows the results of categorial disambiguation.[30] The results have been obtained by using a corpus of 10.000 words that has not been previously used in testing and rule-making:[31]

	Analyses per word	Ambiguity	Correctly interpreted words
Input	1,50	%37,80	%100,00
Output	1,18	%14,12	%99,12

VII.1. Results of categorial disambiguation

Compare this with Table VII.2 below, which displays the data resulting from the complete morphosyntactic disambiguation in the same corpus:

	Analyses per word	Ambiguity	Correctly interpreted words
Input	2,81	%65,75	%100,00
Output	1,76	%33,28	%97,51

VII.2. Results of the complete morphosyntactic disambiguation

[30] The data, wieh were used in Aduriz (2000), correspond to the year 2000.

[31] It has been necessary to manually disambiguate part of the corpus that has been used to calculate the results of the grammar and measure its efficiency. This task was also carried out within the IXA group while creating the rules.

The data show the robustness and power of the disambiguating-grammar when treating real texts. The amount of morphosyntactic ambiguity has dropped down to half: from %65,75 in the input to %33,28. In other words, in the input, from 2,81 possible analyses for each word, it has dropped to 1,7. In this disambiguating process, the correct interpretations have maintained at %97,51. Categorial disambiguation shows results that are even more successful. From 1,50 possible analyses per word, we get almost only one (1,18). In terms of percentages, ambiguity in the input has dropped from %37,80 to %14,12 in the output. Moreover, correct interpretations are maintained at %99,12.

In our view, the amount of errors found in categorial ambiguity is totally acceptable (0,8). It is more serious if we consider the whole morphosyntactic ambiguity. In fact, we know that the existence of non-standard forms considerably raises ambiguity percentages, and that the fact that the grammar is written for the standard mode often provokes errors. In addition, the percentage pertaining to non-disambiguated forms is mostly due to absence of information (the topic of subcategorization). To this, we need to add the cases that are impossible to disambiguate morphosyntactically due to their semantic or pragmatic nature. Also, remember that input ambiguity depends on the linguistic descriptions that have been made. Since the basis for this task, i.e., the database, is constantly being updated, the ambiguity that is added and the efforts of disambiguation do not go in parallel. Thus, the fact that we work on texts, descriptions and real data constantly requires updating. All this clearly shows the cyclicity in our work, as well as the distance from data that we sometimes need to set.

8. First steps in the syntactic analysis: shallow parsing (partial syntax)

Constraint Grammar works on a shallow or partial type of syntax. Thus, the main goal is to assign the corresponding syntactic function(s) to each word, and next, to disambiguate them (as we have just seen for the morphosyntactic level). This results in a shallow analysis of syntax.

As it was just mentioned, the first step is to assign all the possible syntactic function-tags[32] to every word-form. The procedure of assigning syntactic function-tags to words or morphemes is parallel to the assignment of morphological features to words. The next task is to reduce syntactic ambiguity. Syntactic ambiguity refers to situations where a word displays more than one syntactic function-tag (note that, for our purposes, ambiguity refers to syntactic functions rather than to structural ambiguity). Consider the following example in detail:

> Txakurrak (@OBJ @SUBJ)
>
> 'Dogs/the dog' (ambiguous between the object and the subject function respectively)

Again, as in the previous section, the main aim is to reduce ambiguity. For this purpose, we employ the syntactic constraints that are based in context. The goal of

[32] Syntactic function-tags are labeled with @.

applying syntactic constraint-rules is to reduce the number of potential syntactic tags into one, namely the correct one. However, constraint-rules do not attempt at cases where syntactic ambiguity is irreducible. For example, in the sentence "**Txakurrak** (@OBJ @SUBJ) **egunkariak** (@OBJ @SUBJ) ahoan zekartzan" (lit. '**dog-the papers** mouth-in-the was-bringing'), the words in boldface are left without disambiguating. Another option is that a decision would be made for isolating a single syntactic function. Recall that we do not consider structural or semantic-pragmatic ambiguities in this step.

8.1. Syntactic Functions

Syntactic functions-tags inform us about the function of words in sentences. They provide direct information about shallow parsing or shallow syntax, namely information about the surface structure of verb-chains. In fact, they provide information about the existing relations between words. Many of these functions do not correspond to the traditional functions; they are often tags that serve to form phrases or verb-chains. In fact, the grammar requires that every word must necessarily carry some tag. This is the reason why we sometimes find the traditional syntactic tags, and others, mere tags.

There are two main types of function-tags: heads and modifiers. *Main syntactic functions* correspond to the former, for instance, subject, object, indirect object, etc., and also to certain syntactic functions pertaining to verbs (+*JADNAG* e.g., matrix finite verb). In contrast, *modifiers* indicate the direction relative to their head: they express syntactic dependencies between elements within noun phrases or verbal periphrases. For example, adjectives that are followed by a noun that modify it are labeled with the tag @IZLG>, and the tag itself indicates that the head being modified is placed to its right (((*mendiko* (@IZLG>) *tontorretik* (@ADLG)),[33] literally "mountain-of top-from", meaning "from the top of the mountain").

Syntactic function-tags follow the philosophy of the CG formalism, in the sense that they are based on the *functionally labeled dependency syntax.*[34] By adopting the CG formalism, we express the syntactic functions of words and the interdependencies that exist among them.

8.2. Syntactic disambiguation

In the previous section, we have mentioned the importance of syntactic functions in shallow parsing in terms of the information that they provide about the existing syntactic relation among words. Thus, in cases where a reading contains more than one function, i.e., when it is ambiguous, we will need to disambiguate it.

The disambiguation of syntactic functions is carried out by syntactic rules, just as in morphosyntactic ambiguity. Thus, the aim of syntactic rules is to reduce the number of

---

[33] For further information on the later function-tags that we have employed, see the appendices in (Aduriz 2000).

[34] The concept of dependency-syntax has a long tradition in grammatical analyses since the Greco-Roman era. More recently, within the application of formalisms to syntactic theory, among others we find Tesnière (1959), Hays (1964) and Mel'čuk (1988), the ones who have recovered dependency-syntax in theoretical terms.

function-tags in each word-form into one. The set of rules, which includes both morphological and syntactic rules, are related to each other. In fact, syntactic constraint-rules are applied only after morphosyntactic disambiguation has taken place.

Let us consider one syntactic-ambiguity problem: absolutive forms in singular, plural and indefinites may either have subject, object or predicative functions. When faced with this ambiguity problem, several disambiguating rules have been created. One of them is the following:

— REMOVE (@OBJ) (0C ABS) (NOT *-1 NORK) (*-1 (NR_HU)) (1 (PUNT_PUNT));

- Example: Eta bertan agortu zen haren ODOL-JARIOA.

  Lit.: And there dry-up did his/her blood-flow
  "And there ended his/her bleeding."

The above rule can be paraphrased as follows: *delete the object function from the reading, if the word only contains absolutive case (0C ABS); if the sentence contains a verb of the* NOR[35] *paradigm (NR_HU), if it contains no auxiliary of the* NORK[36] *paradigm (NORK) and if there is a full stop to the right of the ambiguous form (PUNT_PUNT), i.e., if the sentence ends.* The example above suggests that agreement is of key importance to disambiguate syntactic functions, because it includes information on verbal subcategorization. This device will be particularly useful in disambiguating verbs including subordinating affixes.

9. Conclusions

We have developed a constraint-grammar for Basque in terms of shallow parsing with two aims: first, to obtain disambiguation of words that appear in real texts, and second, to develop the first steps in syntax by defining the existing surface relations between words. The main contributions of this grammar include a systematic analysis of ambiguities related to morphosyntax and to shallow parsing, as well as the specification of the disambiguating rules. However, the results of the syntactic analysis include no explicit phrasal-structure, since it does specify any hierarchies of components of phrasal nature.

Along these lines, several successful grammars that are capable of recognizing phases and verb-chains have been developed with great success (Arriola 2000, Aduriz et al. 2001). These grammars are based on the results of the analysis of CG, and mostly on syntactic functions, and they recognize the basic syntactic functions. The applications of these grammars would follow the grammar that we have presented in this paper.[37] Moreover, the results provided by CG have been the basis for other solutions in syntax, for example in the development of a PATR grammar, as presented in Gojenola (2000).

---

[35] Auxiliaries of the NOR paradigm arise when the verb subcategorizes for a single argument. This argument is marked Absolutive.

[36] Auxiliaries of the NORK paradigm arise in transitive environments, where the verb subcategorizes for a subject and an object, which will be marked Ergative and Absolutive respectively.

[37] These grammars have been created by using the mapping-rules of Constraint Grammar.

Also, the analysis (of grammar and phrasal disambiguation) will invariably serve as a basis and starting point for a deeper analysis. Thus, with the goal of achieving deeper analyses, the latest research on syntax in the IXA group is aimed at creating a corpus that is analyzed both syntactically and semantically. In order to do the syntactic annotation, we are currently working on a dependency-based grammar for Basque (Aduriz et al. 2002).

References

Aduriz, I., 2000, *EUSMG: Morfologiatik sintaxira Murriztapen Gramatika erabiliz.* Doctoral Dissertation, Basque Philology Section, University of the Basque Country.

—, Aldezabal, I., Ansa, O., Artola, X., Díaz de Ilarraza, A., Insausti, J.M., 1998, "EDBL: a Multi-Purposed Lexical Support for the Treatment of Basque". *Proceedings of the First International Conference on Language Resources and Evaluation (LREC'98),* Granada.

—, —, Aranzabe, M., Arrieta, B., Arriola, J.M., Atutxa, A., Díaz de Ilarraza, A., Gojenola, K., Oronoz, M., Sarasola, K., 2002, "Construcción de un hábeas etiquetado sintácticamente para el euskara". *XVII Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural* (SEPLN). Revista n.º 29. Valladolid, España.

—, Arriola, J.M., Artola, X., Diaz de Ilarraza, A., Gojenola, K., Maritxalar, M., 2000, *Euskararako Murriztapen Gramatika: mapaketak, erregela morfosintaktikoak eta sintaktikoak.* UPV/ EHU/LSI/TR 12-2000.

—, —, Díaz de Ilarraza, A., 2003, "Desanbiguazio morfologikoa, azterketa sintaktikoaren lehen urratsak eta aplikazioak Murriztapen Gramatikaren eredu konputazionala jarraituz". In J.M. Makatzaga & B. Oyharçabal (eds.), *Euskal gramatikari eta literaturari buruzko ikerketak XXI. mendearen atarian. Gramatika gaiak,* Iker-14 (I), Euskaltzaindia, Bilbao: 3-35.

Agirre, E., Arregi, X., Arriola, J.M., Artola, X., Insausti, J.M., 1994, "Euskararen Datu-Base Lexikala (EDBL)". Report UPV/ EHU / LSI / TR8-94.

Alegria, I., 1995, *Euskal morfologiaren tratamendu automatikorako tresnak.* Doctoral Dissertation, Language and Computational Systems Section, University of the Basque Country.

—, Artola, X., Sarasola, K., 1997, "Hizkuntzaren tratamendu automatikoa", *Jakin* 102, 61-82.

Arriola, J.M., 2000, Euskal hiztegia-*ren azterketa eta egituratzea ezagutza lexikalaren eskuratze automatikoari begira.* Doctoral Dissertation, Basque Philology Section, University of the Basque Country.

Black, E., Garside, R., Leech, G., 1993, *Statistically-Driven Computer Grammars of English: The IBM / Lancaster Approach.* Black, Garside & Leech (eds.), Rodopi, Amsterdam.

Brill, E. & Wu, J., 1998, "Clasifier Combination for Improved Lexical Disambiguation". *COLING-ACL'98,* Montreal.

Chanod, J.P. & Tapanainen, P., 1995, "Tagging French – comparing a statistical and a constraint-based method". *Proceedings of EACL'95,* 149-156.

Daelemans, W., Zavrel, J., Berck, P., Gillis, S., 1996, "MBT: A Memory-Based Part-of-speech Tagger Generator". *Proceedings of the 4th Workshop on Very Large Corpora,* 14-27. Copenhagen.

Euskaltzaindia, 1993, *Euskal Gramatika Laburra: Perpaus Bakuna.* Euskaltzaindia, Bilbo.

Ezeiza, N., 1997, *EUSLEM, euskararako lematizatzaile/etiketatzaile baten diseinua eta inplementazioa.* Small Thesis Report, Language and Computational Systems Section, University of the Basque Country.

—, 2002, *Corpusak ustiatzeko tresna linguistikoak. Euskararen etiketatzaile morfosintaktiko sendo eta malgua.* Thesis report, Language and Computational Systems Section, University of the Basque Country.

Gojenola, K., 2000, *Euskararen sintaxi konputazionalerantz.* Doctoral Dissertation, Language and Computational Systems Section, University of the Basque Country.

Green, B. & Rubin, G., 1971, *Automatic Grammatical Tagging of English.* Providence: Brown University.

Hays, D.C., 1964, "Dependency theory: a formalism and some observations", *Language* 40, 511-25.

Huddleston, R. & Pullum, G., 2002, *The Cambridge Grammar of the English Language.* Cambridge University Press.

Kaplan, R.M. & Newman, P.S., 1997, "Lexical Resource Reconciliation in the Xerox Linguistic Enviroment". *Proc. of a Workshop on Computational Enviroments for Grammar Development and Linguistic Engineering,* Madrid.

Karlsson, F., Voutilainen, A., Heikkila J., Anttila, A., 1995, *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text.* Mouton de Gruyter, Berlin.

Koskenniemi, K., 1983, *Two-level Morphology: A general Computational Model for Word-Form Recognition and Production.* Doctoral Dissertation, University of Helsinki.

Màrquez, L., 1999, *Part-of-Speech Tagging: A Machine Learning Approach based on Decision Trees.* Doctoral Dissertation, Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya.

Mel'čuk, I.A., 1988, *Dependency Syntax: Theory and Practice.* State University of New York Press.

Michelena, L., 1972, "De la ambigüedad sintáctica", *Revista Española de Lingüística* 2, 237-247.

Oflazer, K. & Kuruöz, I., 1994, "Tagging and morphological disambiguation of Turkish Text". *Proceedings of ANLP-94,* 144-149.

Padró, L., 1997, *A hybrid environment for syntax-semantic tagging.* Doctoral Dissertation. Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya.

Sánchez, F., 1997, *Análisis morfosintáctico y desambiguación en castellano.* Doctoral Dissertation, Department of Linguistics, Modern Languages, Logic and Philosophy of Science. Universidad Autónoma de Madrid.

Tesnière, L., 1959/66, *Eléments de Syntaxe Structurale,* (2nd revised edition). Paris, Klincksieck.

Urkia, M., 1997, *Euskal morfologiaren tratamendu automatikorantz.* Doctoral Dissertation, Basque Philology Section, University of the Basque Country.

—, 2002, (forthcoming), "XX. mendeko euskararen corpus estatistikoa", in *Hizkuntza-corpusak. Oraina eta geroa.* Donostia.

— & Sagarna, A., 1991, "Terminología y lexicografía asistida por ordenador. La experiencia de UZEI". *Actas del VII congreso de la SPLN.* Donostia.

Voutilainen, A., 1994, *Designing a Parsing Grammar.* Publications of the Department of General Linguistics, 22. University of Helsinki.

Zabala, I. & Odriozola, J.C., 1994, "'Adjektiboen' eta 'adberbioen' arteko muga zehatzik eza", *ASJU* 28.2, 525-541.

THE TRANSITIVITY OF BORROWED VERBS IN BASQUE: AN OUTLINE

Xabier Alberdi Larizgoitia

Abstract

In this paper I will study the main patterns that Basque speakers follow when they choose the auxiliary used with verbs borrowed from neighbouring languages.

*In order to do that, the method I follow consists in analyzing the way borrowed verbs appear in the Standard Lexicon (*Hiztegi batua*) of the Basque Language Academy.*

First, after exposing the theory of unaccusativity and proposing a typology of Basque verbs, the main tendancies observed in the borrowing of verbs from Spanish will be analyzed groups by groups: 1) borrowed verbs which unique auxiliary is du *'have' (a.* adoratu *'adore',* bisitatu *visit'...; b.* estudiatu *'study',* eskribitu *'write'...; c)* abdikatu *'abdicate',* abortatu *'abort'...); 2) borrowed verbs which alternate the* da *'be'/* du *'have' auxiliaries (a. causative alternation:* sikatu *'dry'; b. intransitive/reflexive alternation:* obligatu *'oblige'; 3) borrowed verbs that only have the* da *'be' auxiliary (*burlatu *'mock',* konformatu *'be satisfied with'). Among the results of this survey, it appears that the hypothesis which states that Basque borrowed monadic verbs strongly tend to be unergative must be abandoned: when verbs having only one argument are borrowed, the tendancy to use them with the* du *'have' auxiliary is mostly restricted to agentive intransitive verbs that don't have the clitic* se *in Spanish (*saltatu *'jump',* deskantsatu *'rest',* dantzatu *'dance',* olgatu *'play about'...).*

1. Introduction

The main aim of this article is to refute a widespread belief or hypothesis according to which verbs borrowed from Spanish or French enter Basque principally as *du*-type verbs. According to this supposition, borrowed verbs are generally conjugated by means of the same set of auxiliaries (e.g. *du* 'has') which are also used to conjugate transitive verbs of all types. While use of these auxiliaries is expected where syntactically transitive verbs are concerned (such as *ikusi du* '(he/she/it) saw (it/her/him)', *jan du* '(he...) ate (it)', *estudiatu du* '(he...) studied (it)'), I am interested in the claim that they are also being assigned to Romance-origin single-argument (intransitive) verbs. The number of single-argument *du*-type verbs, called unergatives in this paper, would thus seem to be on the increase in modern Basque. Such verbs contrast with Basque "pure" intransitives, which are conjugated with another set of auxiliaries (e.g. *da* 'is': *etorri da*

'(he/she/it) came', *erori da* '(he...) fell', *pasatu da* '(it) happened'). Examples of unergative borrowed verbs are *dudatu du* '(he) doubted'; *erregutu du gu guztion alde* '(he) prayed for all of us'; *ez du funtzionatu* '(it) didn't work'.

If indeed the hypothesis of a general trend towards unergatives turns out to be wrong, what criteria *do* Basque speakers apply to assign transitivity to borrowed verbs? Does any general principle determine the choice between *da* and *du* in such cases? A second aim of the article will be to answer this question.

I shall take as my starting point the borrowed verbs listed in a normative dictionary for modern Basque, *Hiztegi Batua* [Standard Dictionary] (henceforth *HB*). I shall not consider all the borrowed verbs listed but only those with a transparently foreign etymology. I shall supplement that information with additional examples and statistics obtained from other dictionaries and corpora, drawing from the following sources: Sarasola's *Euskal Hiztegia* [Basque Dictionary] *(EH)*; the text corpus for the *Orotariko Euskal Hiztegia* [General Dictionary of Basque] *(OEH)* of the Basque Language Academy (Euskaltzaindia); and *Egungo Euskararen Bilketa-lan Sistematikoa* [Systematic Corpus of Contemporary Basque] *(EEBS)*. I shall limit myself to citing but a few examples for want of space.

It will be necessary to consider how Spanish and French classify verbs, particularly with regard to unaccusative and unergative types. In this I shall follow Mendikoetxea (1999), a summary of whose findings falls outside the scope of this paper. I will also refer to the framework proposed by Levin and Rappaport (1995), the standard work on unaccusativity. The present paper represents part of a larger project, omitting numerous points and illustrations, but it is hoped it will nevertheless prove intelligible and useful as it stands.

2. Unaccusativity and the classification of the Basque verb

Linguists classify verbs by different criteria. This section will be concerned with one particular classification developed and used in syntactic studies. In this classification, grammatical terms originating from older traditions, such as 'transitive' and 'intransitive', appear side by side with others arising from generative grammar, such as 'unaccusative' and 'unergative'.[1] This classification aims at a refinement of the traditional 'transitive'-'intransitive' dichotomy. Adopting Gràcia's criteria (2000: 8-10), in the present section I attempt to provide definitions for each verb class in a way that is useful and valid for the analysis of Basque.

Transitive

Only verbs having two arguments will be considered transitive. Thus I exclude single-argument verbs conjugated with *du*, such as *irakin* 'boil (intr.)' and *iraun* 'last, endure', traditionally classed as transitives in some Basque grammars. Transitive verbs are those that take an argument having the syntactic function of direct object which in Basque is assigned absolutive case (zero). Transitive verbs assign the ergative case *(-k)* to the subject. The object is generated on a sister branch to the V[erb]. Some examples of

[1] Cf. *inter alia* Gràcia et al. (2000: 8-10) and Artiagoitia (2000: 169-184).

verbs classified as transitive are *erail* 'kill, murder', *esan* 'say', *maitatu* 'love', *kezkatu* 'worry (tr.)', *ikasi* 'learn, study' and *estudiatu* 'study'.

I shall also consider transitive the type of verb, such as *jan* 'eat', *edan* 'drink', *ikasi* 'learn, study' and *estudiatu* 'study', which normally meets these criteria but allows an 'absolute construction' wherein a prototypical object is not overtly present, as in *Peruk gauez ikasten du* 'Peru (Erg.) studies [something] at night', cf. *Peruk frantsesa ikasten du* 'Peru (Erg.) studies French (Abs.)'.

Unergative

Unergative verbs are intransitive (single-argument) verbs that have an external argument, generally either an agent or an experiencer (cf. Lafitte 1944/1979: §411), like transitives, they are conjugated using *du*-type auxiliaries in Basque. An unergative verb's single argument is both the syntactic subject and the notional subject. Such verbs usually denote an activity or process having an agent. The label 'unergative' is retained even though this sounds somewhat awkward in Basque grammar since the single argument of an unergative verb in Basque is assigned the ergative case *(-k): urak irakin du* 'the water (Erg.) boiled'. The following are examples of Basque verbs that are classified as unergatives: *irakin* 'boil (as of water)', *iraun* 'last', *dudatu* 'doubt', *alde egin* 'go away'.

Unaccusative

Unaccusative verbs are those having a single subject argument generated in internal argument position; in other words, intransitive verbs lacking external arguments. The surface subject of unaccusative verbs is a derived subject promoted from object position, so it is an internal argument. The argument with the function of syntactic subject is the notional object argument or 'theme' at the semantic level. In Basque, unaccusative verbs assign absolutive (zero) case to the subject and are conjugated using *da*-type auxiliaries: *Peru gaur etorri da* 'Peru (Abs.) came today'.

While many linguists (e.g. Perlmutter 1978, Levin and Rappaport 1995) employ 'unergative' and 'unaccusative' to distinguish between the two foregoing types of intransitive verb, others use 'unergative' and 'ergative' to refer to the same opposition (Burzio 1981, 1986), substituting 'ergative' for 'unaccusative'. Still other authors (Campos 1999, de Miguel 1992) refer to the two subclasses of intransitive verbs 'intransitive' and 'unaccusative' respectively, calling unergative verbs 'intransitives' (or 'pure intransitives').

In this paper I shall normally employ 'unergative' and 'unacusative' respectively as designations of the two major types of intransitive verb. However, in Part 3, I shall distinguish between 'unaccusatives' and 'pure unaccusatives'. The former label will refer to verbs possessing a transitive alternant, such as *apurtu da/du* 'break (intr./tr.)', *sartu da/du* 'enter, go in/insert, put in', *normalizatu da/du* 'be(come) normalized'/'normalize'. Those lacking any such alternation (e.g. *jaio da/*du* 'be born', *erori da/*du* 'fall', *egon da/*du* 'be, stay') will be identified as 'pure unaccusatives'.[2]

[2] Gràcia et al. (2000) propose distinguishing the two subclasses simply as 'unaccusative' (e.g. *egon* 'be, stay', *etorri* 'come') and 'ergative' (e.g. *sartu* 'go in/put in', *jarri* 'stand, sit/put, place').

(1) Pure unaccusative verbs:
a. erori *(da/*du)* 'fall'
b. jaio *(da/*du)* 'be born'
c. egon *(da/*du)* 'be, stay'
d. heldu *(da/*du)* 'arrive'
e. ahalegindu *(da/*du)* 'try'

(2) Unaccusative verbs:
a. sartu *(da/du)* 'go in/put in'
b. apurtu *(da/du)* 'break'
c. bainatu *(da/du)* 'bathe'
d. ireki *(da/du)* 'open'
e. hedatu *(da/du)* 'expand, spread'

3. Transitivity of Basque borrowed verbs

To classify Basque's borrowed verbs we shall begin by looking at their formal features. As in the *H[iztegi] B[atua]*, we shall start with three types of verb according to the auxiliaries with which they are conjugated: a) *du* verbs; b) *da/du* verbs; c) *da* verbs. I will draw from Mendikoetxea's (1999) and Levin and Rappaport's (1995) theoretical analysis of Spanish unaccusative structures for comparison and explanations.

3.1. Borrowed verbs that only take *du* (A. *adoratu, bisitatu...*; B. *estudiatu, eskribitu...*; C. *abdikatu, abortatu...*)

First let us look at borrowed verbs that only admit the *du* auxiliary. As we have already seen, not all verbs of the *du*-type are transitive. Furthermore, some transitive verbs fall not into the *du*-type but the *da/du*-type because of the alternation that they admit. The *du*-verbs may therefore be subdivided into the following three groups:

a) Transitives that only take *du*: *adoratu* 'adore', *bisitatu* 'visit', *fabrikatu* 'manufacture', *estimatu* 'esteem, consider', *kantatu* 'sing', *kobratu* 'charge, receive (payment), recover', etc.
b) A subset of transitives that may be used absolutely (without a specified direct object) as described in the preceding section, thus permitting transitive/unergative alternation: *estudiatu* 'study', *eskribitu* 'write', etc.
c) Unergatives (intransitives conjugated with *du): abdikatu* 'abdicate', *abortatu* 'have a miscarriage or abortion', *abusatu* 'abuse, go too far', etc.

These three subtypes are next considered in turn.

3.1.1. *Transitive verbs that only take* du (adoratu, bisitatu, fabrikatu...)

The borrowed verbs in the following list are transitives that are always conjugated with *du*:

BORROWED TRANSITIVE VERBS THAT ONLY TAKE *DU*

abandonatu 'abandon'
abisatu (*Nor.*) 'notify'
absolbitu 'absolve'
adjudikatu 'adjudicate'
administratu 'administer'
adoratu 'adore'
aglutinatu 'agglutinate'
akabatu 'slaughter, murder, kill'
amortizatu 'pay off'
analizatu 'analyse'
?apailatu 'prepare, get ready'
apartatu 'separate, take away, select'
aplikatu 'apply'
aprobetxatu 'take advantage of, make use of'
apuntatu 'aim at, point out, note'
arnegatu (*fedea*) 'abnegate'
artxibatu 'file'
aseguratu 'insure'
atrapatu 'catch'

autorizatu 'authorise'
baloratu 'value, appraise'
baltsamatu 'salve, annoint'
bernizatu 'varnish'
bernuzatu 'carve, sculpture'
bisitatu 'visit'
bonbardatu 'bomb'
bordatu 'embroider'
brodatu 'embroider'
dedikatu 2 'dedicate'
deduzitu 'deduce'
defendatu 'defend'
defenditu 'defend'
definitu 'define'
deklaratu 2 'declare, testify'
dekoratu 'decorate'
deportatu 'deport'
desarmatu 'disarm'
deseatu 'desire, wish'
desenkusatu 1 'excuse'
desinfektatu 'disinfect'
desiratu 'desire, wish'
deskribatu 'describe'
desobeditu 'disobey'
desordenatu 'throw into disorder'
despeditu 2 ("kaleratu") 'dismiss, fire'
desterratu 'exile, banish'
destilatu 'distill'
determinatu 'determine'
diktatu 'dictate'
diskriminatu 'discriminate (against)'
doblatu 'double, dub'
drainatu 'drain'
ebanjelizatu 'evangelise'
endredatu 'confuse'
engainatu 'deceive'
entenditu 'understand'
erlazionatu 'relate'
erratu 'err'
errebelatu 2 'rebel'
erreferatu 'respond, rejoin'
erreformatu 'reform'
?*errefusatu* 'refuse, turn down'
erregalatu 'give away, make a present of'
erremediatu 'remedy, redress'
erreparatu 1 'make reparations for'
errepikatu 'repeat'
errespetatu 'respect, honour'
errezatu 'pray'
errezibitu 'receive'
eskarniatu 'mock, taunt'
eskribitu 'write'
esleitu 'assign, allocate'
esplikatu 'explain'
esplotatu ("ustiatu") 'exploit'
esportatu 'export'
estandarizatu 'standardise'
esterilizatu 'sterilise'
estimatu 'esteem, consider'
estreinatu 'use or perform for the first time'
estudiatu 'study'
etsaminatu 'examine'
?*faboratu* 'favour'
fabrikatu 'manufacture'
fidatu 2 'entrust'
filmatu 'film'
formalizatu 'formalise'
formulatu 'formulate'
fotografiatu 'photograph'
frijitu 'fry'
fundatu 'found'
fusilatu 'kill (by a firing squad)'
gaztigatu 2 ("zigortu") 'punish'
gillotinatu 'guillotine'
gobernatu 1 'govern'
gomendatu 1 'recommend'
gomitatu 'invite'
gonbidatu 'invite'
grabatu 'record'
gurutzefikatu 'crucify'
heredatu 'inherit'
hidratatu 'hydrate'
hipotekatu 'morgage'
identifikatu 'identify'
imajinatu 'imagine'
imitatu 'imitate'
inauguratu 'inaugurate'
inportatu 'import'
inposatu 'impose'
inprimatu 'print'
instalatu 'install'
interpretatu 'interpret'
intsentsatu 'incense'
izkiriatu 'write'
jujatu 'judge'
juzgatu 'judge'
kabitu 2 'fit, have room'
kalibratu 'calibrate'
kalifikatu 'qualify'
?*kalitu* 'stone'
kalkatu 'stuff, cram, press'
kalkulatu 'calculate'
kanonizatu 'canonise'
kantatu 'sing'
kapitalizatu 'capitalize'
?*karesatu* 'caress'
kardatu 'comb (wool)'
?*kartzelatu* 'imprison'
kastigatu 'punish'
katalizatu 'catalyse'
katalogatu 'catalogue'
kausatu 'cause'
?*kitatu* 'pay up, settle accounts'
koblatu 'compose (verses)'
kobratu 'charge, receive (payment), recover'
kolektibizatu 'collectivize'
kolonizatu 'colonise'
?*koloratu* 'colour'
konbidatu 'invite'
konbinatu 'combine'

konektatu 'connect'
konfirmatu 1 [*h. Berretsi*] 'confirm'
konfiskatu 'confiscate'
konfitatu 'preserve, sweeten'
konparatu 'compare'
konposatu 'compose'
konprenitu 'understand, comprehend'
kontatu 'count, narrate, tell'
kontrolatu 'control'
kontsakratu 'consecrate'
?*kontseilatu* 'advise, counsel'
kontsideratu 'consider'
kontsultatu 'consult'
kontsumitu 'consume'
kopiatu 'copy'
koplatu 'compose, sing or recite verses'
kreatu 'create'
kritikatu 'criticise'
kromatu 'electroplate (with chromium)'
laboratu 'farm, till (the soil)'
letreiatu 'spell'
libratu 1 'free, liberate'
lisatu 'iron'
lixibatu 'bleach'
lorifikatu 'glorify'
manatu 'command, order'
maneatu 2, 3 'till (soil), season (food)'
maneiatu 'handle, use, operate'
martirizatu 'martyrize'
masakratu 'massacre'
mastekatu 'chew'
mespretxatu 'scorn, slight'
metrailatu 'machine-gun'
molestatu 'bother, annoy'
monopolizatu 'monopolize'
muntatu 'assemble, achieve'
?*musikatu* 1 'put to music'
nazionalizatu 'nationalize'
neutralizatu 'neutralize'
obratu 'bring about, do'
ofenditu 'offend'
ofentsatu 'offend'
ordenatu 1 ("antolatu") 'order, organise'
organizatu 'organise'
orkestratu 'orchestrate'
pagatu 'pay, pay for'
paratu 'put'
partitu 1 'divide'
patentatu 'patent'
pausatu 2 'put, place'
pentsatu 'think'
pertsegitu 'pursue, harass, persecute'
pintatu 'paint'
pipatu 'smoke'
pisatu 'weigh'
plagiatu 'plagiarise'
planteatu 'raise, pose (a problem, question)'
plantxatu 'iron'
polikopiatu 'run off, make multiple copies (of)'
polimerizatu 'polymerize'
polinizatu 'pollenize'
populatu 'populate'
praktikatu 'practise'
predikatu 'preach'
prentsatu 'press'
presentatu 'present, introduce'
probatu 'try, test, prove'
produzitu 'produce'
profanatu 'profane'
profetizatu 'prophesy'
profitatu 'profit, benefit'
programatu 'programme'
proiektatu 'project, plan'
promestu 'promise'
prometatu 'promise'
proposatu 'propose, suggest'
publikatu 'publish, make public'
purgatu 'purge'
pusatu 'push'
saneatu 'clean, disinfect, drain'
seduzitu 'seduce'
segatu 'cut (with a scythe)'
seinalatu 'point out, indicate'
sentitu 1 'feel'
setiatu 'besiege, lay siege (to)'
sinatu 'sign'
sinplifikatu 'simplify'
sintetizatu 'synthesize'
?*soberatu* 2 'be left over, be too much'
sobratu 2 'be left over, be too much'
sokorritu 'come to (someone's) aid'
soldatu 'solder'
solfeatu 'read (music)'
sostengatu 'sustain, maintain'
sozializatu 'socialize'
sufritu 'suffer'
tapizatu 'upholster, carpet'
tentatu 'tempt'
tindatu 'dye, paint'
tiranizatu 'tyrannize'
tiratu 'throw; draw close, pull'
tormentatu 'torture, torment'
torturatu 'torture'
tragatu 'swallow, gulp'
transkribatu 'transcribe'
transmititu 'transmit'
tratatu 'treat'

trazatu 'plan, design, outline'
urbanizatu 'urbanize, develop, build'
urkatu 'hang'
xukatu 'suck'
zentralizatu 'centralize'
zentsuratu 'censor'
zimendatu 'cement, provide a foundation for'
zirkunzidatu 'circumcise'
zitatu 'cite, quote'
zundatu 'sound, probe'
zurratu 'beat, give a beating (to)'

Most transitive verbs that only take *du* do not have a causative meaning, and come from verbs not used reflexively in the source language (or in any case Basque has not taken over any such reflexive use), e.g. Sp. *adorar* 'adore', *avisar* 'notify, warn', *apuntar* 'aim (at), point out, note', *amortizar* 'pay off', *analizar* 'analyse', *archivar* 'file away', *exportar* 'export', *estrenar* 'use or perform for the first time'. (At a formal level, reflexive verbs in Romance take an object-pronoun clitic agreeing with the subject, or *se* for the third person and in citation forms. Semantically, true reflexivity is only one of the notions denoted by such verbs; a great many have intransitive or medio-passive meanings. Basque possesses a true reflexive construction formed analytically and analogous to English *myself*, etc., but has nothing in its verbal morphology corresponding directly to Romance languages' formal reflexives. Regarding causatives, see 3.2.1. below.)

In some cases *HB* only lists the major sense of a borrowed verb. *aprobetxatu* 'take advantage (of), make use of', from Sp. *aprovechar*, is listed as a *du*-type verb (taking an ergative subject), showing the most frequent use in Basque, but the case of the subject of *aprobetxatu* in the following example demonstrates, even though the verb is in a non-finite form without an auxiliary, that there are also occurrences in which, reflecting the verb's reflexive character in another Spanish usage (Sp. *aprovecharse*), it may be conjugated with *da* in Basque (with an absolutive subject):

(3) APROBETXATU 'take advantage of':

Zer nai dezu, Jauna, nigandik? O! Badakit, Zuk zer nai-zun! Ni zure odolaz *aprobetxatu*, ta nik Zu amatzea. (*OEHtc*: Cardaberaz, *San Ignacioren ejerzizioak*, 64).
'What do you want from me, Lord? Ah! I know what you want! That I [absolutive] should profit from your blood, and that I [ergative] should love you.'

Some verbs given in *HB* as *du*-types do admit causative alternation, and would be more appropriately labeled as *da/du*-type verbs:

(4) a. AGLUTINATU 'agglutinate': [*EH: da/du* vb.] (reflexive use: Sp. *aglutinarse*)
b. SENTITU 'feel': *haurdun sentitu zen* 'she felt pregnant' (Sp. *sentirse*)

3.1.2. *Transitives with absolute uses: transitive/unergative alternation* (estudiatu, eskribitu...)

As already mentioned, certain transitive verbs, i.e. those which prototypically take a direct object, at times accept unergative constructions with no specified direct object:

TRANSITIVES WITH ABSOLUTE USES:

aprobetxatu 'take advantage'
apuntatu 1 "keinatu, destatu (arma bat)" 'take aim (with a weapon)'
desobeditu 'disobey'
diktatu 'dictate'
errezatu 'pray'
eskribitu 'write'
estimatu 'estimate'
estudiatu 'study'
izkiriatu 'write'
kantatu 'sing'
kobratu 'charge'
laboratu 'farm, till the soil'
meditatu 'meditate'
pagatu 'pay'
pipatu 'smoke'
predikatu 'preach'
segatu 'cut with a scythe'
sentitu 'feel, feel sorry'
sinatu 'sign'
sufritu 'suffer'
tiratu 1 "jaurti" 'throw'
tiratu 2 "nork beregana *erakarri*" 'draw close, pull'

A couple of examples showing unergative uses follow:

DIKTATU 'dictate':

Ikasleentzat ez da lagungarria hitzez hitz *diktatzea*; prozedura honek, izan ere, testa baliogabetu egiten du. (*EEBS*: 1968-1990; Euskara Batua; Saiakeraliburuak; U. Larramendi; Orria = 0049; Paragrafoa = 007).

'It is not helpful to pupils to dictate [no object] word by word; indeed, this procedure invalidates the test.'

ERREZATU 'pray':

Musulmanek egunean bost aldiz *errezatzen dute*, etxean edo meskitan. (*EEBS*: 1991; Euskara Batua; Ikasliburuak; J. Ossa; Orria = 0040; Paragrafoa = 032).

'Muslims [ergative case] pray [*du*-type aux.] five times a day, at home and at the mosque.'

3.1.3. Du-*type intransitives: unergatives* (abdikatu, abortatu, abusatu...)

A few years ago Sarasola (1977: 79), discussing the trend towards unergative patterns observed in present-day Basque, voiced the following opinion about the possible future of Basque's morphological system:

> (5) Volviendo a los verbos de tipo *iraun, irakin,* su número aparte de ser importante, aumenta continuamente. Ya nos hemos referido anteriormente a que verbos formados a base de nombre más verbos como *egin,* etc., están pasando a la misma situación. Por otra parte, los dobles de procedencia románica de verbos como *jarraiki* 'seguir', *desagertu* 'desaparecer', es decir, *segitu, desaparezitu,* poseen la misma característica de falta de objeto. Así tenemos *Jonek Mikeli segitu dio* frente a *Jon Mikeli jarraikitu zaio* 'Juan ha seguido a Miguel', y *Jonek desaparezitu du* frente a *Jon desagertu da* 'Juan ha desaparecido'.[3] Parece así que existe una ten-

[3] He adds the following footnote: «Leemos en Juan Bautista Aguirre, escritor vasco fallecido en 1823, en el primer tomo de sus *Eracusaldiac,* pag. 41, *saltatu da,* donde actualmente la mayoría de vascos diría, *saltatu du, salto egin du* 'ha saltado'.» [In the first volume of the *Eracusaldiak* of Juan Bautista Aguirre, a Basque writer who died in 1823, we read (p. 41) *saltatu da* 'jumped', whereas today most Basques would say *saltatu du* or *salto egin du*].

dencia en vasco actual a sentir todos los verbos (menos los que designan rigurosamente un estado) como transitivos, es decir, necesitados de elemento ergativo. Esta tendencia se realiza más fácilmente en préstamos, al no sentirse el peso de la tradición lingüística. En consecuencia no sería, quizás, excesivamente aventurado establecer la previsión de que podría llegar el día en el que el sujeto de todos los verbos vascos presentara la marca de ergativo, y la división transitivo/ intransitivo se realizara según tuviera o no sentido para un verbo determinado la noción de objeto. Estaríamos así, utilizando la terminología de Fillmore (1968: 14), ante el paso de un sistema "ergativo" (como el vasco actual) a un sistema "acusativo" (sistema indoeuropeo).

Translation: Coming back to verbs of the *iraun* ['last'], *irakin* ['boil'] type, their number is not only significant but rising all the time. We have already mentioned above that verbs formed from a noun plus a verb such as *egin* ['make, do'] etc. show a similar tendency. Moreover, equivalents of Romance origin for verbs such as *jarraiki* 'follow', *desagertu* 'disappear', namely *segitu, desaparezitu,* have the same characteristic of lack of an object. Thus we have *Jonek Mikeli segitu dio* by the side of *Jon Mikeli jarraikitu zaio* 'Jon followed Mikel', and *Jonek desaparezitu du* by the side of *Jon desagertu da* 'Jon has disappeared.'[3] Thus there appears to be a trend in present-day Basque to perceive all verbs (except those that strictly denote a state) as transitives, that is, requiring an ergative argument. This trend is manifested most easily in loans, in the absence of the weight of linguistic tradition. In consequence it would perhaps not be too risky to forecast that the day may come when the subjects of all Basque verbs will be in the ergative, and the transitive/intransitive distinction will depend on whether or not the notion of an object makes sense for a given verb. Thus we would be witnessing, to use Fillmore's (1968: 14) terminology, the passage from an "ergative" system (such as present-day Basque) to an "accusative" one (the Indo-European system).

We wish to clarify this putative trend towards unergativity in this subsection. First let us see which verbs have been borrowed into Basque as unergatives. The following is a tentative list of verbs appearing in *HB* that have been borrowed as unergatives in at least one of their senses:[4]

(6) LIST OF UNERGATIVE BORROWED VERBS:

abdikatu 'abdicate'
abortatu 'have a miscarriage/an abortion'
abusatu 'go too far, abuse'
arnegatu 'renounce, abnegate'
blasfematu (?) 'blaspheme'
bogatu 'row'
deklaratu 'declare, testify'
desertatu 'desert'
dudatu 'doubt'
emigratu 'emigrate'
ernegatu 1 "arnegatu" 'abnegate'

[4] It is worth noting that only in a few cases —*eskiatu* 'ski', *funtzionatu* 'work (as of a machine)'— does *HB* recognise a verb's unergativity through the observation 'no absolutive case'. Owing to the scarcity of this syntactic information it is not easy to list all the unergative borrowed verbs. Sometimes, moreover, a borrowed verb only acts unergatively in one of its senses: *deklaratu* 'declare', for instance, is only used unergatively in the technical legal sense of 'testify, make a declaration': *hiru lagunek deklaratu zuten atzo epaitegian* 'three people [ergative case] testified [*du*-type aux.] in court yesterday'.

erratu (?) 'err'
erregutu 'pray, beg, appeal'
erreinatu 'reign'
erreparatu 2 "ohartu" 'notice'
eskiatu 'ski, go skiing'
esperimentatu "esperimentua egin" 'experiment'
faltatu 1 "huts egin" 'miss, fail'
funtzionatu 'work, function'
komuniatu 'take communion'
korritu 'run'
kurritu 'run'
kotizatu 'have a quoted price'
manifestatu Nor. "manifestazioan parte hartu" 'demonstrate, protest'
profetizatu (?) 'prophesy'
protestatu 2 "arbuioa adierazi" 'protest'
segitu 'continue, follow'
topatu 2 "topa egin" 'drink (a toast) to'
toreatu 'bullfight'
trafikatu 'traffic (in something)'
tratatu 2 "harremanak izan" 'deal, do business'
zirkulatu 'circulate'

The subclass of borrowed unergative intransitives seems to be associated with the agentive character of a verb's meaning. Let us recall Levin and Rappaport's (1995) suggestion that agentive unergative verbs constitute a subset of verbs expressing internal causation.[5]

(7) INTERNALLY CAUSED EVENTUALITIES (Mendikoetxea 1999: 1597)

AGENTIVE (UNERGATIVE)	NONAGENTIVE
jugar 'play', *reír* 'laugh', *hablar* 'speak'	a) physical or emotional reaction: *temblar* 'tremble'. b) verbs of emission: *brillar* 'shine', *chirriar* 'squeak, creak', *apestar* 'stink', *emanar* 'emanate' (unergatives) c) internally caused change of state: *florecer* 'flower, flourish', *envejecer* 'get old, age' (unaccusatives)

[5] Many unaccusative verbs denote a change of state or change of place. Such verbs express causative events and may be of two types: some can be said to express 'external causation' as part of their meaning, others 'internal causation'. Thus beside the Spanish transitive construction {*Juan/el viento/la llave*} *abrió la ventana* (1a) '{Juan/the wind/the key} opened the window' we find the unaccusative construction *La ventana se abrió* (1b) 'The window (was) opened'; we might say that (1b), like (1a), implies external causation which can be considered the source of the process denoted by the verb, but which in (1b) is not expressed. In (2) on the other hand, *El rosal floreció* 'The rose tree bloomed', the event referred to has 'internal causation': the process of "blooming" results from an intrinsic property of the syntactic subject, since only things having the necessary characteristics are able to bloom, even though external factors (springtime, fertilizers) may also be required to trigger the process.

(1) a. External causation (transitive construction): *{Juan/el viento/la llave} abrió la ventana.*
b. External causation (unaccusative construction): *La ventana se abrió.*
(2) Internal causation (unaccusative construction): *El rosal floreció.*

Most of the borrowed unergative verbs listed in (6), such as *abdikatu* 'abdicate', *abortatu* 'have a miscarriage, have an abortion', *abusatu* 'abuse, go too far', *arnegatu* 'renounce, abnegate', *blasfematu* 'blaspheme', *bogatu* 'row', *deklaratu* 'declare, testify', *desertatu* 'desert', *dudatu* 'doubt', *emigratu* 'emigrate', etc., are typical agentives. Most such events relate to an individual's will and are controlled by "someone", and that "someone" is usually [+human] or at least [+animate]. Thus agentivity seems to be an important semantic feature determining the type of borrowed intransitive verbs in Basque. The transitivity of some verbs that are not completely assimilated but nonetheless occur frequently, particularly in spoken Basque, such as the following, appears to be determined by this agentivity feature: *almortzatu* 'have lunch', *bozkatu* or *botatu* 'vote', *debutatu* 'make one's début', *dimititu* 'resign', *erreakzionatu* 'react', *insistitu* 'insist', *kolaboratu* 'collaborate', *meriendatu* 'have an afternoon snack', etc.

However, a few borrowed verbs in (6) such as *funtzionatu* 'work (as of a machine etc.), function', *kotizatu* 'have a quoted price, sell (at)', *zirkulatu* 'circulate', are non-agentive and require another explanation. Here we may invoke the hypothesis of grammaticalisation of verbs of internal causation (Mendikoetxea 1999: 1602-5; Levin & Rappaport 1995: 136, 146). The verbs *funtzionatu, kotizatu* and *zirkulatu* express "internally caused eventualities": when we say *makina horrek funtzionatzen du* 'that machine works', the machine is understood to possess some property or characteristic which makes it work, i.e. some property of the argument itself is considered to be the cause of the event. Thus the notion of internal causation is involved in determining the type of some borrowed intransitive verbs, and some verbs expressing internal causation are borrowed into Basque as unergatives. This tendency seems to be most noticeable in examples such as the following, typical of spoken Basque: *autoak ez du arrankatzen* 'the car [erg.] doesn't/won't start', *pilotak txarto botatu du* 'the ball [erg.] bounced the wrong way ("misbounced")', etc.[6]

In addition to the aforementioned semantically determined pattern or tendency whereby verbs of agentivity or internal causation are incorporated into Basque as unergatives, I believe we should also take into account a syntactic criterion involving the source of the borrowed verb. According to this hypothesis, only verbs that are non-reflexive intransitives in Spanish or French (or non-reflexive uses or senses of verbs), such as Sp. *abdicar* 'abdicate', *abortar* 'abort', *abusar* 'abuse', *renegar* 'renounce', *blasfemar* 'blaspheme', *bogar* 'row', *declarar* 'declare, testify', *desertar* 'desert', *dudar* 'doubt', *emigrar* 'emigrate' are candidates for borrowing into Basque as unergatives. However, reflexive verbs such as Sp. *atreverse* 'dare', *burlarse* 'make fun (of)', *descuidarse* 'neglect', *divertirse* 'have fun', *fiarse* 'trust', *quejarse* 'complain', *conformarse* 'be satisfied (with)', etc., are not. It would appear that in these languages the reflexive syntactic pattern signals either unaccusative constructions or reflexive meaning and is related to the expression of causativity (Mendikoetxea 1999: 1603). Reflexive verbs tend to be borrowed into Basque with the *da*-type auxiliary, e.g. *atrebitu da* 'dared',

[6] *Bibratu (du)* 'vibrate', a verb used in physics listed in the Elhuyar (2000) dictionary, would also belong here.

burlatu da 'made fun, derided', *deskuidatu da* 'neglected, didn't bother' (cf. 3.3 below).[7]

Problems and uncertainties in Basque arise in cases such as Sarasola's example of *saltatu da* 'jumped', in older usage, versus *saltatu du* 'ditto', the newer tendency. As it happens, the verb "jump" in other languages —Sp. *saltar,* Fr. *sauter,* Eng. *jump*— is a standard example of an unergative verb. Here are a few of the borrowed verbs whose transitivity is subject to uncertainty and vacillation in present-day Basque:

(8) a. DANTZATU 'dance' (*HB: da/du* vb.: *arratsalde osoan dantzatu naiz* (not **dantzatu dut*) 'I danced [*da*-type] all night'; *aurreskua dantzatu dut* 'I danced [*du*-type] the *aurresku* [a Basque dance]'*; makila dantzatu dut* 'I did the *makila dantza* [the "stick dance"]', literally 'I danced [*du*-type] the stick').
b. DESKANTSATU 'rest' (*HB:* 1 *da* vb. syn. *atseden hartu* 'have a rest'. 2 *Nor. da/du* vb. syn. *lasaitu* 'calm down').
c. KOMULGATU, KOMEKATU 'take communion' (*HB: da* vb.) # KOMUNIATU (*HB: du* vb.; *EH:* formerly *da).*
d. OLGATU 'have fun' (*HB: da/du* vb.).
e. SALTATU 'jump' (*HB: da* vb. syn. 'salto egin' 'ditto').
f. NABIGATU 'sail, navigate' (*HB: da/du* vb.).
g. PASEATU 'go for a walk or ride' (*HB: da/du* vb.: *karrozan paseatu ziren* 'they went for a ride [*da*-type aux.] in a carriage'*; bere astoa paseatzen du* 'he takes his donkey for a walk [*du*-type aux.]').

As Perlmutter (1978: 164) observes, the most problematic cases are verbs of movement, such as (among others) *saltatu, nabigatu* and *paseatu* in the preceding list. In accordance with Mendikoetxea's (2000: 1606-7) study of Italian data, agentive verbs denoting manner of movement such as *correr, saltar, navegar, pasear* appear to be unergative in Spanish, but past and present Basque usages involving such verbs vary.

Sarasola's *E[uskal] H[iztegia]* lists *nabigatu* 'sail, navigate' as either a *da*-type or a *du*-type verb, but that doesn't make it a *da/du*-type; and similarly for several other verbs listed in the same dictionary – *apelatu* 'appeal', *komuniatu* 'take communion', *emigratu* 'emigrate', *eskatu* 'request, demand, ask (for)', *itsastatu* 'sail', where the same shift from *da*-type to *du*-type can be observed. Hence the indication *da/du* that *HB* gives for *nabigatu* is potentially misleading, referring not to the usual causative alternation seen in ordinary *da/du* verbs such as *animatu* 'cheer up (intr.), become emboldened etc./cheer up (tr.), encourage', but merely to the verb's uncertain transitivity in the present-day language.

---

[7] *(Bilboko kaleetan) manifestatu zuten* 'they demonstrated (on the streets of Bilbao)', an example of northern Basque usage with a *du*-type auxiliary listed in *HB,* evidently reflects the French verb *manifester* which is non-reflexive in this use, while the southern Basque usage which would require *manifestatu ziren* here, with a *da*-type auxiliary, corresponds to the Spanish *manifestarse,* which is reflexive. Such contrasts are rare because assignments of reflexivity in French and Spanish generally coincide. This example thus provides useful corroboration of the influence of Romance transitivity on Basque loans.

For the verbs *saltatu* 'jump' and *paseatu* '(go for a) walk or ride', the literary tradition strongly favours use with *da*. The treatment of *paseatu* as a *da*-type, incidentally, is supported by the Spanish verb's reflexive use (*DRAE,* 'pasear'). Yet as Sarasola himself asserts and *EEBS* confirms, many present-day speakers treat both as unergatives. Although *HB* leans towards the older usage for both of these verbs, their shift to unergatives (such as Sarasola acknowledges for *apelatu, komuniatu, emigratu, eskatu* and *itsastatu*) seems not to be an entirely novel development.[8]

The doublet *komuniatu (du)* 'take communion' ~ *komulgatu (da)* 'ditto' is curious. The phonology reveals the former to derive from a French source and the latter from Spanish. Now in the modern languages, neither French *communier* nor Spanish *comulgar* is reflexive. Possibly the use of *komulgatu da* is explained by the fact that Spanish *comulgar* could formerly be used reflexively (cf. *DRAE,* 'comulgar'). As the following quotation from Mendiburu demonstrates, at one time Basque *komulgatu* admitted causative alternation as did its Spanish counterpart:

(9) Eta Jangoiko berak biztu zuen S, *Eloi, *komulga zęzan* [1] ilzeko beldurrez horduraño *komulgatu etzen* [2] eri gaiso bat. (*OEHtc:* Mendiburu, *Idazlan argiragabeak,* 1. lib., 188).
'And the same God brought S to life, Eloi, in order that a sick patient fearing death who had not been administered communion [*da*-type aux.] [2] until then might take communion [*du*-type aux.] [1].'

Olgatu 'have fun, amuse oneself' was in older tradition almost exclusively of the unaccusative *da*-type. Significantly, the obsolescent Spanish equivalent *holgar* had reflexive uses in some senses (in the senses "alegrarse de una cosa" ['be pleased about something'] and "divertirse, entretenerse con gusto en una cosa" ['entertain or amuse oneself with something'], cf. *DRAE*). *EH* and *HB* give *olgatu* as having *da/du* alternation based on a single example: *Leengo eguneko zure berbeldijak guztiz asko olgau ninduban. OEHtc:* Uriarte, *Euskal Dialogoak,* 3 'Your conversation of the other day amused me [*du*-type aux.] greatly'. In my opinion this is an isolated, idiosyncratic use; on the whole this verb does not have causative alternation. On the other hand, some present-day speakers have abandoned the traditional usage and employ the verb as an unergative. The verb *deskantsatu* 'rest' has a similar story.

Thus the verbs listed in (8), either owing to reflexive use in the source language or for some other reason, have for the most part been interpreted as *da*-type verbs in the literary tradition. Many present-day Basque speakers, on the contrary, treat such verbs as unergatives. The unergative trend seems to me not to be universal, but largely restricted to verbs that are not reflexive today in the source language and may be considered agentives, such as *saltatu* 'jump', *olgatu* 'have fun', *dantzatu* 'dance', *paseatu* 'go for a walk or ride', etc.

---

[8] This is seen in Leizarraga: "Begira zaitezte, Baldin hire anaiek hire kontra *saltatu badu,* reprehendi ezak hura, eta baldin emenda badadi barkha iezok" (*OEHtc:* Leizarraga, *Testament berria,* 139). 'Beware, if thy brother [ergative] hath jumped [*du*] at thee, reprimand him, and if he calmeth down, forgive him]'.

So a Basque speaker should have no trouble with a Spanish verb like *entrenar* 'train (tr.)'/*entrenarse* 'train (intr.)' that undergoes causative alternation, as seen in (10a-b). Integrated into Basque, this will become a *da/du* verb as in (11). If however a Basque-Spanish bilingual habitually says in Spanish *Los jugadores entrenan* as in (10c), rather than *se entrenan* as in (10b), for 'The players train', as is increasingly common (cf. Torrego 1998: 22, 38-39), he may vacillate concerning the verb's transitivity when borrowing it into Basque: see (12).[9]

SPANISH

(10) a. El profesor *entrena* a los alumnos. [not reflexive] 'The teacher trains the pupils.'
b. Los alumnos *se entrenan.* [reflexive] 'The pupils train.'
c. Hugo Sánchez *entrenó* solo durante una hora. 'Hugo Sanchez only trained for one hour.' (Exceptional use, equivalent to *...se entrenó...*).

BASQUE

(11) a. Irakasleak bere ikasleak *entrenatu ditu.* [*du*-type auxiliary] 'The teacher trained his students.'
b. Ikasleak eskolan *entrenatzen dira.* [*da*-type auxiliary] 'The students train at school.'

(12) a. Gogor *entrenatu naiz* jokatu ahal izateko. [*da*-type auxiliary] 'I trained hard to be able to play.'
b. ?Gogor *entrenatu dut* jokatu ahal izateko [*du*-type auxiliary] 'ditto'

As already noted, the treatment of Spanish and French agentive intransitives[10] as unergatives in Basque appears to be a tradition of some antiquity. This is suggested by the well-established change of certain verbs like *apelatu* 'appeal', *komuniatu* 'take communion', *emigratu* 'emigrate', *itsastatu* 'sail, navigate' from *da* to *du* type, and also by the relative antiquity of the majority of the borrowed unergative verbs listed, such as *abusatu* 'abuse, go to far', *arnegatu/ernegatu* 'renounce, abnegate', *blasfematu* 'blasphemе', *erratu* 'err', *erregutu* 'pray', *erreinatu* 'reign', *erreparatu* 'notice' *faltatu 1* 'miss, fail', *komuniatu* 'take communion', *korritu* 'run', etc.

Some examples of unergative verbs follow:

---

[9] The following sentence appeared recently in the Basque-language newspaper *Euskaldunon Egunkaria* (2001-IX-13, p. 31): «Asko gara, baina titular izateko *entrenatzen dut*» 'There are many contestants, but I am training for the post' [unergative construction]. In an article about the auxiliaries used with certain verbs, Arrieta (2001: 19) claims that *entrenatu*, and also *eboluzionatu* 'evolve', take *du* as auxiliary.

[10] Certain non-agentive intransitives are also borrowed into Basque as unergatives, e.g. *AGUANTATU* 'hold out, endure': *Han geratuko dira, aguantatzen duen arte.* 'They will stay there as long as he holds out'. *TARDATU* 'take (of time)': *Asko tardatu zuen* 'He took a long time'): in such cases the argument is human or animate, but experiences the event rather than controlling it.

ABDIKATU 'abdicate' [*EH:* 1977, *nart.*]:

Erakundea, beraz, jaiotzen da, Isabel II.a tronutik bota ondoren eta Amadeo de Saboya-k *abdikatu* ondoren... (*EEBS:* 1991; Euskara Batua; Saiakeraliburuak; Artistas 1995; Orria = 0024; Paragrafoa = 002; abdikatu).

'Thus the institution was born after Isabella II was ousted from the throne and Amadeus of Savoy [ergative case] abdicated...'

BLASFEMATU 'blaspheme' [*EH:* 1571; *arch.*]:

Kaifas'ek itz oiek aditutakoan, soñekoak urratu zituen, eta deadar egin zuen: *Blasfematu du*; añenekoa bota du. (*OEHtc:* Lardizabal, *Testamentu zarreko edo berriko kondaira,* 447).

'When Caiphas had heard these words, he rent his clothes, and shouted: He has blasphemed [*du*]; he has uttered a curse.'

DEKLARATU 'testify' [*HB:* 1. epaiketaren inguruko adierarekin batez ere 'mainly in relation to court cases']:

Era berean, enpresako beste hiru buruk Durangoko Hirugarren Epaitegian *deklaratu zuten* atzo. (*EEBS*: 1991; Sailkatu gabeak; Egunkariak; Egunk 1991; Orria = 0008; Paragrafoa = 005).

'Furthermore, the directors [ergative case] of three companies testified [*du*-type aux.] in the Third Court in Durango yesterday.'

3.2. Borrowed *da/du* verbs *(alfabetatu, animatu...; defendatu...)*

HB lists these borrowed verbs as having alternation between *da* and *du* (in the following list, where intransitive and transitive glosses differ, the "*da*-meaning" precedes a slash and the "*du* meaning" follows it, e.g. *borratu* intr. 'vanish'/tr. 'erase'; where a single gloss or set of glosses is given below, its intransitive and transitive senses correspond to Basque uses with *da* and *du* respectively, e.g. *bainatu* 'bathe (intr./tr.)'):

BORROWED *DA/DU* VERBS

abiatu 'set out, start out/ set going, set in motion'
alfabetatu 'learn/teach literacy'
altxatu 'rise, get up/raise, lift up'
animatu 'cheer up (intr.), become emboldened etc./ cheer up (tr.), encourage'
armatu 'arm (i.e. acquire/ provide arms'
arrimatu 'come/bring close'
bainatu 'bathe'
beztitu '(get) dress(ed)'
borratu 'vanish/erase'
debaluatu '(be) devalue(d)'
degeneratu '(cause to) degenerate'
deklinatu 'decline (a noun etc.)'
desengainatu 'undeceive/ see the light'
despeditu 1 ("agur egin") 'take one's leave/send on one's way'
disolbatu Kim. 'dissolve'
entretenitu 'amuse (onself); dally/delay'
errendatu 'surrender'
errenditu 'surender'
erretiratu 'withdraw'
espantatu '(be) frighten-(ed)'
esposatu 'marry'
estonatu 'astonish'
formatu '(be) train(ed), (be) educate(d)'
fosildu 'fossilize'
funditu 'melt'
gastatu 'wear out'

gorde 'hide'
gozatu 'become sweet/sweeten'
helenizatu '(be) hellenize(d)'
immunizatu '(be) immunize(d)'
industrializatu 'industrialize'
informatu 'find out/inform'
ionizatu '(be) ionize(d)'
jiratu 'turn, rotate'
juntatu 'join'
justifikatu 'be justified/justify'
kanbiatu 'change'
kargatu '(become) load(ed), charge(d)'
konbertitu '(be) convert-(ed) into; turn into'
kondenatu '(be) condemn-(ed), condemn (oneself)"
kondentsatu 'condense'
konfesatu 'confess'
konfirmatu 2 *Erl.* '(be) confirm(ed) (religious ceremony)'
konplitu 'come/bring to completion, (be) fulfill(ed)'
kontentatu 'become/make happy'
kontserbatu '(be) conserve(d)'
kontsolatu '(be) console(d)'
kontzentratu '(be) concentrate(d)'
koordinatu '(be) coordinate(d)'
kurbatu '(become) curve(d), bend'
makillatu 'put make-up on (oneself)'
maneatu 1 "atondu" '(be) arrange(d), prepare(d)'
mantendu 'stay/keep'
matrikulatu 'sign up'
mudatu 'change'
normalizatu '(be) normalize(d)'
obligatu 'force (oneself)'
ordenatu 2 '(be) ordain(ed)'
orientatu 'orientate (oneself)'
oxidatu '(cause to) rust'
oxigenatu 'oxygenate'
palatalizatu 'palatalize'
penatu '(be) trouble(d)'
perfekzionatu 'improve, perfect (oneself)'
plegatu 'fold'
presatu 'hurry'
profesionalizatu 'professionalize'
sakrifikatu 'devote oneself/sacrifice'
salbatu '(be) save(d)'
santifikatu 'be sanctified/sanctify'
sikatu 'dry'
sosegatu 'calm down'
tapatu '(be, become) cover(ed), obstruct(ed)'
tatuatu 'get (tattoo)ed'
trabatu '(become) block(ed)'
transformatu '(be) transform(ed)'
traumatizatu '(be) traumatize(d)'
tronpatu '(be) deceive(d), confuse(d)'
umiliatu '(be) humiliate(d)'
zibilizatu '(become) civilize(d)'

3.2.1. *Causative alternation: transitive/ergative opposition* (sikatu du/sikatu da)

The verbs presently under consideration are transitive and have causative meanings:

(13) CAUSATIVE:

a. Mirenek zorua *sikatu du.* 'Miren [erg.] dried [*du*] the floor.'
b. <Miren caused> zorua *sikatu da* 'the floor [abs.] dried [*da*] (spontaneously)'
c. Zorua *sikatu da* 'The floor [abs.] dried [*da*].' (anti-causative or ergative variant: "spontaneously")

These are intrinsically transitive verbs with external causation, whose unaccusative variant (13c) results from decausitivization, where the notional subject —the external cause of the event denoted by the verb— is not expressed. Only some causative verbs —those denoting an event that is able to occur spontaneously, without an agent's

deliberate participation— permit the unaccusative variant. A causative such as *akabatu* 'slaughter, murder, kill (of an animal, or with violence)', for instance, does not permit causative alternation, for the causative sub-event is highly characterised: **Otsoak akabatu dira (berez)* '*The wolves [abs.] slaughtered [*da*-type aux.] (by themselves, spontaneously)'. Only causative verbs whose causative sub-event is unspecified in the verb lexeme's semantic structure admit an unaccusative variant (Mendikoetxea 1999: 1591).

In the source languages, verbs with causative alternation are clearly identified syntactically by the use of a clitic reflexive pronoun for the non-causative sense: Sp. *alzar* 'raise'/*alzarse* 'rise', *animar* 'cheer, cheer up (tr.)'/*animarse* 'cheer up (intr.)', *arrimar* 'bring close'/*arrimarse* 'come close', *borrar* 'erase'/*borrarse* 'vanish', etc. Torrego (1998: 21) makes the following insightful comment on verb-pairs with causative alternation in Spanish:

> (14) En efecto, si tomamos como referencia los casos de 3.1.1 es evidente que los verbos *dormir, levantar, separar* no significan lo mismo con el pronombre que sin él. Así, en *Juan durmió al niño* hay un significado *factitivo* parecido al de "Juan hizo que el niño durmiera". No es éste el significado de "Juan se durmió", pues no es desglosable en *"Juan hizo que Juan se durmiera" o en *"Juan se hizo dormir". Eso quiere decir que una posible estructura profunda de *Juan se durmió* nunca podría ser la de "Juan durmió a Juan": por tanto, difícilmente podríamos considerar a *se* (al pronombre) con función nominal de 'c. directo', pues no es equivalente a *lo,* ya que en *Juan lo durmió* aparece el primer significado de *dormir.* Una vez más, el significado como control es necesario para valorar los resultados de una conmutación. La conclusión parece clara: en nuestra lengua hay un verbo *dormir* y un verbo *dormirse,* un verbo *levantar* y un verbo *levantarse,* un verbo *separar* y un verbo *separarse,* etc. Y en los diccionarios deberían aparecer como entradas diferentes.
>
> *Translation:* Indeed, in reference to the examples in 3.1.1 it is clear that the verbs *dormir* 'sleep', *levantar* 'lift', *separar* 'separate' do not have the same meaning with and without the [reflexive] pronoun. Thus *Juan durmió al niño* 'Juan made the child sleep', 'J. sent the child to sleep' has a *factitive* meaning equivalent to "John caused the child to sleep". This is not the meaning of "Juan se durmió" 'Juan fell asleep', for this cannot be broken down into *"John caused John to sleep" or *"John made himself sleep". Thus the deep structure of *Juan se durmió* cannot possibly be "Juan durmió a Juan" 'John made John sleep': so we can hardly consider *se* (the [reflexive] pronoun) as having the nominal function of 'direct object'; it is not equivalent to *lo* 'him', since in *Juan lo durmió* 'John made him sleep' the first meaning appears. Once more, the control meaning is needed to evaluate the results of a commutation. The conclusion seems clear: in our language there is a verb *dormir* and a verb *dormirse,* a verb *levantar* and a verb *levantarse,* a verb *separar* and a verb *separarse,* etc., and in dictionaries these ought to appear as separate entries.

A few observations remain to be made. Although most borrowed verbs in Basque with *da/du* alternation have equivalents in the source language admitting reflexivization, that is not always so. Sometimes such an alternation has developed within Basque itself. There are no reflexive verbs in the source language corresponding to Basque un-

accusative uses of *deliberatu* 'decide, determine', *deskantsatu 2* 'rest', *ernegatu* 'abnegate, deny', *eskarmentatu* 'learn (from experience)', *finitu* 'finish', or *pasatu* 'pass', i.e. there is no such thing as Spanish **deliberarse*,[11] **renegarse, *escarmentarse*, French *se finir*, etc.

In the case of some of the verbs of the *da/du* type listed in *HB*, the causative alternation results not from decausitivization but from a reverse process. Such verbs are basically unaccusatives that are occasionally used transitively. Verbs such as *abiatu* 'set out, start out', *paseatu* 'go (/take) for a walk or ride', and *dantzatu* 'dance' are intransitives that occasionally permit a marked transitive use, not decausitivized or detransitivized transitives.

3.2.2. *Transitive/reflexive alternation:* obligatu du/obligatu da

In the case of some of the verbs that are labeled as *da/du* in *HB*, such as *armatu* 'arm (oneself)', *beztitu* '(get) dress(ed)', *makillatu* 'put make-up on (oneself)', *obligatu* 'force (oneself)', the choice is not between a causative verb and its inchoative variant, but rather between a transitive verb and its reflexive counterpart. This is because some Basque verbs express reflexivity by adopting the *da*-type auxiliary (cf. Etxepare 2003):

(15) DA/DU? > 'TRANSITIVE/REFLEXIVE' (DU > DA) ALTERNATION

a. ARMATU 'arm (oneself)'
b. BEZTITU '(get) dress(ed)'
c. MAKILLATU 'put make-up on (oneself)'
d. OBLIGATU 'force (oneself)'

It ought to be noted that there are some verbs in *HB* labeled as only conjugating with *du*, such as *abandonatu* 'abandon', *gobernatu* 'govern', *libratu* 'free, liberate', *defendatu* 'defend', *engainatu* 'deceive, trick', *konparatu* 'compare', which do permit *du/da* alternation to express reflexivity:

(16) 'TRANSITIVE (*DU*) > REFLEXIVE (*DA*)' ALTERNATION:

a. ABANDONATU 'abandon (oneself)'
b. GOBERNATU 'govern (oneself)'
c. LIBRATU 'free, liberate (oneself)'
d. DEFENDATU 'free, liberate (oneself)'
e. ENGAINATU 'deceive, trick (oneself)'
f. KONPARATU 'compare (oneself)'

3.3. Borrowed verbs that only conjugate with da: "pure" unaccusatives *(burlatu, konformatu...)*

Some of the verbs labeled as only conjugating with *da* in *Hiztegi Batua* are the following:

[11] However, in Basque *deliberatu* has acquired the meaning 'decide', and in Spanish *decidir* may be used reflexively: *decidirse*.

abusatu 2 (*Nor.* "jostatu") 'enjoy'
ailegatu 'arrive'
antsiatu 'worry'
atrebitu 'dare'
atrofiatu 'atrophy'
biziatu 'become addicted, degenerate'
burlatu 'make fun'
dedikatu 'devote oneself'
deskantsatu 'rest'
deskuidatu 'neglect'
dibertitu 'have fun, be entertained'
dibortziatu 'get divorced'
dutxatu 'have a shower'
errabiatu 'be(come) furious, get mad'
eskapatu 'escape, get away'
espezializatu 'specialize'
existitu 'exist'
faltatu 2 ("falta izan") 'be lacking'
federatu 'federate'
fidatu 'trust'
gobernatu 2 ("jokatu") 'behave, act'
herratu 'err'
interesatu 'be(come) interested'
kabitu 'fit, have room'
kanpatu 'camp'
katigatu 'get caught, stuck, tangled'
kexatu 'worry'
komekatu 'take communion'
komulgatu 'take communion'
konfederatu 'enter into a confederation'
konformatu 'be satisfied (with)'
kostumatu 'become accustomed (to), get used to'
libratu 2 (*coll.* "kaka egin") 'move one's bowels'
manifestatu 1 (*South*) 'demonstrate'
mutinatu 'rise up, rebel, mutiny'
paratu 'put'
partitu 'set out, depart'
pasatu 2 ("gertatu") 'happen'
pausatu 'perch, rest, pass away'
portatu 'behave, act'
prezatu 2 ("harrotu") 'be proud of (oneself)'
saltatu 'jump, leap'
soberatu 'be left over, be too much'
sobratu 'be left over, be too much'
tokatu 'correspond (to), devolve (upon), fall to'
trabailatu 'work'
usatu 'be(come) accustomed to, be or get used to'
zeinatu 'cross oneself'

Generally, then, 'other' intransitives —those which have no causative alternation and are not considered agentives— are borrowed into Basque as *da*-only verbs. The single argument of such verbs is the semantic 'theme' although it is syntactically the subject, as in (17a), and they do not allow causative alternation (17b):

(17) a. Mikel [THEMA] horretara atrebitu da. 'Mikel [abs.] dared [*da*] to (do) it.'
b. *Maisuak Mikel horretara atrebitu du. 'The teacher [erg.] dared [*du*] Mikel to (do) it.'

Syntactically, most borrowed verbs that only take *da* in Basque are reflexive in form in the source language. The verb may be intrinsically reflexive in the source language as in (18), or Basque may have borrowed the verb in reflexive uses (or mainly so), as in (19).

(18) INTRINSICALLY REFLEXIVE, OR USED REFLEXIVELY:

a. Fr. SE MUTINER/*mutiner → mutinatu (*da*) 'rise up, rebel, mutiny'
b. Sp. ATREVERSE/*atrever → atrebitu (*da*) 'dare'
c. QUEJARSE/*quejar → kexatu (*da*) 'complain'
d. CONFORMARSE # conformar → konformatu (*da*) 'be satisfied (with)'
e. FIARSE # fiar → fidatu (*da*) 'trust'
f. ANSIARSE # ansiar → antsiatu (*da*) 'worry'

g. PORTARSE # portar → portatu (*da*) 'behave'
h. SIGNARSE/? signar → zeinatu (*da*) 'cross oneself'

(19) BORROWED IN A (CHIEFLY) REFLEXIVE USE:

a. descuidar/DESCUIDARSE → deskuidatu (*da*) 'neglect, not bother'
b. divertir/DIVERTIRSE → dibertitu (*da*) 'have fun'
c. divorciar/DIVORCIARSE → dibortziatu (*da*) 'get divorced'
d. interesar/INTERESARSE → interesatu (*da*) 'take an interest'
e. ocupar/OCUPARSE → okupatu (*da*) 'occupy oneself, busy oneself'
f. dedicar/DEDICARSE → dedikatu (*da*) 'devote oneself (to), be engaged (in)'

Sometimes the source verb is originally transitive, but Basque has borrowed it in a reflexive use (or this is the only use recognised by *HB):*

(20) REFLEXIVE USE:

a. duchar/DUCHARSE → dutxatu (*da*) '(have a) shower'
b. federar/FEDERARSE → federatu (*da*) 'federate'
c. confederar/CONFEDERARSE → konfederatu (*da*) 'enter into a confederation'

Verbs of existence and appearance, which are considered unaccusative in the source languages, are also treated as *da*-verbs:

(21) VERBS OF EXISTENCE AND APPEARANCE:

a. EXISTITU 'exist'
b. FALTATU 2 'be lacking'
c. KABITU 'fit, have room'
d. PASATU 2 'happen'
e. SOBERATU 'be left over, be too much'
f. SOBRATU 'ditto'
g. SUERTATU 'happen, come about, turn out'

Some verbs of movement expressing point of departure or goal are also unaccusative in Basque:

(22) SOME VERBS OF MOVEMENT:

a. AILEGATU 'arrive'
b. PARTITU 'set out, depart'
c. HERRATU 'err'
d. ESKAPATU[12] 'escape, get away'

[12] Also used today as a *du* verb in western Basque.

However, as we mentioned above, doubts arise regarding the conjugation of borrowed verbs denoting manner of movement, e.g.

(23) VERBS OF MANNER OF MOVEMENT: VACILLATIONS

a. KORRITU 'run': unergative (*du*)
b. SALTATU 'jump': *da*, but vacillation with ?*du* in present-day Basque
c. PASEATU '(go for a) walk or ride': *da*, but vacillation with ?*du* in present-day Basque

Some verbs labeled only as *da* verbs in *HB* are really *da/du* verbs:

(24) MISLABELED IN *HB*?: *DA* > *DA/DU*

a. ATROFIATU 'atrophy' (*EH: da/du*).
b. BIZIATU 'become addicted, degenerate' (*EH: da/du*).
c. BURLATU 'make fun (of)'
d. ERRABIATU 'become furious, get mad' (*EH: da/du*).
e. GOBERNATU (*EH: du,* "zuzendu..." 'govern'; *da,* "jokatu, portatu" 'behave, act').
f. ?LIBRATU (1. *du* "atera, askatu" 'liberate, free'; 2 *da coll.* "move one's bowels").
g. PARATU (1 *du* "jarri, prestatu" 'put, prepare'; 2 *da* "jarri" 'place oneself').
h. PAUSATU (1 *da* "atseden hartu, hil; jarri" 'rest, pass away; place oneself'; 2 *du* "jarri, ipini" 'put, place').
i. PREZATU (1 *du* "balioetsi" 'appreciate, value'; 2 *da* 'harrotu" 'be proud of oneself').

Finally, there are a few cases that defy a simple explanation (cf. 3.1.3):

(25) a. DESKANTSATU (*da*) 'rest'
b. KOMEKATU/KOMULGATU (*da*) 'take communion'
c. TRABAILATU (*da: Nor. arch.*) 'work'

4. Conclusions

To sum up, the following principal patterns or strategies are observed in present-day Basque regarding the transitivity of verbs borrowed form neighbouring languages:

1. Verbs that are transitive in the source language are generally conjugated as transitives (i.e. with *du*) in Basque (e.g. *adoratu* 'adore', *bisitatu* 'visit', *fabrikatu* 'manufacture'...; *estudiatu* 'study', *eskribitu* 'write'; *alfabetatu* 'teach literacy (to)', *animatu* 'cheer up, encourage'...; *defendatu* 'defend'...). Verbs borrowed into Basque include many transitives; indeed this is the largest class. Present-day speakers do not appear to have any doubts about the the transitivity of such verbs.

As in Spanish and French, some transitive verbs which place tight selectional restrictions on their object permit an absolute or unergative use, e.g. *gauez ikasten du*

'he studies at night'. As in neighbouring languages, many transitives allow causative alternation, realized in Basque through alternation between the *du* and *da* auxiliaries. Others express reflexive meanings through the same alternation *du* > *da*.

2. When intransitive verbs are borrowed into Basque, a choice regarding transitivity is in order. Some intransitives, such as *bogatu* 'row', *desertatu* 'desert', *dudatu* 'doubt', *emigratu* 'emigrate', *erreinatu* 'reign', *eskiatu* 'ski', become unergative in Basque, conjugating with the *du* auxiliary. Others become unaccusatives, e.g. *atrebitu* 'dare', *kexatu* 'complain', *konformatu* 'be satisfied (with)', *fidatu* 'trust', *portatu* 'act, behave', *dibertitu* 'have fun'; these are conjugated with the *da* auxiliary. Thus, unlike transitive verbs, uncertainties concerning transitivity often occur with intransitives in both traditional and present-day usage.

3. However, the hypothesis (Sarasola 1977: 79) that the overwhelming present-day trend in these cases is to make intransitive verbs unergative is only partially borne out by my study. It is true that many foreign intransitives, such as *eskiatu* 'ski', *korritu* 'run', are borrowed into Basque as unergatives. But this formula is restricted by two conditioning variables: a) on the semantic level, unergativity is mostly associated with verbs which either display agentivity (e.g. *abdikatu* 'abdicate', *abortatu* 'have a miscarriage/abortion', *abusatu* 'go too far, abuse', *arnegatu* 'abnegate', *blasfematu* 'blaspheme', *bogatu* 'row', *deklaratu* 'declare, testify', *desertatu* 'desert', *dudatu* 'doubt', *emigratu* 'emigrate') or else denote an internally caused eventuality (such as *funtzionatu* 'work (as of a machine)', *kotizatu* 'have a quoted price, sell (at)', *zirkulatu* 'circulate'); b) syntactically, the unergative is restricted to intransitive verbs which are not formally reflexive in the source language. Thus, the tendency to make verbs unergative is mostly limited to borrowed intransitive agentive verbs that are not reflexive in the source language, and it is here that the recent trend towards unergativity is most striking, in verbs such as *saltatu du* 'jumped', *deskantsatu du* 'rested', *dantzatu du* 'danced', *olgatu du* 'had fun', displacing the older unaccusative usage (*saltatu da, deskantsatu da, dantzatu da, olgatu da...*).

4. The borrowed unergative verbs are not numerous: they are fewer in number than the other types of verbs, such as unaccusative borrowed verbs (taking *da*), and certainly rarer than borrowed verbs allowing causative alternation (*du/da*). In any case, in my opinion the tendency to borrow verbs as unergatives is not so much a modern development as a trend rooted in the past (cf. *EH, da > du: apelatu* 'appeal', *emigratu* 'emigrate', *komuniatu* 'take communion', *eskatu* 'request, demand', *itsastatu* 'sail') that has recently intensified. As a matter of fact, many borrowed unergatives are loans of long standing, e.g. *abusatu* 'go too far, abuse', *arnegatu* 'abnegate', *blasfematu* 'blaspheme', *bogatu* 'row', *desertatu* 'desert', *dudatu* 'doubt', *emigratu* 'emigrate', *erregutu* 'pray, beg, appeal', *erreinatu* 'reign', *erreparatu* 'notice', *faltatu 1* 'fail, miss', *komuniatu* 'take communion', *korritu* 'run', *protestatu* 'protest', *segitu* 'follow, continue', *toreatu* 'bullfight', *trafikatu* 'traffic (in)'.

5. Loans, then, that are candidates for treatment as unergative verbs in Basque are generally Spanish or French non-reflexive, agentive intransitives. While we have observed a modern development *da* > *du* in some verbs of this class (*saltatu* 'jump', *deskantsatu* 'rest', *dantzatu* 'dance', *olgatu* 'have fun', *komulgatu* 'take communion'...) there is no vacillation concerning the transitivity of verbs of other types. It is out of the question to treat as unergatives verbs that are reflexive in the source language (e.g. Fr. *se*

mutiner 'mutiny', Sp. *atreverse* 'dare', *quejarse* 'complain'...) or borrowed in a reflexive use (e.g. *conformarse* 'be happy (with)', *fiarse* 'trust', *portarse* 'behave', *signarse* 'cross oneself', *descuidarse* 'neglect, not bother', *divertirse* 'have fun', *interesarse* 'be interested', *ocuparse* 'occupy oneself, busy oneself', *dedicarse* 'devote oneself, be engaged in', *ducharse* 'have a shower'...). Verbs of existence or appearance (such as *existir* 'exist', *faltar* 'be lacking', *caber* 'fit', *pasar* 'happen', *sobrar* 'be left over, be too much'...) are hardly ever borrowed into Basque as unergatives. In my opinion Mendikoetxea's (1999) analysis of Spanish unaccusative verbs is supported by Basque, given that its treatment of borrowed verbs generally patterns with the classificiation proposed for Spanish. The Basque data, by showing that the bilingual speaker recognises implicitly the unaccusative syntactic nature of reflexives, support the analysis of reflexives in Spanish and French.

Bibliography

Arrieta, A., 2001, "Lagundu laguntzaileei", *Administrazioa Euskaraz* 32, p. 19.

Artiagoitia, X., 2000, *Hatsarreak eta Parametroak lantzen.* Gasteiz: Euskal Herriko Unibertsitatea, Arabako Foru Aldundia.

Bosque, I. & V. Demonte, (eds.), 1999, *Gramática descriptiva de la lengua española.* Madrid: Espasa Calpe.

Burzio, L., 1981, *Intransitive Verbs and Italian Auxiliaries.* Doctoral Dissertation, MIT Cambridge, Massachusetts.

—, 1986, *Italian Syntax. A Government-Binding Approach.* Dordrecht, Kluwer.

Campos, H., 1999, «Transitividad e intransitividad». In Bosque & Demonte (eds.), II: 1519-74.

Etxepare, R., 2003, «Valency and Argument Structure in the Basque Verb». In Hualde & Ortiz de Urbina (eds.).

Euskaltzaindia, 1993, *Euskal Gramatika Laburra: Perpaus Bakuna.* Bilbo: Euskaltzaindia.

Gràcia, Ll., Cabré, M.ª T., Varela, S., Azkarate, M., et al., 2000, *Configuración morfológica y estructura argumental: léxico y diccionario.* Bilbo: Euskal Herriko Unibertsitateko Argitalpen Zerbitzua.

Grimshaw, J.B., 1990, *Argument Structure.* Cambridge, Massachusetts: MIT Press.

Hualde, I. and J. Ortiz de Urbina (eds.), in press, *A Grammar of Basque.* Mouton de Gruyter.

Lafitte, P., 1944, *Grammaire Basque (Navarro-Labourdin Littéraire),* (references are to the 3rd edition: Donostia, Elkar, 1979).

Levin, B., 1983, *On the Nature of Ergativity.* Doctoral dissertation, MIT Cambridge, Massachusetts.

— & M. Rappaport, 1995, *Unaccusativity at the Syntax-Lexical Semantics Interface.* Cambridge, Massachusetts: MIT Press.

Mendikoetxea, A., 1999, «Construcciones inacusativas y pasivas». In Bosque & Dèmonte (eds.), II, 1575-1629.

Miguel, E. de, 1992, *El aspecto en la sintaxis del español: perfectividad e impersonalidad.* Madrid: Ediciones de la Universidad Autónoma de Madrid.

Perlmutter, D. M., 1978, "Impersonal Passives and the Unaccusative Hypothesis". In *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society,* 157-89. Berkeley: Berkeley Linguistics Society, University of California.

Sarasola, I., 1977, "Sobre la bipartición inicial en el análisis en constituyentes". *ASJU-International Journal of Basque Linguistics and Philology* 11, 51-90.

van Valin, R. D., 1988, "Semantic parameters of split intransitivity", *Language* 66, 221-60.

Torrego, E., 1989, "Unergative-unaccusative alternations in Spanish". *MIT Working Papers in Linguistics,* 10, MIT, Cambridge, Mass., 253-272.
—, 1998, *Valores gramaticales del 'se'.* Madrid: Arco Libros.
Vendler, Z., 1967, *Linguistics in Philosophy.* Ithaca (Nueva York): Cornell University Press.
van Voorst, J., 1988, *Event Structure.* Amsterdam: John Benjamins.

Corpora and dictionaries (with their abbreviations):

EEBS: Egungo *Euskararen Bilketa-lan Sistematikoa* (nowadays *XX. mendeko Euskararen Corpus Estatistikoa:* http://www.euskaracorpusa.com).
Elh.: Elhuyar 2000, *Elhuyar Hiztegia. Euskara-gaztelania. Castellano-vasco,* Elhuyar, Usurbil.
OEHtc: Euskaltzaindia 1984. *OEH-ko testu corpusa.*
DRAE: Real Academia Española 1992. *Diccionario de la lengua española* (21. edition), Madrid, Real Academia Española.
EH: Sarasola, I. 1996. *Euskal Hiztegia,* Kutxa.
HB: Euskaltzaindia 2000. *Hiztegi Batua.*
OEH: Michelena, L. 1987-1999. *Orotariko Euskal Hiztegia,* Bilbo: Euskaltzaindia.

PATRIXA: A UNIFICATION-BASED PARSER FOR BASQUE AND ITS APPLICATION TO THE AUTOMATIC ANALYSIS OF VERBS

Izaskun Aldezabal, M.ª Jesús Aranzabe, Aitziber Atutxa,
Koldo Gojenola, Kepa Sarasola

Abstract*

In this paper we describe a computational grammar for Basque, and the first results obtained using it in the process of automatically acquiring subcategorization information about verbs and their associated sentence elements (arguments and adjuncts). The first part of this article (section 1) will be devoted to the description of Basque syntax, and to present the grammar we have developed. The grammar is partial in the sense that it cannot recognize every sentence in real texts, but it is capable of describing the main syntactic elements, such as noun-phrases (NPs), prepositional phrases (PPs), and subordinate and simple sentences. This can be useful for several applications. Next, the syntactic grammar will be used by a syntactic analyzer (or parser) to automatically acquire information on verbal subcategorization from texts (section 2). The results will later be used by a linguist or processed by statistical filters.

1. The syntactic analyzer

1.1. A brief introduction to computational syntax

The computational treatment of syntax has long been an area of research. From 1950, when the first automatic translation systems were created, many researchers have studied the syntactic relationships among words and the way they are combined to

---

* This work has been done by the IXA Natural Language Processing research group, centered on the application of automatic methods to the analysis of Basque. Comparing to other languages (English, German, French...) Basque can be considered as a minority language due to the following constraints:

— Limited number of language users. This fact implies a reduced number of researchers/developers of computational linguistic tools.

— Limited number of language resources, in the form of computational lexicons, grammars, corpora, annotated treebanks or dictionaries.

These are the main reasons that have compelled the IXA group to the development of automatic methods for the analysis of linguistic data. The work described in this chapter is a part of this effort.

form sentences. However, the task was more difficult than expected. Nowadays, there is no system capable of syntactically analyzing any sentence in real texts, such as newspapers. At the moment, the best syntactic analyzers have been developed for English, but they find an unsolvable obstacle in the form of ambiguity, because many common sentences can produce tens or even hundreds of different syntactic analyses. In this context, we can distinguish two approaches to computational syntax, according to their main objective:

— Full parsing. The aim is to construct more accurate and complete grammars and parsers, with the objective of syntactically analyzing any sentence. As we have noted earlier, the state of the art is still far from this objective.
— Partial parsing. In many systems the objective is not to completely analyze a sentence, but to detect several syntactic elements, such as NPs, verb chains or simple sentences. These pieces of information, also called *chunks* (Abney 1997), are useful for several linguistic applications, as information retrieval or speech synthesis.

Regarding the main kind of knowledge employed, we can classify syntactic analyzers in four groups:

— Unification-based analyzers (Shieber 1986). These systems are based on context-free grammars (Chomsky 1957) with the addition of information to syntactic elements and rules by means of feature structures (see subsection 1.2.).
— Finite state analyzers (Karttunen et al. 1997). They are mainly dedicated to partial parsing, that is, they typically distinguish the different components of a sentence. Grammars are defined using regular expressions.
— Constraint grammar (Karlsson 1995). To analyze a sentence, this formalism begins with all the options to analyze each individual word-form, and the task of the grammar is to discard as many options as possible until each word contains a single analysis that gives information about number, case, person and syntactic category. This formalism is called reductionistic because it starts from all the possibilities and it ends only when the correct one is selected.
— Statistical methods. These systems automatically acquire syntactic information (in the form of context-free grammars or regular expressions) from big corpora. The information thus obtained is used to analyze new sentences. Usually, statistical methods are not used in isolation, but combined with other methods (Collins 1997).

The IXA natural language processing group has developed two syntactic analyzers for Basque, one using a unification-based formalism and another one based on a Constraint Grammar. Work on this second formalism is described in (Aduriz et al. 1997, Arriola 2000, Aduriz 2000, Aduriz and Arriola 2003b). In this chapter we will describe a unification grammar for Basque together with its application to the task of automatically extracting verbal information from text corpora.

Regarding computational grammars and syntactic analyzers for languages other than Basque we can cite the following:

— Natural Language Software Registry: http://registry.dfki.de
— Computational Linguistics (on-line presentations): http://www.ifi.unizh.ch/CL/InteractiveTools.html#as-h2-3296

Or else, if we want to experiment directly with a syntactic analyzer:

— Syntactic analyzer for English: http://www.conexor.fi
— Syntactic analyzer for Spanish (CliC): http://clic.fil.ub.es/equipo/index_en.shtml

1.2. Unification-based grammar formalism and PATR

Unification-based grammar formalisms are based on context-free grammars (CFG). CFGs were formalized by Chomsky (1957), and they define a grammar as shown in Table 1.

English grammar			Basque grammar		
S	→	NP VP	S	→	NP VP
VP	→	Verb NP	VP	→	NP Verb
NP	→	Noun	NP	→	Noun
NP	→	Det Noun	NP	→	Pronoun

Table 1. Two examples of context-free grammars

Context-free rules are of the form 'a → b' or 'a → b c', where *a* is a non-terminal syntactic category and *b*, *c* are terminals (lexical elements) or non-terminals. Non-terminal symbols (S, NP, PP...) are syntactic categories, while terminals are words or morphemes from a lexicon. The chains of terminal symbols that can be derived from the first symbol (or axiom) of the grammar (*S* or sentence in the example) will be the sentences of the language. A sentence belonging to the grammar will be typically described by a tree. For example, Figure 1 shows an analysis tree of a sentence derived using the rules for the Basque grammar in Table 1.

The formalism of context-free grammars is simple, but there are problems to describe many linguistic phenomena. For example, if we want to specify the agreement between subject and verb in number and person, then the 'S → NP VP' rule would have to be replaced by a number of similar rules, such as 'S → NP_subj_sing_3 VP_subj_sing_3_abs' or S → NP_subj_pl_3 VP_subj_pl_3_abs', and many others.

Unification-based formalisms (Shieber 1986) were defined to overcome this problem. The main idea is to add information to each syntactic element of context-free grammars by means of feature-structures, and to express the syntactic relationships and constraints using equations on that information. Unification is a useful mechanism for the treatment of Basque syntax, due to its rich word-level information and also to the complexity of the syntactic structures that must be dealt with.

This is an example of a rule, given by Shieber (1986):

S → NP VP
S head = VP head
S head subject = NP

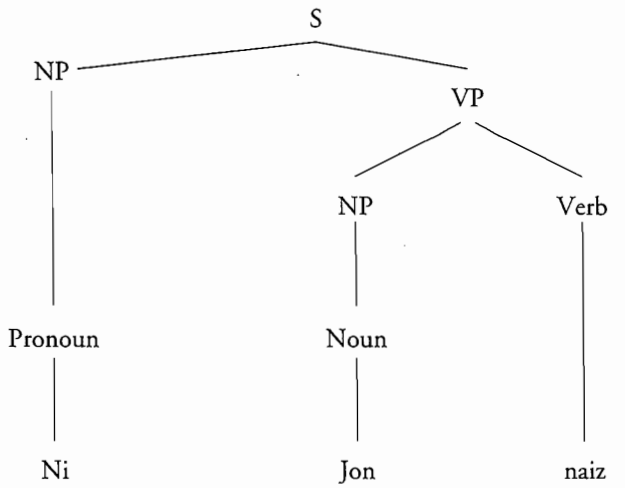


Figure 1. Analysis tree for the sentence *ni Jon naiz* (My name is Jon)

The base is a context-free rule that expresses one way of forming a sentence. Two unification equations are used to specify constraints among the sentence components. The first equation states that the head of the sentence is that of the VP, while the second one says that the subject of the sentence corresponds to the NP appearing before the VP. The application of these equations will create a feature structure describing the information in the sentence, as in Figure 2, which corresponds to the sentence "The man runs".

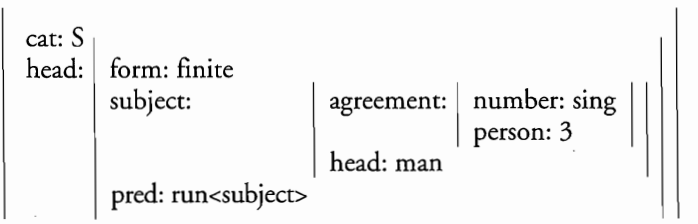


Figure 2. Example of a feature structure

Several kinds of unification-based formalisms have been defined, such as PATR (Shieber 1986), Generalized Phrase Structure Grammar (GPSG, Gazdar et al. 1985), Lexical Functional Grammar (LFG, Bresnan 1982) and Head-Driven Phrase Structure Grammar (HPSG, Pollard and Sag 1994). When developing a computational grammar, there is always a compromise between depth and breadth of analysis. Sometimes the objective is to develop a formal theory of complex linguistic phenomena. The resulting grammar can serve as a tool for the investigation of linguistic phenomena, but will not be very helpful to analyze real texts, because many linguistically inter-

esting sentences do not appear often in common texts. For example, Abaitua (1988) and Zubizarreta (1992) described several kinds of linguistic phenomena of Basque using the LFG formalism. On the other hand, there is another approach, named shallow parsing (Abney 1997), that is based on the analysis of the most frequently occurring phenomena. This allows, using limited resources, to obtain automatic tools capable of doing several tasks, such as information extraction or machine translation.

We opted for this second option, choosing PATR for the description of Basque syntax, mainly for two reasons:

— To build a computational grammar, we must use the lexical database of Basque (EDBL, Agirre et al. 1995; Aduriz et al. 1998), and this database does not contain all the information required by rich formalisms such as GPSG, LFG or HPSG.

— PATR is a flexible and simple formalism, which can serve in the first attempt to develop a computational syntactic analyzer for Basque. More complex formalisms as LFG and HPSG will be left for future developments.

We will illustrate the main characteristics of the PATR formalism with the grammar in Table 2.

R1. X0 → X1 X2 X0 cat = S X1 cat = NP X2 cat = S X1 case = erg X2 subcat erg agr = X1 agr X0 = X2 X0 subcat erg head = X1	R2. X0 → X1 X2 X0 cat = S X1 cat = S X2 cat = NP X2 case = erg X1 subcat erg agr = X2 agr X0 = X1 X0 subcat erg head = X2
R3. X0 → X1 X2 X0 cat = NP X1 cat = noun X1 type = common X2 cat = case-morpheme X0 head = X1 X0 case = X2 case X0 agr = X2 agr	R4. X0 → X1 X0 cat = S X1 cat = sv X0 subcat = X1 subcat X0 root = X1 root

Table 2. Example PATR grammar of Basque

The first rule (R1) combines a sentence (S) with an NP, giving an S (in a context-free grammar it would correspond to the rule 'S → NP S'). The X0 component (parent) is formed combining X1 and X2. The unification equations serve two purposes:

— They express syntactic constraints among the sentence elements.
— They also tell how to combine the information from the sentence components (NP and S in the right part of the rule) to form a new element (S at the left of the rule).

The first three equations of rule R1 define the categories of the syntactic elements participating in the rule. The fourth equation ('X1 case = erg') is a constraint imposing that the subject NP must be in the ergative case. The fifth equation ('X2 subcat erg agr = X1 agr') determines whether the NP and the S agree in number, definiteness and person. The sixth equation ('X0 = X2') asserts that the sentence (X0) is a projection of the simpler S appearing in the right hand of the rule, that is, they share the same information. Finally, the last equation ('X0 subcat erg head = X1') of rule R1 states that the NP corresponds to the subcategorized ergative argument.

Rule R2 expresses the same phenomenon as in R1, but changing the order of the sentence components ('S → S NP'). This is how the grammar reflects the free order of Basque. Similar rules must be defined for NPs in absolutive and dative cases, and for subordinate sentences and PPs as well (in our grammar PPs have the same syntactic structure as NPs, differing only in the grammatical case: absolutive, dative and ergative in NPs, and the remaining ones for PPs).

The second line of the table shows rule R3, which defines that an NP can be composed by a noun followed by a case-morpheme ('NP → noun case-morpheme'). This rule links a noun with a morpheme containing information about number, definiteness and case. For example, "*etxe* (house) + *-ari* (to)" (to the house).

Rule R4 defines that, in its simplest form, an S is formed by a synthetic verb (sv). Beginning from this basic S, a sentence is formed linking NPs and PPs to it (either to the right or to the left of the verb).

L1. X0 entry = ***dakarte*** X0 cat = sv X0 root = ekarri X0 subcat erg agr num = 3p X0 subcat abs agr num = 3s	L2. X0 entry = ***dakartza*** X0 cat = sv X0 root = ekarri X0 subcat erg agr num = 3s X0 subcat abs agr num = 3p
L3. X0 entry = ***-ak*** X0 cat = case-morpheme X0 case = abs X0 agr num = 3p X0 agr def = d	L4. X0 entry = ***-ek*** X0 cat = case-morpheme X0 case = erg X0 agr num = 3p X0 agr def = d
L5. X0 entry = ***gizon*** X0 cat = noun X0 type = common	L6. X0 entry = ***txakur*** X0 cat = noun X0 type = common

Table 3. Example of a lexicon in the PATR formalism

Table 2 shows an example lexicon. The L1 and L2 entries define verbal forms: *dakarte* ((they) bring (it)) and *dakartza* ((he) brings (them)). For each verb the lexicon defines its category (synthetic verb, abbreviated to *sv*) and information about subcategorization. L1 is defined as a subject-object verb (ergative + absolutive) where the NP in ergative case must be the third person plural (3p) and the absolutive NP must be third person singular (3s). L2 defines that the ergative and absolutive NPs must be respectively third person singular and plural. L3 and L4 describe case-marking morphemes: absolutive-plural (*-ak*) and ergative-plural (*-ek*). The last line of Table 2 defines two noun entries: *gizon* (man) and *txakur* (dog).

Taking this lexicon and the grammar in Table 2, the syntactic analyzer can determine that *gizonek dakarte* (the men bring (it)) or *dakartza txakurrak* ((he) brings the dogs) are correct sentences and, conversely, that sentences such as **gizonek dakartza* (*the men brings (them)) are incorrect, because in this case it does not obey the agreement constraint in R1. Figure 3 presents the syntactic tree representing the analysis of the sentence *gizonek dakarte.*

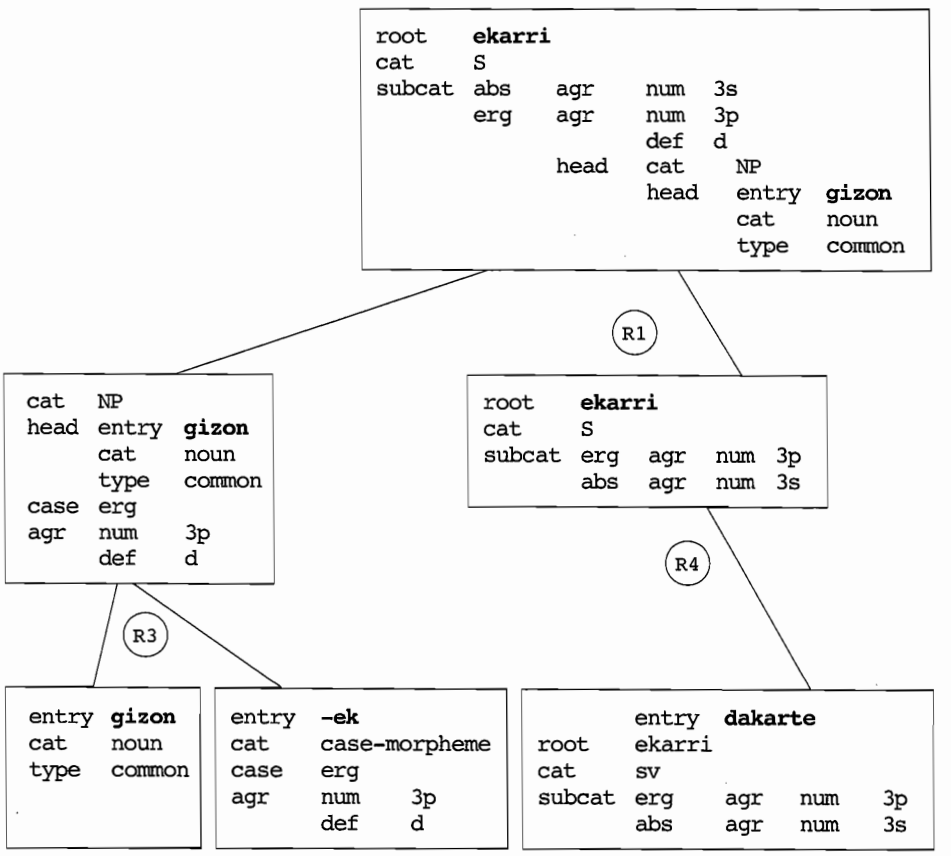


Figure 3. Analysis of *gizonek dakarte* (the men bring (it))

After explaining the basics of the PATR formalism by an example grammar we will, in the next section, describe the grammar we have developed for Basque.

1.3. A computational grammar for Basque using the PATR formalism

If we want to describe Basque syntax, we must take the following facts into account:

- The morpheme is the basic unit of analysis (Goenaga 1980, Abaitua 1988, Abaitua et al. 1992). This implies that both morphology and syntax will be integrated in the grammar, without a sharp limit between them, as it happens in agglutinative languages. This will differ from most European languages, such as English or French. For example, in the NP "*gizon* + *handi* + *-a*" (the big man), the case-morpheme "*-a*" at the end is not syntactically linked to the adjective "*handi*" (big) but to the whole noun phrase ("*gizon handi*"). This way the syntactic description is more general and simpler.
- Lexical information is rich. Every lexical entry (and the syntactic elements projected from it) contains information about number, definiteness, case or syntactic functions. The main objective of the grammar will be to adequately combine all this information.
- The lexicon does not contain full subcategorization information. Verbs are the central elements in syntax, both in syntactic theories and in applied systems. From the verbal information, subcategorization is the most complex, specifying how each verb combines with other kinds of elements. In Basque the auxiliary verb conveys information about the subject, object and indirect object (case, number and person), but the lexical database we are using lacks information about main verbs.
- There is agreement between the verb and subject, object and indirect object (corresponding roughly to the ergative, absolutive and dative cases).
- Free order of sentence components. In Basque the order of the main sentence elements (NPs and PPs) is relatively free. This means that in the following example changing the order of subject, object and indirect object gives 24 possible permutations, which are correct sentences in some context:

Txakurrak	*egunkaria*	*ahoan*	*zekarren.*
The-dog	the-newspaper	in-his-mouth	brought
ergative-3-s	absolutive-3-s	inessive-3-s	
subject	object	modifier	verb

(The dog brought the newspaper in his mouth)

We must also say that this flexibility at sentence level is much more restricted for other syntactic elements (for example, inside NPs or subordinated sentences).

Next, we will begin a description of the grammar, showing the structure of NPs and PPs, and then we will continue with the sentence structure. We have described three main types of NPs (PPs):

1. NPs and PPs with a common noun as head. NPs and PPs end with a case-morpheme (it contains information about case, number and definiteness). Before the noun there could be optional genitive NPs (similar to PP-of in English) and determiners. After the noun there could be one or more adjectives

and determiners (optional). Unification equations are in charge of checking constraints on order or number:

(NP-gen) + (det) +	**noun +**	**(adj) +**	**(det) + case-morpheme**
etxeko	*gauza*	*zahar*	*hori*_____*ekin*
of-the-house	thing	old	those_____with (3rd-pl)

(with those old things of the house)

etxeko	*lau*	*gauza*	*zahar*_____*etan*
of-the-house	four	thing	old_____in (3rd-pl)

(in four old things of the house)

etxeko	*gauza*	*zahar*_____*ari buruz*
of-the-house	thing	old_____regarding (3rd-sg)

(regarding the old thing of the house)

2. NPs (or PPs) with a proper noun as head. There are optional genitive NPs, but neither adjectives nor determiners are accepted:

(NP-gen)	**+**	**proper-noun**	**+**	**case-morpheme**
Donostiako		*Jon*_____		*ri*
of-Donostia		Jon _____		to

(to Jon of Donostia)

3. NPs with a pronoun as head. They only admit the case morpheme:

pronoun	**+**	**case-morpheme**
ni _____		*ri*
I _____		to

(to me)

These descriptions are relatively simple but not 100% complete, because there are exceptions to some of the principles stated. For example, in NPs formed by a proper noun it could be correct to use adjectives in some contexts, but the inclusion of this fact would have several disadvantages:

— The grammar would be considerably more complicated.
— The resulting ambiguity would increase. It is usual to have tens of analysis for many sentences, due to lexical ambiguity (several analysis per word-form) and syntactic ambiguity (when a part of a sentence can be analyzed by different rules). The inclusion of exceptional cases has the effect of dramatically increasing ambiguity.
— The introduction of new possibilities, although correct in some context, only would account for a very small fraction of sentences in real texts. As our objective is to use the analyzer as a tool for the analysis of written texts, we decided not to include the special rules in the grammar, as most of them would describe phenomena that do not have even a single instance in the corpora we have studied.

In order to accept the described kinds of syntactic structures, we have defined several auxiliary syntactic categories np1, np2 and np3, starting from the simplest categories to the most complex ones. Finally, adding a case-morpheme to the highest-level structure (np3) forms the category npc (NP + case), that corresponds to an English NP or PP (in fact, they are distinguished by their case: absolutive, ergative and dative for NPs, and the rest of the cases for PPs). We have taken a broad definition of a case-morpheme. It will describe a suffix containing information about number, case and definiteness. Moreover, we have defined complex suffixes (postpositions) formed by the combination of a suffix with a different word (for example, we take *-ri_buruz* as a suffix, as in *zinemari buruz* (about the cinema)).

The following rules show the structure of NPs and PPs:[1]

	Rule	Examples
1	np1 → noun adj	*etxe EDER* (NICE house)
	| noun	*etxe* (house)
2	np2 → det np1	*ZENBAIT etxe eder* (SEVERAL nice houses)
	| det np1	*HIRU etxe eder* (THREE nice houses)
	| np1 det	*etxe eder BAT* (ONE nice house)
	| proper-noun	*JOHN*
	| np1	*etxe eder* (nice house)
3	np3 → np-gen np2	*MENDI HORRETAKO zenbait etxe eder* (several nice houses OF THAT MOUNTAIN)
	| pronoun	*ZU* (you)
	| np2	*zenbait etxe eder* (several nice houses)
4	npc → np3 case-morpheme	*etxe ederrEKIN* (WITH the nice houses) *mendiko zenbait etxe ederrAK* (several nice houses of the mountain)
		mendiko zenbait etxeRI BURUZ (REGARDING several nice houses of the mountain)
5	np-gen → np3 case-morpheme(gen/gel)	*mendi horretaKO* (OF that mountain)

Table 4. Grammar rules for NPs

[1] The example rules are a simplification of the actual rules. As we have explained before, each rule will have an associated set of unification equations describing syntactic restrictions among its components.

The structure of the genitive NP (np-gen in rule 5) is the same as for a general NP, where the case must be one of the two genitives (*gen* (possessive) and *gel* (locative)). In the analysis of a sentence, we do not distinguish the subject from other NPs. A sentence will be a projection of a verb-phrase (VP). The simplest VP is formed by a verb (synthetic or formed by a main verb plus an auxiliary verb). After recognizing the verb, its dependents will be added one by one either to the left or to the right, using the rules in Table 5.

	Rule		Examples
6	vp → synthetic-verb		*dakartza* ((he) brings (them))
7	vp → main-verb	aux-verb	*ikusi dute* ((they) have seen (him))
8	vp → npc(erg)	vp	*GIZONEK ikusi dute* (THE MEN have seen (it))
	| npc(abs)	vp	*GIZONAK ikusi dituzte* ((they) have seen THE MEN)
	| npc(dat)	vp	*GIZONARI eman dio* ((he) has given (it) TO THE MAN)
	| vp	npc(erg)	*ikusi dute GIZONEK* (THE MEN have seen (it))
	| vp	npc(abs)	*ikusi dituzte GIZONAK* ((they) have seen THE MEN)
	| vp	npc(dat)	*eman dio GIZONARI* ((he has given (it) TO THE MAN)
9	vp → (npc *not abs, erg or dat*)	vp	*GIZON HORREKIN ikusi dute* ((they) have seen (him) WITH THAT MAN)
10	vp → adb	vp	*GAUR egin dut* ((I) have done (it) TODAY)
11	vp → subord-modal-temp	vp	*HONA NENTORRELA ikusi dut* ((I) saw (it) WHILE COMING HERE)
	| subord-ind-interrog.	vp	*EA JOAN DEN galdetu du* ((he) asked WHETHER HE WAS GONE)
	| subord-completive	vp	*ETORRI DIRELA jakin da* ((it) has been known THAT THEY HAVE COME)
12	subord-completive → vp	subord-suffix	*Hona nentorrELA* (THAT (I) was coming here)

Table 5. Grammar rules for sentences

1. Rules 6 and 7 express the simplest way to form an VP, that is, a sentence. It is formed either by a synthetic verb or by the combination of a main verb with an auxiliary verb.

2. Rules for analyzing the grammatical cases (rule 8 in Table 5). NPs in the ergative, absolutive and dative case must agree with the verb in number, case and person. The three rules are duplicated in order to account for free constituent order.
3. Rules for adjuncts (rule 9). These rules account for all the cases (instrumental, inessive...) apart from the grammatical ones.
   As before, there will be a corresponding rule that accepts an adjunct after the verb.
4. Rules for adverbs (see rule 10).
5. Rules for linking subordinated sentences to a verb: completive, indirect interrogative, modal and temporal (see rule 11).
6. Rules for subordinated sentences. They are formed by adding a subordination suffix to a sentence (see rule 12).

The grammar contains a total of 90 rules, each one with an average of 15 equations. As we have explained before, the rules are more complex than the ones presented. Example 1 shows a part of the rule "np3 → np-gen + np2".

X0 → X1, X2

X0 cat	= np3
X1 cat	= np-gen
X2 cat	= np2
X0 sint agr	= X2 sint agr
X0 lexhead	= X2 lexhead
X0 sint elements np-gen	= X1 sint np-gen
X0 sint elements adj	= X2 sint elements adj
X0 sint elements determiner	= X2 sint elements determiner
X0 sint head agr	= X2 sint head agr

...

Example 1. Grammar rule

As the resulting grammar uses a broad-coverage lexical database, we can say that the analyzer is capable of analyzing any NP (or PP) in real texts, also verifying agreement among the component elements, added to the proper use of determiners. This also happens with sequences of the following syntactic elements not separated by punctuation marks:

— Verbs and verb chains.
— NPs (grammatical cases: ergative, absolutive and dative).
— Adjuncts (NPs in cases other than the three grammatical ones).
— Adverbs.
— Nominalized verbs.
— Relative, completive and modal subordinate clauses.
— Temporal subordinate clauses.
— Indirect interrogatives.

— Simple sentences using all the previous elements. The rich agreement between the verb and the main sentence constituents (subject, object and second object) in case, number and person is verified. As we explained before, sentence analysis is performed up to the level of phenomena that can be described using only syntactic information now included in the lexicon.

1.3. Examples

Figure 4 shows the analysis of the NP *'gure etxe polit hark'* (that nice house of us). The union of np-gen (of us) and np2 (that nice house) gives an element of category np3, and adding the final case-morpheme *(-ak)* gives the final NP (npc).

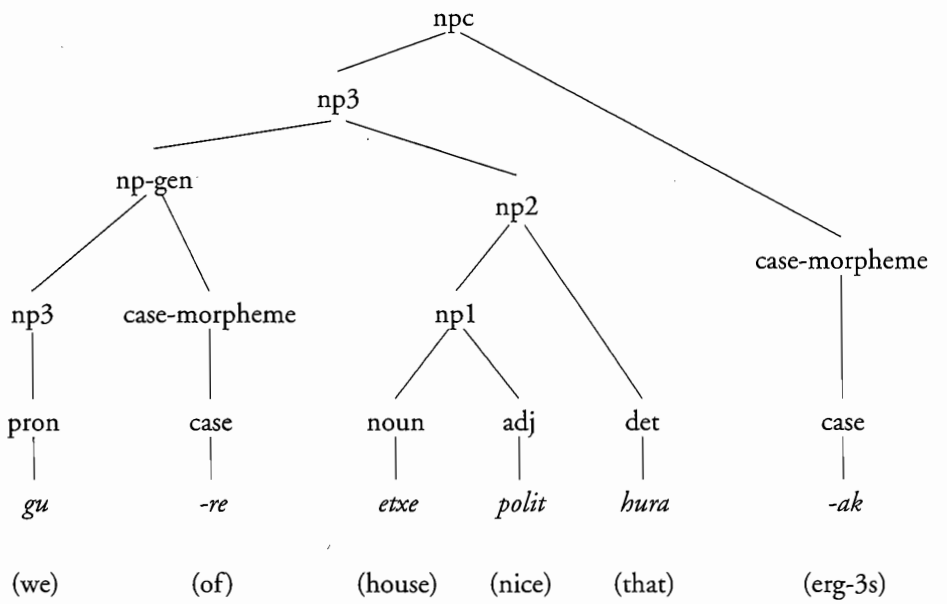


Figure 4. Analysis of *'gure etxe polit hark'* (that nice house of us)

Figure 5 shows the analysis of the sentence *'etxera zetorrela jakin du'*. In this example, a completive subordinated sentence *'etxera zetorrela'* (that he came to the house) is linked to the main sentence *('jakin du')*.

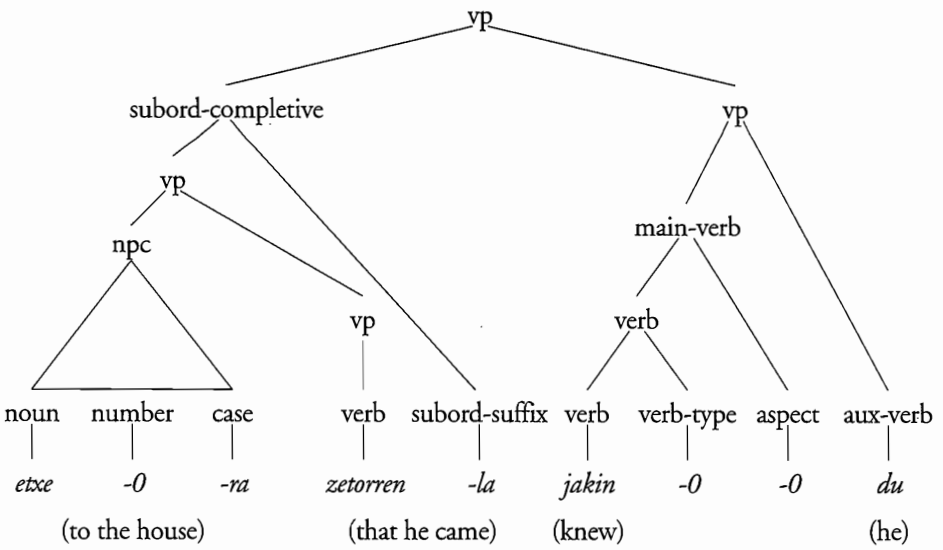


Figure 5. Analysis of *'etxera zetorrela jakin du'* ((he) knew that he came to the house)

1.4. Summary

In the first part of this chapter we have presented the core of PATRIXA, a computational syntactic grammar of Basque. As the lexical coverage is very robust (more than 70.000 lexical entries from the Lexical Database of Basque are used), we can say that the syntactic analyzer provides a good coverage of syntactic elements for the analysis of real texts (newspapers or written texts). The grammar describes extensively NPs, PPs, subordinate sentences and simple sentences.

The grammar can be useful from two perspectives. First, it can be a tool for linguists, helping them in the examination of corpora. The analyzer will give the possibility of finding the syntactic structures present in written texts. Second, it can also be useful for several applications, such as information retrieval or machine translation, where it is crucial the determination of basic syntactic units as the ones found by the analyzer.

In order to obtain deep syntactic analysis of sentences, we think that the next step should be the inclusion of verbal subcategorization information in the grammar. For that reason, we have used the syntactic analyzer to automatically acquire information about verbs and their complements from text corpora. These experiments will be described in section 2.

2. Application to the automatic analysis of text corpora.

In this section we will describe the application of the syntactic analyzer to the extraction of information about 1.400 verbs from a newspaper corpus, followed by a preliminary evaluation of the results. These results will be used for both manual and automatic examination (Aldezabal et al. 1998, 2000, 2001).

The acquisition of lexical information is an ineludible step in many applications, ranging from lexicography (construction of dictionaries) to automatic systems, such as machine translation or automatic text understanding. Most of the recent syntactic theories project syntactic structure from the lexicon, where every verbal entry will contain information about predicate subcategorization, including the number and type of arguments, semantic selectional preferences, and so on (Briscoe and Carroll 1997). Manual acquisition of lexical information is reliable and accurate in general, but it is also a costly enterprise, because of the need of highly specialized experts (linguists) in a very time-consuming process. Moreover, manual encoding also faces the problems of errors, such as omission of relevant information or, conversely, adding information based on a linguist's intuitions which do not match with real occurrences. To that we must also add that predicate subcategorization is associated with lexical senses, which vary with the corpus or domain. The huge size of the now available corpora demands successive extensions of the lexicons, to include corpus-specific information or to augment the available lexical information.

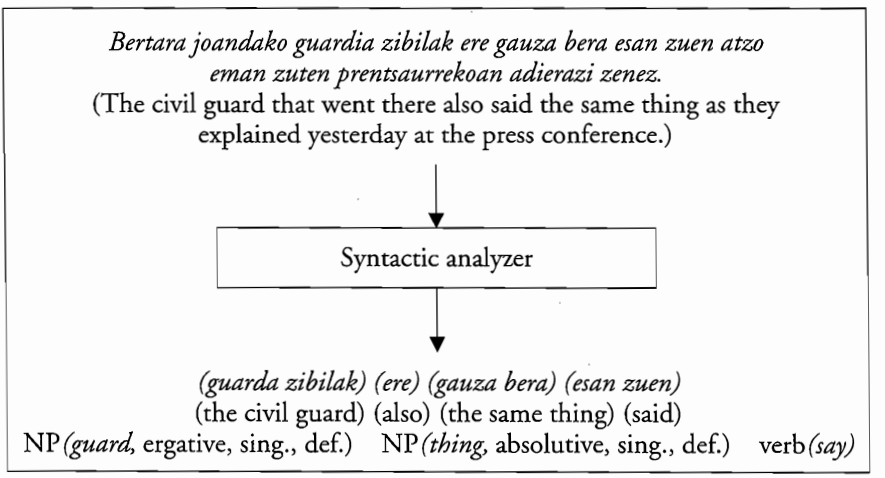


Figure 6. Input sentence and result for the verb *esan* (say)

For that reason, we have explored the possibility of using computers to help in the process of lexical acquisition. Automatic methods will never get the reliability of a linguist expert, but they can be helpful in several cases:

— The information gathered automatically can be validated by experts. This way, the linguist gets rid of the most mechanical task of examining hundreds of text sentences.

— In cases where it is not feasible to dedicate people to the task of lexical acquisition, automatically collected information could serve as an approximation useful for several applications. The reliability of the approximation can be evaluated examining a small fraction of the extracted information.

In our experiment, we have automatically examined more than 1.000.000 words of newspaper text obtaining, for each of 1.400 verbs, the set of sentences containing each verb and the elements associated with it (arguments and/or adjuncts), marking each element with information about case, number or type of subordinated sentence. Figure 6 shows the result obtained by the system when examining the verb *esan* (say). The syntactic analyzer first tries to analyze the whole sentence. As the grammar is partial and the sentences long, many times the analyzer does not find an analysis for all the sentence, but it can obtain the main syntactic components. In a second phase of the process, the analyzer looks up the syntactic elements surrounding the target verb (*esan*) and determines which of them are the most plausible arguments or adjuncts. This way, the result is the last line in Figure 6, where the verb is linked with two NPs (ergative and absolutive). This kind of information can be useful for an ulterior manual or automatic determination of subcategorization frames.

Subsection 2.1 will review previous works on the automatic acquisition of subcategorization information. Next, we will describe the architecture of the system (subsection 2.2), together with the linguistically relevant aspects of the experiment. In subsection 2.3 we will examine the results.

2.1. Previous work on the acquisition of subcategorization information

Concerning the acquisition of verb subcategorization information, there are proposals ranging from manual examination of corpora (Grishman et al. 1994) to fully automatic approaches. (Briscoe and Carroll 1997, Carroll et al. 1998) describe a grammar based experiment for the extraction of subcategorization frames with their associated relative frequencies, obtaining 76,6% precision and 43,4% recall.

Kuhn et al. (1998) compare two approaches for the acquisition of subcategorization information: a corpus query pattern based approach (no grammar, using regular expressions on morphologically analyzed word forms) and a grammar based approach (in a way similar to Briscoe and Carroll (1997). Both are applied to the problem of acquiring subcategorization instances of 3 subcategorization frames, showing that the grammar based approach improves results specially in recall, due mainly to the higher-level knowledge encoded in the grammar. Comparing with our work, we think that our system is situated between the two approaches, as we will use a partial parser. Our objective is more ambitious in the sense that we try to find all the subcategorization instances, rather than distinguishing among 3 previously selected frames.

On the statistical side, Carroll and Rooth (1998) present a learning technique for subcategorization frames based on a probabilistic lexicalized grammar and the Expectation Maximization algorithm using unmarked corpora. The results are promising, although the method is still computationally expensive and requires big corpora (50 million words).

2.2. Description of the process

We have developed a parsing system divided in several main modules: the unification-based parser that we have presented in section 1 is the core of the system (see Figure 7). Prior to parsing, there is another step concerned with morphological

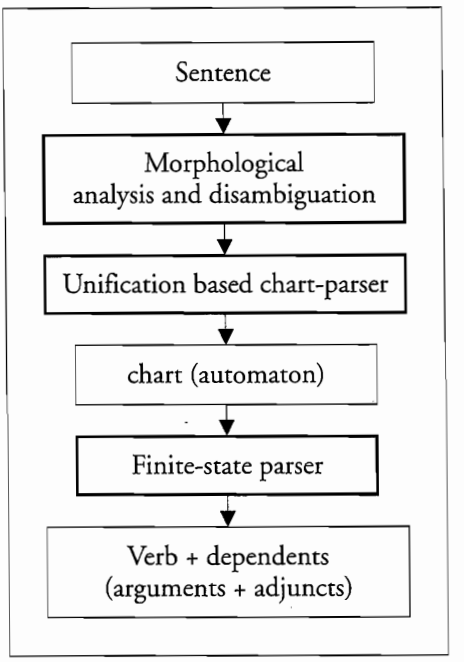


Figure 7. Description of the system

analysis and disambiguation, using the basic tools for Basque that have been developed in previous projects. These are the main modules of our system:

— The lexical database. As we have commented earlier, it is a large repository of lexical information, with about 70.000 entries (including lemmas and declension/derivational morphemes), each one with its associated linguistic features, like category, subcategory, case and number, contained in a commercial database management system.

— Morphological analysis and segmentation. Inflectional morphology of Basque was completely described in (Alegria et al. 1996). This system applies Two-Level Morphology (Koskenniemi 1983) for the morphological description and obtains for each word its segmentation(s) into component morphemes, where each morpheme is associated with its corresponding features in the lexicon. The segmentation module has full coverage of free-running texts in Basque, and it is capable of treating unknown words and non-standard forms, such as dialectal variants and typical errors (Aduriz et al. 2003a).

— Morphological disambiguation. A disambiguation system was implemented for the assignment of the correct lemma and part-of-speech to each token in a corpus (Ezeiza et al. 1998) taking the context into account, by means of statistical (Hidden Markov Models) and hand-crafted rules in the Constraint Grammar formalism (Samuelsson and Voutilainen 1997, Karlsson et al. 1995,

Aduriz et al. 1997). This tool reduces the high word-level ambiguity from 2,65 to 1,19 interpretations, still leaving a number of interpretations per word.

— Unification-based chart parsing. The syntactic analyzer presented in section 1 recognizes the main syntactic units of the sentence, described in the unification-based PATR grammar of Basque.

— After the partial parser has obtained the main syntactic components of the sentence, there are multiple readings for each sentence, as a result of both morphological ambiguity (1,19 interpretations per word-form after morphological disambiguation) and syntactic ambiguities introduced by the partial parser. For this reason, we have also developed a finite-state grammar that performs syntactic disambiguation and filtering of the results. This grammar consists of a set of regular expressions and transducers for both disambiguation and determination of clause boundaries, in order to exactly delimit the syntactic elements corresponding to each verb (Aldezabal et al. 2003a). The finite state filter has been implemented using the *Xerox Finite State Tool* (XFST, Karttunen et al. 1997).

2.2.1. *Size and type of the corpus*

In the present work we have used newspaper texts from "Euskaldunon Egunkaria", ranging from January 1999 to May 2000. This corpus offers a rich variety of text types, using standard Basque. It contains 111.000 sentences (more than one million words). In a preliminary stage of this work we also used the *Statistical corpus of 20th Century Basque* (UZEI 2003).

2.2.2. *Number of verbs*

We selected a preliminary set of 1.400 verbs appearing in the corpus. From them, 400 had more than 50 occurrences in the corpus, which we have taken as the minimum for the results to be representative.

2.2.3. *Data extraction method*

After doing some preliminary tests and a manual verification of the results, we defined several procedures to be applied, related with specific features of Basque, with the aim of improving the reliability of the results. The resulting procedures are the following:

1. Grouping of cases and subordination suffixes. Basque has a high number of cases and subordination suffixes. In our grammar we have described 61 different types. Concerning the verb, however, several of them perform a similar function. We will not go into details about what we have defined as a "similar function". The grouping was made based mainly on the syntactic function (subject, object...), also taking into account semantic relationships. So, for example, we have grouped subordination suffixes related to time: *-nean* (when), *-t(z)ean* (when), *-rako* (for when), *-terakoan* (while), *-takoan* (after), *-ino* (until), *-netik* (since), *-neko* (of when). We must also say that the grouping could be done in a different way depending on the definition of "similarity". After the grouping, we had 28 groups of elements.

2. Using the auxiliary verb. The auxiliary verb in Basque gives information about the "grammatical cases" (absolutive, ergative and dative). So, even when a sentence does not contain an NP corresponding to one of these cases, the auxiliary verb reflects their occurrence and, therefore, we can assume that the elements exist. This feature is characteristic of *pro-drop* languages. Nevertheless, in unergative verbs the object NP (marked with the absolutive case) does not exist, even when the auxiliary verb marks it. Taking these verbs into account, we have decided not to recover NPs in the absolutive case, because doing it the system would get incorrect information about all of the unergative verbs.
   Summarizing, the recovering of cases has been applied in the following syntactic environments:

   — If the auxiliary is of the type absolutive-ergative (this type of verb is usually represented by the form corresponding to the present indicative in third person singular: DU), the NP in the ergative case will be recovered. This assumption will be wrong for all the verbs associated to weather (to rain, to snow...), because they will never have a subject in the ergative case. However, as these verbs form a reduced set that could be treated separately, we estimated that the application of this heuristic will be useful.
   — If the auxiliary verb is of the type absolutive-ergative-dative (DIO), the ergative and the dative NPs will be recovered.
   — If the auxiliary verb is of the type absolutive-dative (ZAIO), the dative NP will be recovered.

3. Elimination of ill-formed syntactic combinations. Several combinations of cases with the auxiliary verb can never appear in a sentence and, consequently, we eliminated them, because they will always correspond to an error of the syntactic extraction system. Most of the times the errors appear because the main sentence and the subordinated ones are incorrectly delimited:

   — An ergative NP can never appear with an auxiliary verb of the absolutive (DA) or absolutive-dative type (ZAIO).
   — A verb cannot contain two ergative NPs.
   — Syntactic structures with more than five elements (arguments or adjuncts) are not common, and most of the times are a result of errors of our analyzer. For that reason, we did not take them into account.

2.3. Results

Table 6 presents an example of the results obtained by the system when applied to the verb *hurbildu* (to near). The second column contains the input sentence, where the subsentence corresponding to the target verb has been marked in bold type. The third column presents the result obtained by our system. For each instance of the target verb the system gets its auxiliary verb and, for each dependent, its case, head and number. For example, in sentence 1 the system finds NPs in the absolutive, inessive and adlative cases. The result will be the set of candidate dependents, where some of them will be arguments and the rest will correspond to adjuncts. For example, in sentence 1 the

	Input sentence	Output
1	*Bideoa bezalako euskarri berrien abantailak azpimarratu zituen Villotak, eta* ***ildo horretan dokumentala bideo-sorkuntzara hurbildu dela*** *deritzo.* Villota stressed the advantages of new media such as video, and **in a similar way** he thinks that **documental has neared towards video creation.**	**verb:** *hurbildu* **auxiliary:** *dela* (DA) **absolutive:** *dokumental* **head:** documental (sing.) **inessive(in):** *ildo horretan* **head:** way (sing.) **adlative(to):** *sorkuntzara* **head:** creation (sing.)
2	*Unionista amorratuenek eta, gezurtia deitu zioten Trimbleri,* ***UUPko burua sarrerara hurbildu zenean.*** And the most stubborn unionists called Trimble liar, when **the head of UUP neared the entry.**	**verb:** *hurbildu* **auxiliary:** zenean (DA) **absolutive:** *UUPko burua* **head:** head (sing.) **adlative(to):** *sarrerara* **head:** entry (sing.)
3	*Baina* ***jendea frontoira hurbiltzen ari da,*** *erantzuten ari da.* But **people is nearing the fronton**, they are responding.	**verb:** *hurbildu* **auxiliary:** *da* (DA) **absolutive:** *jendea* **head:** people **adlative(to):** *frontoira* **head:** fronton (sing.)
4	*Garaipena eskuan,* ***Pascual Jover minutu batzuk beranduago hurbildu zen Vital kutxaren aretora.*** With the victory in his hands, **Pascual Jover neared the Vital hall several minutes later.**	**verb:** *hurbildu* **auxiliary:** zen (DA) **absolutive:** *minutu batzuk* **head:** minute (pl.)
5	*Manifestazioa Hernani kaletik zihoala,* ***pertsona bat ondoko kale batetik hurbildu zen presoen aldeko oihalarekin*** *eta eskuak goraturik, bake seinalean.* When the demonstration crossed Hernani street, **one person neared with a sheet in favour of prisoners from a street nearby** and his hands up, in sign of peace.	**verb:** *hurbildu* **auxiliary:** *zen* (DA) **absolutive:** *pertsona bat* **head:** person (sing.) **ablative(from):** *ondoko kale batetik* **head:** street (sing.)

Table 6. Examples of input sentences and their corresponding output

inessive NP *ildo horretan* (in the same way) is an adjunct, while the other NPs correspond to arguments.

Sentence 4 is an example where the system gets an incorrect result, because the syntactic analyzer does not recognize the temporal modifier *minutu batzuk beranduago* (several minutes later) as a single unit, due to a gap in the partial grammar. As a result, it incorrectly proposes *minutu batzuk* (several minutes, absolutive) as the subject of the target verb.

Finally, sentence 5 shows how sometimes the system does not obtain the complete list of dependents of a verb. In this example, the analyzer correctly identifies two dependents, but misses a third one: *presoen aldeko ohialarekin* (with a sheet in support of prisoners). This is due to unresolved ambiguity of the auxiliary verb *zen*, which can be both sentence final and a verb in the past tense. In this example, the correct reading corresponds to the past tense, which would imply that this element is a dependent. However, as the morphosyntactic disambiguation process is not able to decide about which one is the correct interpretation, the system, in case of doubt, does not take any risk, and discards the element, taking into account the sentence final interpretation. This strategy tries to maximize precision (that is, to minimize the number of incorrect dependents) at the cost of lowering recall (some correct elements will also be discarded).

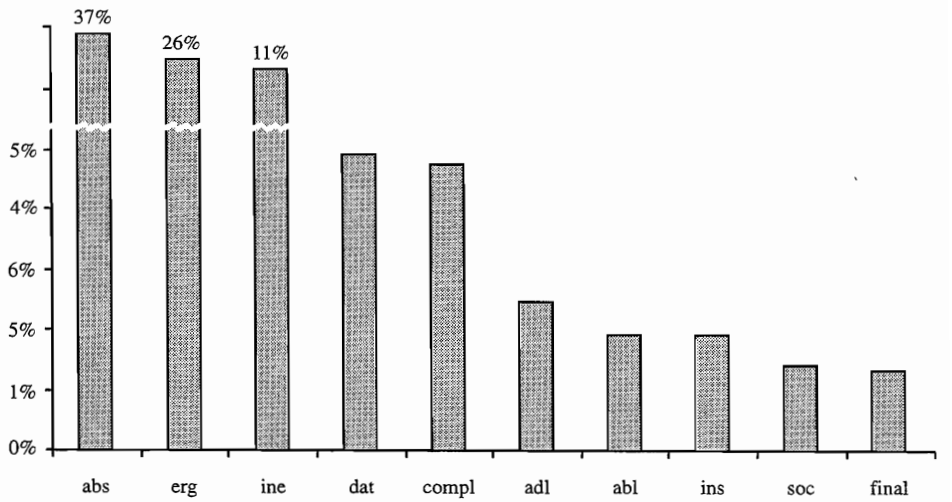


Figure 8. Most frequent cases and subordination suffixes appearing in the corpus

In order to estimate the results obtained by our system, we tested three different approximations:

1. General frequency of dependents. With the aim of obtaining a general view of the corpus, we measured the relative frequency of each type of dependent, including all the cases for NPs (PPs) and each type of subordinated sentence. Figure 8 shows the ten kinds of dependents appearing most in the corpus (those that appear in more than 1%

of the sentences). Table 7 shows the correspondence of the abbreviations in the table with their associated syntactic element.

Case	Abbreviation	Example
absolutive	abs	THE HOUSE (object)
ergative	erg	THE MAN (subject)
inessive	ine	IN THE HOUSE
dative	dat	TO THE MAN
completive subordinated sentence	compl	(I know) THAT SHE WOULD COME
adlative	adl	TO THE HOUSE
ablative	abl	FROM THE HOUSE
instrumental	ins	WITH THE HAMMER
sociative	soc	WITH THE MAN
final subordinated sentence	final	(I did it) FOR YOU TO COME

Table 7. Different types of dependents

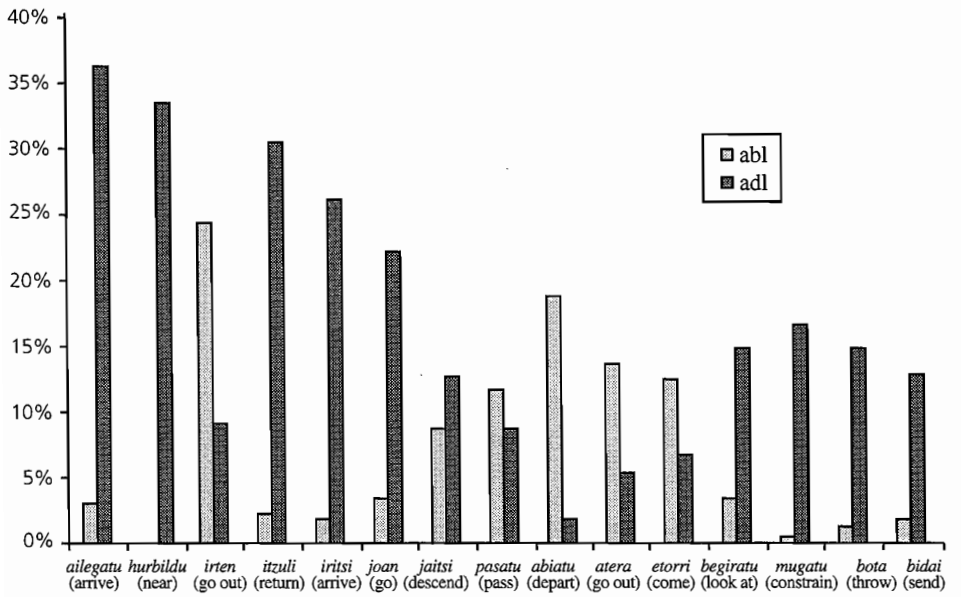


Figure 9. Verbs with high frequency of ablative and adlative cases

Figure 8 shows that three types of dependents appear most frequently: NPs in the absolutive, ergative and inessive case. The high frequency of the absolutive case can be considered normal, as this is the case used to represent the subject of intransitive verbs

as well as the object of transitive ones, that is, this case will appear with most of the verbs. Similarly, the ergative case is used as the subject of transitive and unergative verbs. The high frequency of the inessive case can be explained if we take into account that the corpus is formed by newspaper texts, which must be situated both in time and location.

If we look at the next most frequent types of dependents, we find the dative case, typically representative of goal, and completive sentences. These can also be derived from the type of corpus, because many communicative verbs are used, containing a message that has to be transmitted (and sometimes has an associated goal). This is the case with verbs expressing volition, desire or preference.

Next to these elements we find the locative cases: ablative, adlative and instrumental (by, by means of), followed by the sociative and the subordination suffix *-t(z)eko*, which can be both final and completive.

2. In a second approximation we wanted to investigate the validity of the results regarding the ability of the system to detect certain types of verbs from their associated dependents. In our experiment we tried to select verbs corresponding to motion taking those verbs with the highest frequencies of the ablative (from) and adlative (to) cases. Figure 9 shows the 15 verbs with a highest frequency of these two cases in the corpus.

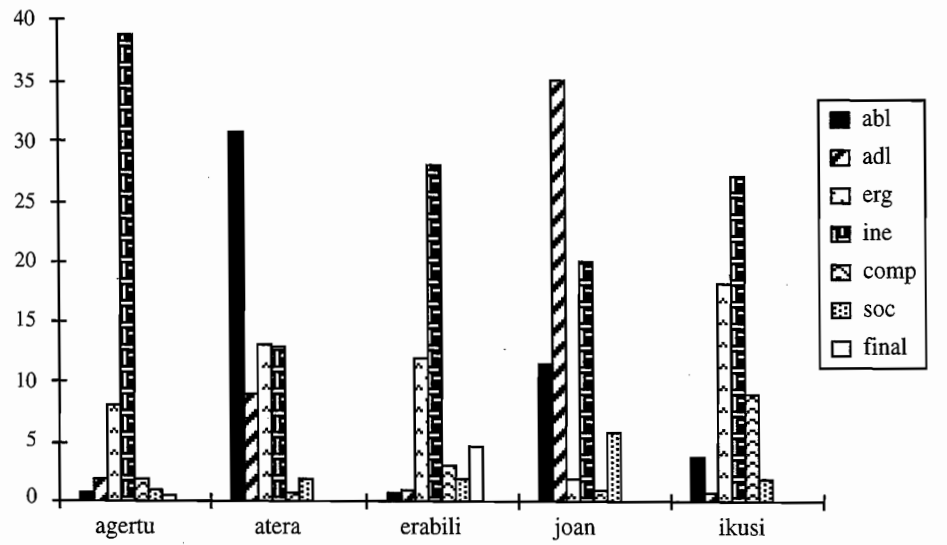


Figure 10. Frequency of elements appearing with each of five verbs

The results show the usefulness of the system to find verbs with similar characteristics. From the 15 verbs with highest proportion of the cases ablative and adlative, 13 correspond to typical motion verbs. The two exceptions are *mugatu* (to constrain) and *begiratu* (to look at).

Even when all the verbs admit both cases, many times the verb shows preference for one of them. For example, the verb *hurbildu* (to near), in one extreme, is rarely ac-

companied by the ablative case. This asymmetry could be explained defining two subclasses of motion verbs:

— Verbs expressing source or beginning by means of the ablative case (from). This set would contain the following verbs, sorted by descending order of frequency: *irten* (to go out), *abiatu* (to depart), *atera* (to go out), *etorri* (to come) and *pasatu* (to pass).
— Verbs expressing destiny, which express a goal or arrival by means of the adlative case: *ailegatu* (to arrive), *hurbildu* (to near), *itzuli* (to return), *iritsi* (to arrive) and *joan* (to go).

3. Finally, we studied the frequencies of dependents for five common verbs: *agertu* (to appear), *atera* (to go out), *erabili* (to use), *joan* (to go) and *ikusi* (to see).

Figure 10 shows the frequencies of elements appearing with each verb. The absolutive case has been omitted, because it is the most frequent one in all the verbs, due to the reasons explained before. The inessive is predominant, as it situates the sentences in temporal and spatial coordinates. The ergative case gives the subject of actions. After these elements, we can see how each verb shows preference for different kinds of subcategorized elements. For example, the verb *erabili* (to use) contains a high proportion of subordinated sentences with the *-t(z)eko* suffix, expressing finality.

These results show that the tool is useful for the automatic selection of possible subcategorized elements. The information obtained can then be used by a linguist or processed by statistical methods to select subcategorization frames for verbs.

3. Conclusion

In this work we have presented PATRIXA, a syntactic analyzer for Basque based on a unification-based formalism (PATR), and its application to the automatic analysis of texts, in order to extract information on verbal subcategorization.

These are the main features of the syntactic analyzer:

— Lexical coverage. As the system is based on a wide-coverage lexical database of Basque (EDBL) with more than 70.000 entries, the system is very robust, capable of analyzing almost any word occurring in texts.
— Grammatical coverage. The system correctly analyzes NPs, PPs, simple sentences and subordinated sentences. However, the grammar does not address several linguistic phenomena such as coordination or complex sentences.
— Ambiguity. Many times the syntactic analyzer obtains more than one analysis for a piece of text. For example, *gizonak* can be both "the man"(subject) and "the men"(object). This has been dealt with by means of special disambiguation rules (Aldezabal et al. 2003a).

In the second part of the work (section 2), we have presented the application of the grammar to the automatic analysis of texts, with the objective of obtaining information on verbal subcategorization. These are the main characteristics of the experiment:

— The corpus contains more than a million words of newspaper texts, with the objective of obtaining information about 1.400 verbs.

— The system obtained, for each verb and sentence, a list of its corresponding dependents (arguments and adjuncts). For evaluation we measured precision (the number of correctly selected elements/all the elements returned by the parser) and recall (the number of correctly selected elements/all the elements present in the sentence). The results are reliable, with 87% precision (this corresponds to the proportion of correctly selected dependents) and 66% recall (that is, the system obtained an analysis for 66% of the sentences). Although there is always a balance between recall and precision, we tried to maximize the latter, sometimes at the cost of lowering recall.

The following are the lines of work to continue in the future:

— Extension of the grammar. We plan to extend the grammar in two ways. First, including syntactic constructions not treated at the moment, such as coordination or complex sentences. Second, including subcategorization information, not present at the moment in the lexical database.
— Regarding the results of the analyzer, the information gathered will be used to manually and automatically extract subcategorization information about verbs.
— We also plan to compare the results with other works on extraction of subcategorization information. For example (Arriola 2000) has studied the extraction of this kind of information from a dictionary (Sarasola 1997).

Acknowledgements

This research has been supported by the European Commission (MEANING IST-2001-34460), the Spanish Ministry of Science and Technology (Hermes TIC2000-0335-C03-03), the University of the Basque Country (9/UPV00141.226-14601/2002) and the Basque Government (ETORTEK2002/HIZKING21).

Bibliography

Abaitua, J., 1988, *Complex predicates in Basque: from lexical forms to functional structures.* PhD disertation, University of Manchester.

—, Aduriz, I., Agirre, E., Alegria, I., Arregi, X., Artola, X., Arriola, J.M., Díaz de Ilarraza, A., Ezeiza, N., Gojenola, K., Maritxalar, M., Sarasola K., Urkia, M., Zubizarreta, J.R., 1992, *Estudio comparativo de diferentes formalismos sintacticos para su aplicación al euskara.* Technical report, UPV/EHU/LSI.

Abney S. P., 1997, "Part-of-Speech Tagging and Partial Parsing". In S. Young and G. Bloothooft, (eds.), *Corpus-Based Methods in Language and Speech Processing,* Kluwer, Dordrecht.

Aduriz I., 2000, *Morfologiatik sintaxira Murriztapen Gramatika baliatuz.* PhD disertation, Department of Basque Philology, University of the Basque Country.

—, Aldezabal, I., Ansa, O., Artola, X., Díaz de Ilarraza, A, Insausti, J.M., 1998, "EDBL: a Multi-Purposed Lexical Support for the Treatment of Basque." *Proceedings of the First International Conference on Language Resources and Evaluation,* LREC'98, Granada.

—, Alegria I., Arriola J., Díaz de Ilarraza, A., Ezeiza, N., Gojenola, K., 2003a, "Finite State Applications for Basque", *EACL'2003 Workshop on Finite-State Methods in Natural Language Processing,* Hungary.

— and Arriola, J.M., Díaz de Ilarraza, A., 2003, "Euskararen murriztapen gramatika. Desanbiguazio morfologikoaren tratamendua, azterketa sintaktikoaren lehen urratsak eta aplikazioak Murriztapen Gramatikaren eredu konputazionala jarraituz". In J.M. Makatzaga and B. Oyharçabal (eds)., *Euskal gramatikari eta literaturari buruzko ikerketak XXI. mendearen atarian. Gramatika gaiak,* Iker-1 (I), Euskaltzaindia, Bilbao, 3-35.

—, Artola, X., Díaz de Ilarraza, A., Gojenola, K., Maritxalar, M., 1997, "Morphosyntactic disambiguation for Basque based on the Constraint Grammar Formalism". *Conference on Recent Advances in Natural Language Processing,* RANLP'97, Bulgaria.

Agirre, E., Arregi, X., Arriola, J.M., Artola, X., Díaz de Ilarraza, A., Insausti, J.M., Sarasola, K., 1995, "Different issues in the design of a general-purpose Lexical Database for Basque". *First Workshop on Applications of Natural Language to Databases,* France.

Aldezabal, I., Goenaga, P., Gojenola, K., Sarasola, K., 1998, "Subcategorización verbal vasca: propuesta inicial y herramienta de validación". *Proceedings of SEPLN'98,* Alicante.

—, Gojenola, K., Sarasola, K., 2000, "A Bootstrapping Approach to Parser Development". *Proceedings of the International Workshop on Parsing Technologies,* IWPT'2000, Trento.

—, Aranzabe, M., Atutxa, A., Gojenola, K., Sarasola, K., Goenaga, P., 2001, "Extracción masiva de información sobre subcategorización verbal vasca a partir de corpus". *Proceedings of SEPLN'2001,* Jaén.

—, Atutxa, A., Aranzabe, M., Gojenola, K., Oronoz, M., Sarasola, K., 2003, "Application of finite-state transducers to the acquisition of verb subcategorization information". To appear in the Special Issue on Finite State Methods in NLP, *Journal of Natural Language Engineering,* Cambridge University Press.

Alegria, I., Artola, X., Sarasola, K., Urkia, M., 1996, "Automatic morphological analysis of Basque", *Literary & Linguistic Computing* 11.4, 193-203. Oxford University Press. Oxford.

Arriola, J.M., 2000, Hauta-Lanerako Euskal Hiztegi-*ko informazio lexikalaren erauzketa erdi-automatikoa eta bere integrazioa sistema konputazional batean.* PhD disertation, Department of Basque Philology, University of the Basque Country.

Bresnan, J., 1982, *The Mental Representation of Grammatical Relations.* The MIT Press, Cambridge, Massachusetts.

Briscoe T., Carroll, J., 1997, "Automatic Extraction of Subcategorization from Corpora", *Proceedings of the Conference on Applied Natural Language Processing, ANLP-97,* EEUU.

Carroll, G., Rooth, M., 1998, "Valence Induction with a Head-Lexicalized" PCFG", *Proceedings of the Conference on Empirical Methods in Natural Language Processing,* Granada.

—, Minen, G., Briscoe, T., 1998, "Can Subcategorisation Probabilities Help a Statistical Parser?". In *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora,* Montreal.

Chomsky, N., 1957, *Syntactic structures.* The Hague: Mouton.

Collins, M., 1997, "Three New Probabilistic Models for Statistical Parsing". In *Proceedings of the Conference of the Association for Computational Linguistics, ACL'97,* Madrid.

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

Gazdar, G., Klein, E., Pullum, G., Sag, I., 1985, *Generalized Phrase Structure Grammar.* Harvard University Press.

Goenaga, P., 1980, *Gramatika bideetan.* Erein.

Grishman, R., Macleod, C., Meyers, A., 1994, "Complex syntax: building a computational lexicon". In *Proceedings of the Conference on Computational Linguistics, COLING-94,* Japan.

Karlsson, F., Voutilainen, A., Heikkila, J., Anttila, A., 1995, *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text.* Mouton de Gruyter.

Karttunen, L., Chanod, J-P., Grefenstette, G., Schiller, A., 1997, "Regular Expressions for Language Engineering", *Journal of Natural Language Engineering,* Cambridge University Press.

Koskenniemi, K., 1983, *Two-level morphology: A general computational model for word-form recognition and production.* PhD disertation, University of Helsinki.

Kuhn, J., Eckle-Kohler, J., Rohrer, C., 1998, "Lexicon Acquisition with and for Symbolic NLP-Systems: a Bootstrapping Approach". In *Proceedings of the International Conference on Language Resources and Evaluation, LREC'98,* Granada.

Sammuelson, C., Voutilainen, A., 1997, "Comparing a Linguistic and a Stochastic Tagger". In *Proceedings of ACL-EACL'97,* Madrid.

Sarasola, I., 1997, *Euskal hiztegia.* Kutxa Fundazioa. Donostia.

Pollard, C., Sag, I., 1994, *Head-Driven Phrase Structure Grammar.* The University of Chicago Press.

Shieber, S.M., 1986, *An Introduction to Unification-Based Approaches to Grammar.* CSLI Lecture Notes, 4 zenbakia, Stanford.

UZEI Institute and the Academy of the Basque Language, 2003, *Statistical corpus of 20th Century Euskara.* http://www.euskaracorpusa.net/XXmendea/index.html

Zubizarreta, J.R., 1992, *Un modelo funcional de diálogo para diálogos orientados por la tarea.* PhD disertation, Department of Computer Languages and Systems, University of the Basque Country.

LEARNING ARGUMENT/ADJUNCT DISTINCTION FOR BASQUE

Izaskun Aldezabal, M.ª Jesús Aranzabe, Koldo Gojenola,
Kepa Sarasola, Aitziber Atutxa

Abstract

This paper presents experiments performed on lexical knowledge acquisition in the form of verbal argumental information. The system obtains the data from raw corpora after the application of a partial parser and statistical filters. We used two different statistical filters to acquire the argumental information: Mutual Information, and Fisher's Exact test. Due to the characteristics of agglutinative languages like Basque, the usual classification of arguments in terms of their syntactic category (such as NP or PP) is not suitable. For that reason, the arguments will be classified in 48 different kinds of case markers, which makes the system fine grained if compared to equivalent systems developed for other languages.

This work addresses the problem of learning subcategorization frames by distinguishing arguments from adjuncts, being the last ones the most significant source of noise in subcategorization frame acquisition.

Introduction

In recent years a considerable effort has been done on the automatic acquisition of lexical information. As several authors point out, this information, mostly subcategorization information, is useful for a wide range of applications. For example, Carroll et al. (1998) show how adding subcategorization information improves the performance of a parser (automatic syntactic analyzer). With this in mind, our aim is to build a system that automatically obtains subcategorization frames. The following figure shows the general schema of a subcategorization acquisition system.

The basic idea behind any system like the one presented in this paper is the following. Starting from a corpus, syntactic information is attained as a result of a parsing phase. As a consequence, each verb will get a set of frames assigned to it. These frames represent the different syntactic environments in which the verb appeared in the corpus. Once these frames are available, statistical filters apply to distinguish subcategorized elements from non-subcategorized ones. As we can see in figure 1. there are two ways to perform this filtering. (A) consists in applying the filters to verb-case pairs to distinguish subcategorized elements (arguments) and non-subcategorized ones (adjuncts). (B) consists in applying statistics directly to the frames to distinguish

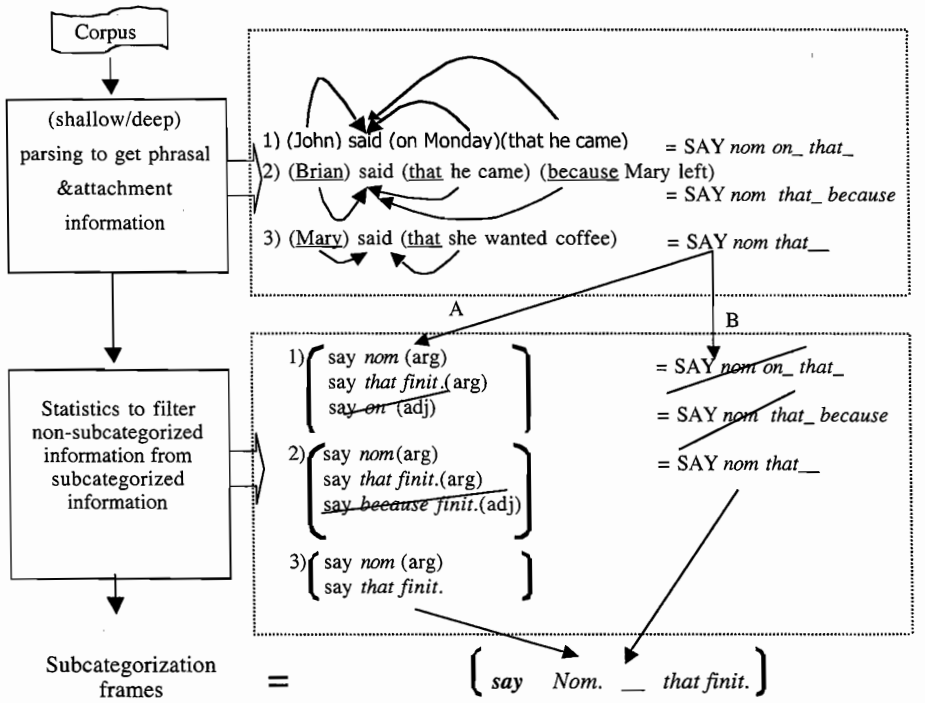


Figure 1. A general schema for a subcategorization acquisition system

subcategorization frames from appearing frames. Following the first filtering, an additional step is required to achieve subcategorization frames; to go back to the original frames and eliminate the elements considered to be adjuncts, because original frames without adjuncts are supposed to be subcategorization frames. The second filtering yields to subcategorization frames directly. The system presented here employs the first filtering approach. We will explain the reasons for this choice in section 2.2.

As we just said, the statistical filters included in the system will perform the argument/adjunct discrimination. But it is well known that this is not a trivial task since there is no clear cut between arguments and adjuncts. However, we decided to pursue it, but under certain limitations, both theoretic and pragmatic.

As for the evaluation, we first evaluated performance of the statistical filters in the argument/adjunct distinction. Second, we evaluated the quality of whole acquired subcategorization frames. We approached the first evaluation (the filter evaluation for the argument/adjunct distinction) in two different fashions; one way consisted in evaluating the resulting list of verb-case marker pairs (tagged either as argument or adjuncts), with the values a human would assign to each verb-case marker pair in the list automatically obtained. Note that the annotator did not have more context than the list of verb and the cases. The second way consisted in selecting some sentences and evaluating over these sentences (that is to say, within a sentential context); again, the statistical filter marked each case phrase the parser attached to the verb in the sentence as argument or adjunct. We compared this marking with the values (argument/adjunct)

assigned by the human annotator to those same verb and case-phrases, but note that the annotator was provided with the sentence and therefore could make use of the sentential context to establish the meaning of the verb. Both methods of evaluation yield significantly different results. Evaluating this way, we wanted to reach some conclusions on the importance of the context for the argument adjunct distinction task. Finally, we also evaluated subcategorization frames obtained using the results of the statistical filter by manually annotating each subcategorization frame obtained by the machine as correct or incorrect. In this case we did not make use of the sentential context.

The paper is divided into five sections. The first section is devoted to explain the theoretical motivations underlying the process. The second section is a description of the different stages of the system. The third and fourth sections present the results obtained by the application of statistical filters to discriminate arguments from adjuncts, and the results of the whole subcategorization frame acquisition, respectively. The fifth section reviews previous work on automatic subcategorization acquisition. And last but not least, we present the main conclusions.

1. The argument/adjunct distinction

As said before, Carroll et al. (1998) showed how adding subcategorization information improves the performance of a parser. Their experiment was developed for English, which is considered to be a fix word order language and head initial.

1. Josuk alde egin zuen etxetik bere amarekin jateko.
   Josu-erg left aux home-from his mother-with eat-to

   "Josu left home to eat with his mother"

2. Josuk alde egin zuen seietan bere amarekin jateko
   Josu-erg left aux. six-at his mother-with eat-to

"Josu left at six to eat with his mother" or "Josu left to eat with his mother at six"

Both *etxetik* (from home) and *seietan* (at six) are postpositional phrases superficially appearing in between *alde egin* (to leave) and *jateko* (to eat), so in principle, and without the help of any subcategorization information, the parser would not be able to decide where to attach in each case. It would treat both the same way. Either it would consider that in both cases these intermediate postpositional phrases are attached to both verbs, or either it would have to make a heuristic decision. For example attach them to the first verb.

Subcategorization information would allow performing the right attachment of the ablative case (*from*) since the ablative is subcategorized by *alde_egin* (to leave) and not by *jan* (to eat). It would also make possible to attach correctly the inessive case to both verbs because the inessive case (at) is not subcategorized by either *alde egin* (to leave) or *jan* (to eat). At this point, we hope we have shown the importance of learning and applying subcategorization information. But such an enterprise is as difficult as important. The argument/adjunct distinction is probably one of the most unclear issues in linguistics. The distinction was presented in the early days in the following

way: subcategorized elements (arguments) are those elements appearing obligatorily while non subcategorized elements (adjuncts) are not. Nowadays we know that this definition is too naive. Several problematic cases are not considered under this definition, for example under-specified elements, elements showing dative case, object shift constructions and so on.

a. Under specified elements

3. I arrived safely.

In principle, *arrive* is taken to be an unaccusative verb, with a single argument.

4. I arrived safely at the station

But in this sentence, *at the station* seems to be an argument too.

b. Object shift constructions

5. I loaded the truck
6. I loaded bricks on the truck

Would we say that the subcategorization is different for these last two cases?

Another definition considers as subcategorized elements those ones participating in the event and as non subcategorized those ones contextualizing or locating the event. This is a semantic definition of what an argument is.[1] It is still not clear, in example 4, whether *at the station* would be an argument or an adjunct. One could say that it participates in the event since it marks the end of the event. Under some aspectual thesis (Tenny 1987) both *the truck* and *on the truck* could also be considered as participants of the event, again because they mark the end of the event. But leaving aside aspectual issues, take a look to the following examples:

7. Yesterday I talked with Mary.
8. Yesterday I played soccer with Mary.

Here, *Mary* is a participant of the event in both cases, therefore under the given definition in both cases *Mary* would be a subcategorized element. But this is contradictory to what traditional views consider in practice. *To play* does not require two participants (though it can have them), while *to talk* (under the sense of communicating) seems to require two participants.

Finer argument/adjunct distinctions have also been proposed differentiating between basic arguments, pseudo-arguments and adjuncts. Basic arguments are those required by the verb. Pseudo-arguments are those that even if they are not required by the verb, when appearing they extend the verbal semantics, for example, adding new

[1] It would be also syntactic because depending wether it is a participant or not the elements will get projected in different positions (external or internal predication).

participants. And finally adjuncts, which would be contextualizers of the event. (For further reference on the argument/adjunct distinction see Gawron 1986, Grimshaw 1990, Schutze 1995, Verspoor 1997).

Though there is an extensive literature on subcategorization, up to day, we did not find a way to establish a clear cut between subcategorized and non subcategorized elements. Nevertheless, from the different diagnostics proposed in the literature some are quite consistent among various authors (Pollard and Sag 1987, Grishman et al. 1994, Verspoor 1997):

1. Obligatoriness condition. When a verb demands obligatorily the appearance of an element, this element is an argument.

   a. John put the book *on the table*
   b. *John put the book

2. Frequency. Arguments of a verb occur more frequently with that verb than with the other verbs.

   a. I came *from home* (argument).
   b. I heard it *from you* (adjunct).

3. Iterability. Several instances of the same adjunct can appear together with a verb, while several instances of an argument cannot appear with a verb.

   a. I saw you in Washington, in the Kenedy Center.
   b. *I saw you in Washington, in N.Y.

4. Relative order. Arguments tend to appear closer to the verb than adjuncts.

   a. I put the book on the table *at three*
   b. *I put *at three* the book on the table

5. Implicational test. Arguments are semantically implied, even when they are optional.

   a. I came to your house (from x)
   b. I heard that (from x)

The third and fourth tests were not very useful to us. Iterability test is quite weak since it seems to rely more on some other semantic notions such as part/whole relation than on the argument/adjunct distinction. For example, sentence 3.a would be grammatical due to semantic plausibility. *The Kennedy Center* is part of *Washington*, therefore to see somebody in *the Kennedy Center* and see him in *Washington* are not semantically incompatible, so it is plausible to say it. In the case of 3.b *N.Y* is not a part of *Washington* and therefore it is not plausible to see (in the same event) somebody in two different places.

The relative order test is difficult to apply on a free word order language like Basque. The first and fifth tests are robust enough to be useful in practice. But only the two first diagnostics can be captured statistically by the application of association measures like Mutual Information. We did not come out with any straightforward way to apply the fifth test computationally.

Before introducing the different statistical measures applied, we will present step by step the whole process we pursued for achieving the argument/adjunct distinction. Talking about Subcategorization Frames (SCF) means talking about arguments. Many existing systems acquire directly a set of possible SCFs without any previous filtering of adjuncts. However, adjuncts are a substantial source of noise and sparseness.[2] If we wanted to acquire directly the right subcategorization frames without making any previous filtering we would need more than the million and a half words that we have. The reason is that on the basis of verb-case markers (of course obtained from the frames appearing in the corpus) we can apply some statistics because the arguments appear more frequently than adjuncts, because they appear in more frames, and the frequency distinction is usually relevant enough as to be able to apply statistics on it. But it is not so frequent to see a bare real subcategorization frame (in other words, a frame where all the cases are only arguments). In most of the cases there is an adjunct, and moreover the range of different adjuncts is huge. This means that the argument and adjunct combination number into frames is very high besides, the frequency distinction between the combinations is not relevant enough. Therefore we decided to pursue the argument/adjunct distinction as a way to obtain real subcategorization frames (option A in Figure 1).

2. The acquisition process

Our starting point was a raw newspaper corpus from of 1.337.445 words, where there were instances of 1.412 verbs. From them, we selected 640 verbs as statistically relevant because they appear in more than 10 sentences.

As we said earlier, our goal was to distinguish arguments from adjuncts. When starting from raw corpus, like in this case, it is necessary to get instances of verbs together with their dependents (arguments and adjuncts). We obtained this information applying a partial parser (section 2.1) to the corpus. Once we had the dependents, statistical measures helped us deciding which were arguments and which were adjuncts (section 2.2).

2.1. The parsing phase

Aiming to obtain the data against which statistical filters will be applied, we analyzed the corpus using several available linguistic resources (for more information see Aldezabal et al., in this volume):

— First, we performed morphological analysis of the corpus, based on two-level morphology (Koskenniemi 1983, Alegria et al. 1996) and disambiguation using the Constraint Grammar formalism (Karlsson et al. 1995, Aduriz et al. 1997).

[2] When the frequency of an event is too distributed into different occurrences, and the frequency of each occurrence is very similar. So statistically there is no occurrence that is more significant than the others.

— Second, a shallow parser was applied (Aldezabal et al. 2000), which recognizes basic syntactic units including noun phrases, prepositional phrases and several types of subordinate sentences.

1. ... *(a) [EEBBetako lehendakariak] (b) [UEko 15 herrialdeetako merkataritza ministroekin] (c) [bazkaldu zehar zuen] (d) [negoziazioen bilgunean]...*
2. ... the president of the USA had to eat with the ministers of Commerce of 15 countries of the UE in the negotiation center...

a) *[EEBB-etako lehendakari-a-k]*
[USA-of president-the-erg.]

NP-ergative (president, singular)
The president of the USA

b) *[UE-ko 15 herrialde-etako merkataritza ministro-ekin]*
[UE-of 15 countries-of Commerce ministers-with]

PP (with)-commitative (minister, plural)
with the ministers of Commerce of 15 countries of the UE

c) *[bazkaldu behar zuen]*

[to eat had]
verb (eat)
had to eat

d) *[negoziazio-en bilgune-an]*
[negotiation-of center-in]

PP (in)-inessive (center, singular)
in the negotiation center

Figure 2. Example of the output of the shallow parsing phase: 1) Input (in Basque), 2) English translation, Below (c) Verb phrase and (a,b,c) verbal dependents (phrases), and also case+head information

— The third step consisted in linking each verb and its dependents. Basque lacks a robust parser as in (Briscoe and Carroll 1997, Kawahara et al. 2001) and, therefore, we used a finite state grammar to link the dependents (both arguments and adjuncts) with the verb (Aldezabal et al. 2001). This grammar was developed using the Xerox Finite State Tool (Karttunen et al. 1997). Figure 2 shows the result of the parsing phase. In this case, both commitative and inessive cases (PPs) are adjuncts, while the ergative NP is an argument.

The linking of dependents to a verb is not trivial considering that Basque is a language with free order of constituents, and any element appearing between two verbs could be, in principle, dependent on any of them. Many problems must be taken into account, such as ambiguity and determination of clause boundaries, among others. We evaluated the accuracy up to this point, obtaining a precision over dependents of 87% and a recall of 66%. So the input data to the next phase was relatively noisy.

2.2. The argument selection phase

In the data resulting from the shallow parsing phase we counted up to 65 different cases (types of arguments, including postpositions and different types of suffixes). These are divided in two main groups:

— 43 correspond to postpositions. Some of them can be directly mapped to English prepositions, but in many cases several Basque postpositions correspond to just one English preposition. This set also contains postpositions that map to categories other than English prepositions, such as adverbs.
— 22 types of sentential complements (For instance, English *that* complementizer corresponds to several subordination suffixes: *-la, -n, -na, -nik).*

This shows to which extent the range of arguments is fine grained, in contrast to other works where the range is at the categorial level, such as NP or PP (Brent 1993, Manning 1993, Merlo and Leybold 2001).

Due to the complexity carried by having such a high number of cases, we decided to gather postpositions that are semantically equivalent or almost equivalent (for example, English *between* and *among*). Even if there are some semantic differences between them they do not seem to be relevant at the syntactic level. Some linguists were in charge of completing this grouping task. Even considering the risk of making mistakes when grouping the cases, we concluded that the loss of accuracy due to having too sparse data (consequence of having many cases) would be worse than the noise introduced by any mistake in the grouping. The resulting set contained 48 cases. The complexity is reduced but it is still considerable.

Most of the work on automatic acquisition of subcategorization information (Carroll and Briscoe 1997, Sarkar and Zeman 2000, Korhonen 2001) apply statistical methods (hypothesis testing). Basically the idea is the following: they get "possible subcategorization frames" from automatically parsed data (either completely or partially parsed) or from a manually annotated corpus. Afterwards a statistical filter is employed to decide whether those "possible frames" are or not real subcategorization frames (option B in Figure 1). These statistical methods can be problematic mostly because they perform badly on sparse data. In most of the cases the systems pursuing this approach (option B) are able to decrease the noise because they already have some subcategorization information coming from dictionaries (Carroll and Briscoe 1997). In our case, there is no dictionary carrying such information, therefore and in order to avoid as much as possible data sparseness, we decided to design a system that starts learning the arguments/adjuncts of a given verb instead of learning whole frames. Frames are combinations of arguments, and considering that our system deals with 48 cases, the number of combinations was high, resulting in sparse data. So we decided to work at the level of the argument/adjunct distinction. Working on this distinction is also very useful to avoid noise in the subcategorization frame, since in this task adjuncts are synonyms of noise. A system that tries to get subcategorization frames without previously making the argument/adjunct distinction suffers of having sparse and noisy data.

To accomplish the argument/adjunct distinction we applied two measures: Mutual Information (MI), and Fisher's Exact Test (for more information on these measures, see

Manning and Schütze 1999). MI is a measure coming from Information Theory, defined as the logarithm of the ratio between the probability of the co-occurrence of the verb and the case, and the probability of the verb and the case appearing together calculated from their independent probability.

$$MI = \log \frac{P(\text{verb, case})}{P(\text{verb})\, P(\text{case})}$$

So higher Mutual Information values correspond to higher associated verb and cases (see table 1).

Table 1. Examples from MI values for verb-case pairs

Verb	case	MI
atera (to take/go out)	Ablative (from)	1,830
atera (to take/go out)	instrumental (with)	–0,955
erabili (to use)	*gisa* (as)	2,255
erabili (to use)	instrumental (with)	–0,783

Mutual Information shows higher values for *atera-ablative* (to go/take out), *erabili-gisa* (to use-as). These pairs were manually tagged as arguments, therefore Mutual information makes the right prediction. On the contrary, *atera-instrumental* (to go/take out-with), *erabili-instrumental* (to use-with) were manually tagged as adjuncts. Mutual Information values in table 1 go along with the manual tagging for these last pairs as well, because these Mutual Information values are low as should correspond to adjuncts.

Fisher's Exact Test is a hypothesis testing statistical measure.[3] We used the left-side version of the test (see Pederssen 1996). Under this version the test tells us how likely it would be to perform the same experiment again and be less accurate. That is to say, if you were repeating the experiment and there were no relation between the verb and the case, you would have a big probability of finding a lower co-occurrence frequency than the one you observed in your experiment. So higher left-side Fisher values tell us that there is a correlation between the verb and the case (see table 2.)

Fisher's Exact values show higher values for *atera-ablative* (to go/take out), *erabili-gisa* (to use-as). These values predict correctly the association between the verbs and cases for these examples. The low values for the *atera-instrumental* (to go/take out-with), and *erabili-instrumental* (to use-with) pairs, should be interpreted as the non-association between the verbs and the cases in these examples, that is to say, they are adjuncts. And again, the prediction would be right according to the annotators.

[3] There are two ways of interpreting Fisher's test, as one or two sided test. In the one sided fashion there is still another interpretation, as a right or left sided test.

Table 2. Examples of Fisher's Exact Test values for verb-case pairs

Verb	case	Fisher
atera (to take/go out)	Ablative (from)	1,0000
atera (to take/go out)	instrumental (with)	0,0003
erabili (to use)	*gisa* (as)	1,0000
erabili (to use)	instrumental (with)	0,0002

These tests are broadly used to discover associations between words, but they show different behaviour depending on the nature of the data. We did not want to make any a priori decision on the measure employed. On the contrary, we aimed to check which test behaved better on our data.

3. Evaluation of the argument/adjunct distinction

We found in the literature two main approaches to evaluate a system like the one proposed in this paper (Briscoe and Carroll 1997, Sarkar and Zeman 2000, Korhonen 2001):

— Comparing the obtained information with a gold standard.
— Calculating the coverage of the obtained information on a corpus. This can give an estimate of how well the information obtained could help a parser on that corpus.

Under the former approach a further distinction emerges: using a dictionary as a gold standard, or performing manual evaluation, where some linguists extract the arguments in a corpus (this would be the gold standard) and compare them with the arguments obtained automatically.

We decided to evaluate the system both ways, that is to say, using a gold standard and calculating the coverage over a corpus. The intention was to determine, all things being equal, the impact of doing it one way or the other.

3.1. Evaluation 1: comparison of the results with a gold standard

From the 640 analyzed verbs, we selected 10 for evaluation. For each of these verbs we extracted from the corpus the list of all their dependents. The list was a set of bare verb-case pairs, that is, no context was involved and, therefore, as the sense of the given verb could not be derived, different senses of the verb were taken into account. We provided 4 human annotators/taggers with this list and they marked each dependent as either argument or adjunct. The taggers accomplished the task three times. Once, with the simple guideline of the implicational test and obligatoriness test, but with no further consensus. The inter-tagger agreement was low (57%). The taggers gathered and realized that the problem came mostly from semantics. While some taggers tagged the verb-case pairs assuming a concrete semantic domain the others took into account a wider rage of senses (moreover, in some cases the senses did not even match). So the tagging was repeated when all of them considered the same semantics to the different

verbs. The inter-tagger agreement raised up to a 80%. The taggers gathered again to discuss, deciding over the non clear pairs.

The list obtained from merging[4] the 4 lists in one is taken to be our gold standard. Notice that when the annotators decided whether a possible argument was really an argument or not, no context was involved. In other words, they were deciding over bare pairs of verbs and cases. Therefore different senses of the verb were considered because there was no way to disambiguate the specific meaning of the verb. So the evaluation is an approximation of how well would the system perform over any corpus. Table 3 shows the results in terms of Precision and Recall.

Table 3. Results of Evaluation 1 (context independent)

	Precision	Recall	F-score
MI	62%	50%	55%
Fisher	64%	44%	52%

Precision measures from the elements marked by the machine as arguments, how many where really arguments, in other words, how many where also tagged as arguments by the human annotators. In this case it tells us that from the elements marked as arguments using MI, 62% were real arguments, the rest either were adjuncts or attachment errors made by the parser that have been considered by the machine as arguments (or elements which were not well attached). As for the elements marked as arguments using Fisher, 64% were real argument, the rest adjuncts or errors. Recall measures, how many of the elements marked as arguments by the human annotators were not marked as such by the machine. That is, how many of the real arguments were left out. F-score is just a way to normalize both precision and recall, so for example MI gets better recall results than Fisher, and Fisher gets better precision results than MI. F-score provides a way to select which one is relatively better considering both precision and recall.

3.2. Evaluation 2: calculation of the coverage on a corpus

The initial corpus was divided in two parts, one for training the system and another one for evaluating it. From the fraction reserved for evaluation we extracted 200 sentences corresponding to the same 10 verbs used in the "gold standard" based evaluation. In this case, the task carried out by the annotators consisted in extracting, for each of the 200 sentences, the elements (arguments/adjuncts) linked to the corresponding verb. Each element was marked as argument or adjunct. Note that in this case the annotation takes place inside the context of the sentence. In other words, the verb shows precise semantics.

We performed a simple evaluation on the sentences (see table 4), calculating precision and recall over each argument marked by the annotators.[5] For example, if a

[4] Merging was possible once the annotators agreed on the marking of each element.

[5] The inter-tagger agreement in this case was of 97%.

verb appeared in a sentence with two arguments and the statistical filters were recognizing them as arguments, both precision and recall would be 100%. If, on the contrary, only one was found, then precision would be 100%, and recall 50%.

Table 4. Results of Evaluation 2 (inside context)

	Precision	Recall	F-score
MI	93%	97%	95%
Fisher	93%	93%	93%

3.3. Discussion

It is obvious that the results attained in the first evaluation are different than those in the second one. The origin of this difference comes mostly, on one hand, from semantics and, on the other hand, from the nature of statistics:

— Semantic source. The former evaluation was not contextualized, while the latter used the sentence context. Our experience showed us that broader semantics (non-contextualized evaluation) leads to a situation where the number of arguments increases with respect to narrower (contextualized evaluation) semantics. This happens because in many cases different senses of the same verb require different arguments. So when the meaning of the verb is not specified, different meanings have to be taken into account and, therefore, the task becomes more difficult.

— Statistical reason. The disagreement in the results comes from the nature of the statistics themselves. Any statistical measure performs better on the most frequent cases than on the less frequent ones. In the first experiment all possible arguments are evaluated, including the less frequent ones, whereas in the second experiment only the possible arguments found in the piece of corpus used were evaluated. In most of the cases, the possible arguments found were the most frequent ones.

At this point it is important to notice that the system deals with non-structural cases. In Basque there are three structural cases (*ergative, absolutive* and *dative*) which are special because, when they appear, they are always arguments. They correspond to the subject, direct object and indirect object functions. These cases are not very conflictive when deciding on their argumenthood,[6] mainly because in Basque the auxiliary bears information about their appearance in the sentence. So they are easily recognized and linked to the corresponding verb. That is the reason for not including them in this work. Precision and recall would improve considerably if they were included because they are the most frequent cases (as statistics perform well over frequent data), and also because the shallow parser links them correctly using the information carried by the auxiliary. Notice that we did not incorporate them because

[6] As we said in section 1, the nature of the dative case is not very clear.

our aim is to use the subcategorization information obtained to help our parser, and the non-structural cases are the problematic ones.

4. Eliminating the adjuncts from then original frames

Until now we presented a part of the system which is able to decide whether a case phrase corresponds to an argument or an adjunct[7] by means of the occurrence frequency of verb-case pairs in the data. Next step consisted in going back to the original case frames obtained by the partial parser, and eliminating the cases tagged by the machine as adjuncts. Remember that the partial parser tries to attach the case phrases surrounding the different verbs to the corresponding verb. This way, for each verb in a sentence, the parser will provide a frame, or in other words, the combination of case phrases attached to it. This is what we would call an original frame. We used the list resulting from the application of MI. Thus, for example, take *bazkaldu* (to have lunch). The frames obtained by the parser are the following ones:

1. occurrences ### 8,3 DU-erg[8]
2. occurrences ### 3,8 DU-erg-ine
3. occurrences ### 2,9 DU-erg-soc
4. occurrences ### 1 DA-abs-ala
5. occurrences ### 1 DU-abl-erg-ine
6. occurrences ### 1 DU-abs-erg
7. occurrences ### 1 DU-abs-erg-ine-soc
8. occurrences ### 0,7 DA-abs
9. occurrences ### 0,2 DA-abs-ine
10. occurrences ### 0,1 DA-abs-soc

Figure 3. Frames obtained by the parser for the verb *bazkaldu* (to eat)

As we said before, by applying the statistical filters the system got for each verb the list of arguments and adjuncts.

bazkaldu ine: 0,504482
bazkaldu soc: 2,065221
bazkaldu ala: 0,210678
bazkaldu abl: 0,430152

Figure 4. List of arguments/adjuncts obtained by the parser for the verb *bazkaldu* (to eat)

[7] Or an error coming from the heuristics applied by the parser to attach the different phrases to the verbs.

[8] Remember that we did not recover the absolutive case when the auxiliaries are DU or DIO since it is quite usual to find incorporation of the internal argument into the verb with some transitive verbs.

These mutual information values tell us that *sociative case* (with) is an argument *bazkaldu(to have lunch)*.[9] Now, as said before, all cases but the *sociative* (with) will be eliminated from the initial frames, and the result is:

1. occurrences ### 13,1 DU-erg[10]
2. occurrences ### 2,9 DU-erg-soz
3. occurrences ### 1 DU-abs-erg
4. occurrences ### 1 DU-abs-erg-soz
5. occurrences ### 1,9 DA-abs
6. occurrences ### 0,1 DA-abs-soz

Figure 5. Frames obtained for the verb *bazkaldu* (to eat) after eliminating adjuncts

4.1. Evaluation

Once, we got these new frames, our goal was to see if these new frames could be considered as the real subcategorization frames. We know that certain cases are always adjuncts for a given verb, but there are also some cases acting either as arguments or adjuncts depending on the frames they appear in. More over, sometimes the frame in which that case is an argument, and the frame in which that same case acts as an adjunct belong to two different meanings of the verb. For example, consider the following case and frames:

atera 6,95061728395062 ### DA-abs-ala-ine
atera 41,3703703703704 ### DA-abs-ine

Figure 6. Examples of frames obtained for the verb *atera* (to go out/to publish)

If the machine was marking the *inessive case* as adjunct, we would go back to these frames and erase the *inessive case* from them, without making a further distinction. The problem comes from the meaning associated to each of these frames. When looking at the examples we noted that for the first frame the *inessive case* is really an adjunct

[9] In this case, it seems that the machine makes a mistake, but when we take a look to the examples one realizes that eat appears meaning *to gather* or *to meet*. So we will go back to the original frames, and the other cases will be eliminated.

[10] Remember that we did not recover the absolutive case when the auxiliaries are DU or DIO since it is quite usual to find incorporation of the internal argument into the verb with some transitive verbs.

because it is associated to *atera* (to go out) and as movement verb the inessive acts as an adjunct. Contrastively, for the second frame, the meaning changes and *atera* is not a movement verb, it would be equivalent to the English *to publish*. In this case, *inessive* would not be an adjunct but an argument.

Going back to the evaluation, the results were obtained as follows: the manual annotators were provided with both the list of these new frames obtained by the machine by deleting the adjuncts and the list of the original frames obtained initially by the parser. The annotators were marking in both lists each frame as correct or incorrect for the given verb. This time they did not have any sentential context to make the decision, again the decisions were made over raw lists of verbs and frames, therefore they could not know the meaning of the verb associated to each frame.

Table 5. Results of the frames evaluation

	Precision	Recall	F-score
Eliminate adjuncts from initial comb. (688 → 144)	52%	75%	61%

In this case precision expresses how many frames, from the number of frames the machine marked as subcategorization frames, are really subcategorization frames. Therefore one could say that precision meassures the quality of the data obtained. Recall meassures how many real subcategorization frames were discovered by the machine. For doing that, we take the original list of frames got initially by the parser and we tagged them as real subcategorization frames or errors. And recall was calculated by taking the number of frames marked as real subcategorization frames from the list obtained after eliminating the adjuncts and dividing this number by the number of frames marked as real subcategorization frames in the original list. This way we can get an idea of the lost of information when eliminating the adjuncts.

4.2. Discussion

The approach of eliminating the adjuncts is useful for acquiring subcategorization frames. We were able to reduce sparseness. After eliminating the adjuncts the total number of frames decreased from 688 to 144. This happens because once the adjuncts are eliminated, we found a lot of combinations that were different because of an adjunct and once that adjunct disappeared, the frames could be merged because they were the same frame. So the frequencies linked to them could be added. This way, we are able to get relevant frequency distinctions and a lower number of case combinations (frames) for each verb.

We also have to consider the loss of information. As the recall measure shows we lost 25% of subcategorization frames. That means that when eliminating adjuncts, due to errors, we eliminated arguments also, and therefore we lost correct subcategorization frames that were originally captured before the argument/adjunct filtering occurred.

5. Related work

Concerning the acquisition of verb subcategorization information, there are proposals ranging from manual examination of corpora (Grishman et al. 1994) to fully automatic approaches. Table 6, partially borrowed from Korhonen (2001), summarizes several systems on subcategorization frame acquisition.

Manning (1993) presents the acquisition of subcategorization frames from unlabelled text corpora. He uses a stochastic tagger and a finite state parser to obtain instances of verbs with their adjacent elements (either arguments or adjuncts), and then a statistical filtering phase produces subcategorization frames (from a set of previously defined 19 frames) for each verb.

Briscoe and Carroll (1997) describe a grammar based experiment for the extraction of subcategorization frames with their associated relative frequencies, obtaining 76,,6% precision and 43,4% recall. Regarding evaluation, they use the ANLT and COMLEX Syntax dictionaries as gold standard. They also performed evaluation of coverage over a corpus. For our work, we could not make use of any previous information on subcategorization, because there is nothing like a subcategorization dictionary for Basque.

Sarkar and Zeman (2000) report results on the automatic acquisition of subcategorization frames for verbs in Czech, a free word order language. The input to the system is a set of manually annotated sentences from a treebank, where each verb is linked with its dependents (without distinguishing arguments and adjuncts). The task consists in iteratively eliminating elements from the possible frames with the aim of removing adjuncts. For evaluation, they give an estimate of how many of the obtained frames appear in a set of 500 sentences where dependents were annotated manually, showing an improvement from a baseline of 57% (all elements are adjuncts) to 88%. Comparing this approach to our work, we must point out that Sarkar and Zeman's data does not come from raw corpus, and thus they do not deal with the problem of noise coming from the parsing phase. Their main limitation comes by relying on a treebank, which is an expensive resource.

Kawahara et al. (2000) use a full syntactic parser to obtain a case frame dictionary for Japanese, where arguments are distinguished by their syntactic case, including their headword (selectional restrictions). The resulting case frame components are selected by a frequency threshold.

Maragoudakis et al. (2001) apply a morphological analyzer and phrase chunking module to acquire subcategorization frames for Modern Greek. In contrast to this work, they use different machine learning techniques. They claim that Bayesian Belief Networks are the best learning technique.

Merlo and Leybold (2001) present learning experiments for automatic distinction of arguments and adjuncts, applied to the case of prepositional phrases attached to a verb. She uses decision trees tested on a set of 400 verb instances with a single PP, reaching an accuracy of 86,5% over a baseline of 74%.

Note that both Manning and Merlo and Leybold's systems learn from contexts with just one PP (maximum) per verb (finite state filter). Our system learns from contexts with up to 5 PPs. Furthermore, we distinguish 48 different kinds of cases, hence the number of combinations is considerably bigger.

Table 6. Summary of several systems on subcategorization information

Method	Number of frames	Number of verbs	Linguistic resources	F-Score (evaluation based on a gold standard)	Coverage on a corpus
C. Manning (1993)	19	200	POS tagger + simple finite state parser	58	
T. Briscoe & J. Carroll (1997)	161	14	Full parser	55	
A. Sarkar & D. Zeman (2000)	137	914	Annotated treebank	—	88
D. Kawahara et al. (2000)	—	23,497	Full parser		82 accuracy
M. Maragoudakis et al. (2001)	—	47	Simple phrase chunker	77	
This paper	—	640	Morph. Analyzer + Phrase Chunker + Finite State Parser	55	95

Regarding the parsing phase, the systems presented so far are heterogeneous. While Manning, Merlo and Leybold and Maragoudakis et al. use very simple parsing techniques, Briscoe and Carroll and Kawahara et al. use sophisticated parsers. Our system can be placed between these two approaches. The result of the shallow parsing is not simple in that it relies on a robust morphological analysis and disambiguation. Remember that Basque is an agglutinative language with strong morphology and, therefore, this stage is particularly relevant. Moreover, the finite state filter we used for parsing is very sophisticated (Karttunen et al. 1997, Aldezabal et al. 2001), compared to Manning's.

Conclusion

This work describes an initial effort to obtain subcategorization information for Basque. To successfully perform this task we had to go deeper than mere syntactic categories (NP, PP...) enriching the set of possible arguments to 48 different classes. This leads to quite sparse data. Together with sparseness, another problem common to every subcategorization acquisition system is that of noise, coming from adjuncts and incorrectly parsed elements. For that reason, we defined subcategorization acquisition in terms of distinguishing between arguments and adjuncts.

The system presented was applied to a newspaper corpus. Subcategorization acquisition is highly associated to semantics in that different senses of a verb will most of the times show different subcategorization information. Thus, the task of learning subcategorization information is influenced by the corpus. As for the evaluation of this

work, we carried out two different kinds of evaluation of the argument/adjunct distinction results. This way, we verified the relevance of semantics in this kind of task.

For the future, we plan to incorporate the information resulting from this work in our parsing system. We hope that this will lead to better results in parsing. Consequently, we would get better subcategorization information, in a bootstrapping cycle. We also plan to improve the results by using semantic information as proposed in A. Korhonen (2001).

Acknowledgements

This research has been supported by the European Commission (MEANING IST-2001-34460), the Spanish Ministry of Science and Technology (Hermes TIC2000-0335-C03-03), the University of the Basque Country (9/UPV00141.226-14601/2002) and the Basque Government (ETORTEK2002/HIZKING21).

References

Aduriz, I., J.M. Arriola, X. Artola, A. Díaz de Ilarraza, K. Gojenola and M. Maritxalar, 1997, «Morphosyntactic disambiguation for Basque based on the Constraint Grammar Formalism». Conference on Recent Advances in Natural Language Processing (RANLP).

Aldezabal, I., K. Gojenola and K. Sarasola, 2000, "A Bootstrapping Approach to Parser Development". *International Workshop on Parsing Technologies (IWPT),* Trento.

—, M. Aranzabe, A. Atutxa, K. Gojenola, M. Oronoz M. and Sarasola K., 2001, "Application of finite-state transducers to the acquisition of verb subcategorization information". *Finite State Methods in Natural Language Processing, ESSLLI Workshop,* Helsinki.

Alegria, A., X. Artola, K. Sarasola and M. Urkia, 1996, "Automatic morphological analysis of Basque", *Literary and Linguistic Computing.* 11.4, Oxford University.

Brent, M.R., 1993, "From Grammar to Lexicon: Unsupervised Learning of Lexical Syntax". *Computational Linguistics* 19, 243-262.

Briscoe, T. and J. Carroll, 1997, "Automatic Extraction of Subcategorization from Corpora", ANLP-97: 356-363.

Carroll, J., G. Minnen and T. Briscoe, 1998, "Can Subcategorization probabilities help a statistical parser?", *Proceedings of the 6th ACL/SIGDAT Workshop on Very Large Corpora,* Montreal.

Gawron, J.M., 1986, "Situations and prepositions", *Linguistics and Philosophy* 9.3, 327-382.

Grimshaw, J., 1990, *Argument Structure.* Cambridge, MA, MIT Press.

Grishman, R., C. Macleod, A. Meyers, 1994, *Complex Syntax: Building a Computational Lexicon.* COLING-94.

Karlsson, F., A. Voutilainen, J. Heikkila, A. Anttila, 1995, *Constraint Grammar: A language-independent system for parsing unrestricted text.* Mouton de Gruyter.

Karttunen, L., J.P. Chanod, G. Grefenstette, A. Schiller, 1997, "Regular expressions for language engineering", *Natural Language Engineering.*

Kawahara, D., N. Kaji and S. Kurohashi, 2000, "Japanese Case Structure analysis by unsupervised construction of a case frame dictionary", COLING-2000, Saarbrucken.

Korhonen, A., 2001, *Subcategorization acquisition.* Unpublished PhD dissertation, University of Cambridge.

Koskenniemi, K., 1983, *Two-level morphology: A general computational model for word-form recognition and production.* PhD dissertation, University of Helsinki.

Manning, C.D., 1993, "Automatic acquisition of a large subcategorization dictionary from corpora". *Proceedings of the 31th ACL.*

— and H. Schütze, 1999, *Foundations of statistical natural language processing.* The MIT Press, Cambridge, Massachusetts.

Maragoudakis, M., K. Kermanidis, N. Fakotakis and G. Kokkinakis, 2001, "Learning automatic acquisition of subcategorization frames using bayesian inference and support vector machines". *The 2001 IEEE International Conference on Data Mining, IMDC'01,* San José.

Merlo, P. and M. Leybold, 2001, "Automatic distinction of arguments and modifiers: the case of prepositional phrases". EACL-2001, Toulouse.

Pederssen, T., 1996, "Fishing for exactness". *Proceeding of the South-Central SAS User Group Conference (SCSUG-96).*

Pollard, C., and I. Sag, 1987, "An information based Syntax and Semantics", volume 13. CSLI lecture Notes, Standford University.

Sarkar, A., and D. Zeman, 2000, "Automatic extraction of subcategorization frames for Czech". COLING-2000, Saarbrucken.

Schutze, C., 1995, "PP Attachment and Argumenthood", *MIT Working Papers in Linguistics.*

Tenny, C., 1994, *Aspectual Roles and the Syntax-Semantic interface,* Kluwer Academic Publishers.

Verspoor, C., 1997, *Contextually-Dependent Lexical Semantics.* PhD dissertation, Brandeis University, MA.

ANALYZING VERBAL SUBCATEGORIZATION AIMED AT ITS COMPUTATIONAL APPLICATION

Izaskun Aldezabal and Patxi Goenaga

Abstract

The verb is one of the most important lexical components: it includes information regarding the necessary components that make up sentences and their features. This is precisely the domain of the analysis of subcategorization. However, specifying the subcategorization of each verb is a difficult task, mainly because of the following reasons: first, because the distinction of the semantic values and the alternations in each verb is problematic; and second, because of the presence of certain phenomena such as ellipsis, unspecification (of general and specific elements), and dependencies between Cases.

This work presents the following: after having reviewed the complex phenomena that are involved in verbal subcategorization, and contextualized these in our research area (i.e., in computational linguistics), we explain the procedure adopted to analyze 100 selected verbs, where Levin (1993) has been taken as point of departure. Once the research has been completed, we have defined what we have considered as subcategorization, namely, all the semantic/syntactic value(s) that we have defined for each verb (ssv), the set of outstanding elements in each ssv, their semantic specifications, and their Case realizations. Thus, we have tried to provide a coherent proposal as a base for grouping verbs depending on the goal.

1. Introduction: the need for sucategorization

Research on lexical components has become increasingly relevant in current theoretical and computational analyses for two reasons: first, because lexical information is the basic information that feeds other levels such as morphosyntax, syntax, semantics, etc., and second, because lexical information imposes conditions that determine the grammaticality and intelligibility of sentences. The verb is one of the most important lexical components. In fact, the verb includes information regarding the necessary components that make up sentences and their features. This is precisely the domain of the analysis of subcategorization.

The fact that, since the advent of generative grammar various proposals have arisen for defining the lexicon, suggests that this task involves many complications. As for us, Computational Linguists, we typically analyze real corpora, i.e., texts that are part of the everyday use of the Basque language. Thus, real corpora are the point of departure for all our analyses. In order to know and define the characteristic features of real

corpora, it is necessary to systematize a big number of phenomena. However, sources offered by general linguistics for this task have proven to be too scarce. The fact that corpora are the starting locus in computational linguistics implies two issues: on the one hand, the sentences under analysis are real sentences, and hence, we will encounter all types of sentences: long, short, grammatical and non-grammatical. On the other hand, sentences in real corpora are set in specific contexts.

All this suggests that we are dealing with components that still need analyzing in theoretical research. In other words, the tools that are available in theoretical linguistics are not sufficient to respond to the demands of automatic resources. One clear example is verbal subcategorization. Thus, computational linguistics adapts its resources by using the information that is available at the time, and it considers other ways in order to continue the research. The latter suggests that computational linguistics sets its own line of research largely.

Along these lines, the computer considers the corpus as a mere string of characters, and thus, the first step usually involves the analysis of the composition of words. Yet, the strings of characters that make up the corpus do not appear in isolation. Rather, they are set in specific contexts, and hence, it is necessary to predict the possible interpretations of words in connection with other surrounding words. Consider the following example: the word *iritziak* ('opinions') may appear in sentences like *Alkatearen iritziak herritarrak harritu ditu* ('The mayor's opinion has surprised the citizens') or in *Egunkariek herritarren iritziak plazaratu dituzte* ('Newspapers have published the opinions of the citizens'). Specifically, the word *iritziak* may appear in Ergative Singular or Absolutive plural. Moreover, *iritzi* has an ambiguous categorical status, and it may be a noun or a verb. To make matters worse, it may appear in a string like *iritzi dio* ('he/she believes'), where *iritzi* surfaces in the participial perfective form. All this implies that, were we to analyze such forms in isolation, they would be ambiguous, i.e., they would have various interpretations. Nevertheless, in order to advance into syntax, we need to cut such ambiguities by disambiguating processes. Among the possible analyses of the word, this process selects a single analysis (i.e., the correct one that corresponds to the context under consideration).

Here are the steps that we have taken to analyze sentences in real corpora:

a) The basis is a database, which includes the necessary information to morphologically isolate and analyze all the words in a sentence: the Basque Lexical Database (i.e., Euskararen Datu-Base Lexikala (henceforth EDBL)) (Aldezabal et al., 2001a). Thus, each item is classified in accordance with its lexical or morphosyntactic category and subcategory. The database is organized to carry out the so-called morphotactic relation (Alegria 1995, Urkia 1997), along the lines of the two-level morphology in Koskenniemi (1983). This means that the combinations between morphemes are included in the database itself. This provides as a result the morphological and morphosyntactic composition of words.
b) In order to reduce ambiguity, we have employed a disambiguating tool for Basque (Aduriz et al. 1997) that was created based on the Constraint Grammar (henceforth CG) formalism in Karlsson et al. (1995). This tool reduces the possible interpretations of words through definitions of rules that are based on

context. This disambiguating tool cuts categorical ambiguity almost entirely. However, ambiguity persists in cases where other factors such as Case or function are considered, which suggests that further information is necessary. One such type of information is verbal subcategorization, namely, specification of elements that are selected by verbs.

c) Yet, computational research has continued into syntax in two directions despite the persistence of ambiguity, but acknowledging the necessity for lexical information. One line of research has created a finite state system by extending the CG formalism (Tapanainen 1996); another line has created the PATR II formalism based on unification (Shieber 1986). The former creates new tags to form phrases based on the function of morphemes. This provides as a result a syntactically tagged sentence (Aduriz 2000, Arriola 2000). The later defines the unification-rules by employing the lexical information of morphemes. These rules meet the relations existing between the word level and phrase level by using the unification-equations (Gojenola 2000, Aldezabal et al. 2003).

d) However, the results obtained by the application of these formalisms are not very successful considering the following facts: first, some interpretations remain ambiguous in the morphological disambiguation process, and second, grammars suggest many combinations among the elements of the sentence, i.e., they create structural ambiguity. In order to minimize this problem, we have applied a Finite State technique based on automata and transducers. As a result, since the verbal context under consideration is reduced, ambiguity percentages are also significantly reduced (Aldezabal et al. 1999b, Aldezabal et al. 2001b). Thus, we are able to get a phrasal analysis of sentences in a corpus, and to use all the morphosyntactic information included in the phrase. In addition, we will often find that we get several interpretations for one sentence.

Let us consider an example of how this process is applied to a particular sentence (excluding ambiguity).

The above results show that there is no relation between the elements surrounding the verb; in other words, we assume that, in principle, the elements surrounding the verb somehow belong to the verb; there is no explicit distinction as to whether elements belong to the verb (the arguments of a verb) or to the sentence (adjuncts).

Things get even more complicated when sentences contain more than one verb, since, in principle the surrounding elements may be related to either predicate. In such cases, apart from not showing the argument/adjunct distinction stated above, there is no way of knowing to which verb phrases relate. This increases ambiguity, since all combinations are considered as legitimate options. Moreover, another arising problem is that clause boundaries within sentences cannot be delimited.

For all these reasons, at this point it is clear that, as is the case in theoretical linguistics, the computational treatment of language requires considering verbal subcategorization.[1] Thus, in this article we will show how subcategorization can be

[1] However, we need to mention that we have taken important steps in retrieving information pertaining to subcategorization by automatic or semi-automatic means (Arriola 2000; Aldezabal et al., 2001b, Aldezabal et al., 2001c).

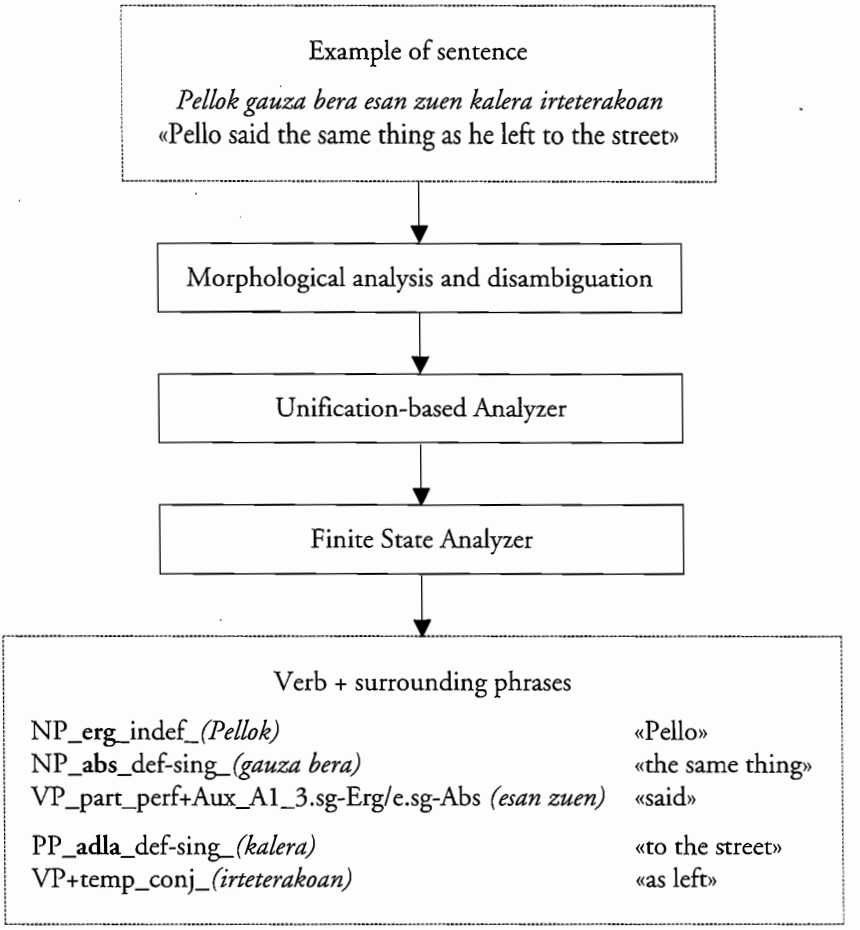


Figure 1. The general picture of the system with an example

analyzed in relation to the perspective of computational linguistics within the IXA group. The presentation of this research is organized as follows. Section 2 includes a presentation of the concept of subcategorization as well as a brief description of the proposals concerning the organization of the lexicon. In section 3, we present an overview of the procedure that we have selected for our work. First, and considering the above stated facts, we will take as a point of departure a proposal that bridges best the theoretical and computational approaches, namely the *English verb classes and alternations* by Levin (1993). Specifically, we will show the viewpoint and methodology included in Levin's work, as well as the gaps that we have detected in them. Next, we will explain the choice we have made for our work. Section 4 includes the overall conclusions drawn from the application of our procedure, the problems we encountered in doing so, the decisions we have made, and the specification of the pheno-

mena that were detected. Finally, section 5 is a summary of the article, and it includes the general conclusion drawn from the research.

2. On subcategorization

So far, in this article, we have suggested that the information in subcategorization pertains to the lexicon. However, in the literature we find various features and expressions that describe and designate this term. The term subcategorization arises parallel to the discussion on the autonomy of the lexicon within generative syntax,[2] which started when Chomsky published his second book *Aspects of the Theory of Syntax* (Chomsky 1965). In Chomsky (1965), the lexicon will become increasingly independent; lexical items include phonological and categorial information, and in the case of verbs, apart from phonological and categorial information, we will find information on subcategorization, selectional restrictions on arguments, and features pertaining to context. Subcategorization information includes the phrasal category (NP, TP, etc.) of the elements that are required by the verb, in other words, the specification of the syntactic realization of arguments. This was precisely what was considered to be in the so-called strict subcategorization.

Additionally, verbs were classified according to one of the two subcategorization structures that were suggested. On the one hand, predicates that had the subcategorization structure 'NP + V + NP' were transitive predicates, and those that displayed the structure 'NP + V' were classified as intransitives. In other words, when predicates contained an object they were called transitive predicates, and otherwise intransitives. Syntactic rules that made up sentences were defined in terms of this parameter. Yet, this first attempt in generative grammar was considered both redundant and too dependent on certain languages. Additionally, contrary to the above prediction, they realized on the existence of predicates that included a transitive auxiliary and a subject but no object (*irakin* ('to boil'), *iraun* ('to last'), *dimititu* ('to resign') and the like). This suggested that the terms transitive and intransitive were not clearly defined. As a solution, Chomsky in his *Lectures on Government and Binding* (Chomsky 1981) presented the influencing framework called Government and Binding Theory (henceforth GB). In this framework, grammars are viewed as computational systems composed of modules that include some universal principles and some parametric variations.

According to this proposal, predicates have the ability to assign a semantic feature called thematic role to each of its arguments (namely, to each of the participants that are necessarily involved in the action denoted by the verb). Additionally, verbs are capable of assigning the Case that will allow the realization of thematic roles in the syntax (the Case Filter). Moreover, thematic roles are hierarchically organized, which defines the function that arguments have in sentences. It is further assumed that thematic roles are invariably realized in specific phrasal categories. Thus, by the principle of Canonical Structural Realization (CSR), each thematic role is assigned the

[2] Before this date, we find the term *government*, which expresses the task of selection of pre/postpositions by the verb, which is a similar concept to current analyses involving verbal selection. Yet, the term subcategorization arises with Chomsky, and we have set our research after the term was suggested.

corresponding grammatical category. Hence, each predicate contains an argument structure in the lexicon, and the hierarchy and the CSR will determine the role and the syntactic realization of arguments.[3]

The new classification of verbs includes the following: unaccusative predicates (those that involve a purely intransitive auxiliary and a single argument, as in *etorri* 'to come'), unergative predicates (which involve a transitive auxiliary and a single argument, as in *iraun* ('to last')), and finally, transitive predicates (which involve a transitive auxiliary and two arguments, as in *eraman* ('to take')). Several theories have arisen attempting to explain single argument predicates.

Some authors started to claim that the structure of the lexicon is more complex than was standardly assumed, and they defended the existence of regularities in it. This attracted the attention of researchers towards the lexicon. It was claimed that such regularities arose from the interaction between semantics and syntax. The first to claim such a relation were Hale and Keyser (1987). Later, the proposal in Jackendoff (1990) has been the most successful one and the one to receive most attention. Jackendoff suggested a more abstract structure to represent the lexicon, namely the Lexico-Conceptual Structure (LCS). This structure is composed of various semantic primitives (among others, GO, STAY, CAUSE, TO, FROM, TOWARD, AWAY-FROM, VIA), an in turn, these primitives correspond to more general conceptual categories (Thing —or Object—, Event, State, Action, Place, Path, Property and Amount). For instance, primitives GO, STAY and CAUSE correspond to the conceptual category Event. In addition, the syntactic correspondence is defined also at this level. Thus, each lexical entry is defined in terms of the conceptual categories, primitives and their corresponding surface syntactic structures.

Other proposals were also suggested. For instance, the generative lexicon in Pustejovsky (1995), who suggests a complex structure for each lexical item (which includes argument structure, event-structure, and qualia structure), and obtains the surface structure through composition of all features that take part in the complex structure; Levin in her *English verbs classes and alternations* (1993) analyzes English verbs. Levin does not specify the entry corresponding to each lexical item (as she herself acknowledges). Rather, she suggests ways of organizing the entries. She notices that verbs that are similar in semantic nature accept the same syntactic structures. Thus, the fact that the ability of language is considered to be innate explains how speakers are capable of knowing what syntactic structures are allowed with predicates. This suggests that the first task is to figure out which syntactic structures we are facing in order to group predicates and to analyze their semantic components. It is clear that the internal composition of lexical items is still being debated and analyzed. Yet, there is a commonality underlying all the theories proposed: lexical items contain various types of features, and the existing relations of such features condition the correct syntactic realization of lexical items.

---

[3] Two clarifications are necessary at this point. First, there are those who support the view that it is necessary to define subcategorization (Grimshaw 1979; Rothstein (1992): among others). Second, the discussion on thematic roles is far from reaching consensus. Some suggest that roles may be distinguished in contrastive pairs (for example, the pairs *agent/cause* and *goal/receiver* through the feature [±animate]). Others have suggested other proposals, among others, we find Dowty (1991) and Van Valin (1993), who compose roles by means of role hierarchies called 'protoroles' —or general roles— and by binary +/- features.

3. The procedure

As for the Basque lexicon, and more specifically, as for the verb, the EDBL defines the category, the subcategory and the word combinatorial options that are accepted within the verb (namely the morphotactic relation). Thus, it is clear that we are far from the complex composition of the lexicon proposed in the previous section, and that there is no reference to the components that are selected. This implies that, if in the near future we engage in completing the lexicon of the verbs for their application in automatic use, we will need to start by positing modest goals. Thus, although the ideal facts about the lexicon may be contained in theoretical proposals, practicality restricts our goals. To start with, our interest is to determine the surface realization of the components at the level of the sentence. This suggests a clear approximation to strict subcategorization.

Second, we need to take into account that the steps that we have taken so far in our group provide us with interesting available information, which includes phrases that compose sentences, including all the information contained in them. This will let us proceed to further analyses or to confirmation-processes.

All this suggests taking into account the work developed by Levin. In our view, the line proposed by Levin is roughly adequate, mainly for two reasons: first, from a computational linguistics perspective, because it engages in analyzing surface structures. Second, because it is aimed at organizing the lexicon of verbs. Thus, we have analyzed her proposal in detail and we have measured the advantages and disadvantages that it offers. In addition, Levin's work has served to analyze verbs in various languages such as Spanish and Catalan (Taulé 1995), French (Saint Dizier 1995), German, Korean and Bangla (Jones et al. 1994). The research on Spanish and Catalan deserves special mention. Because of the cooperative relation that we maintain with them, we have had the chance to get to know their work in detail; moreover, we hope that their experience will serve to guide us in our research (Vázquez et al. 2000).

3.1. Levin as point of departure

Levin claims that native speakers are capable of noticing many phenomena that appear in their language. One of them is the ability to notice among the various syntactic realizations of a particular verb. In other words, speakers are able to establish relations among the various structures —some of which imply semantic differences— that verbs display. They are also able to determine which structure(s) each predicate may accept, and which not. Levin employs the term diathesis alternations to name the different structures or, in other words, to name the pairs of structures of verbs that are related. Quoting:

> Verbs, as argument-taking elements, show especially complex sets of properties. As shown in B. Levin (1985b, in prep.) and other works, native speakers can make extremely subtle judgments concerning the occurrence of verbs with a range of possible combinations of arguments and adjuncts in various syntactic expressions. For instance, speakers of English know which *diathesis alternations* —alternations in the expressions of arguments, sometimes accompanied by changes of meaning— verbs may participate (Levin 1993: 2).

According to her, there is at least one common semantic feature in the syntactic variants of the alternations that verbs admit. This is precisely the reason why it is possible to classify verbs into groups:

> If the distinctive behavior of verb classes with respect to the diathesis alternations arises from their meaning, any class of verbs whose members pattern together with respect to diathesis alternations should be a semantically coherent class: its members should share at least some aspect of meaning (Levin 1993: 2).

Thus, after explaining the theory based on Lexical Knowledge in depth, Levin divides the content of her results into two parts: in the first part, she shows the alternations that she found in English, she provides the list of the verbs that take those alternations, and for each alternation, she describes their syntactic, semantic and (when applicable) morphological features. In total, she presents 80 alternations, and she divides them into 8 groups, which are, in turn, divided into further subgroups. In the second part of her work, and based on these alternations, she suggests 191 semantic subgroups in total, which are organized into 49 larger sections. Yet, we have detected several incoherencies in her procedure of analyzing alternations and grouping verbs. Here is the list of the incoherencies that we found:

— She does not always group verbs according to the alternations that verbs share.

- For instance, verbs of groups 9.1 *(Put verbs)* and 10.1 *(Remove verbs)* admit the same alternations, and yet, she classifies them into distinct groups.
- Another occasion when she turns verbs into distinct subgroups, is when they contain a semantic component introduced into the verb via suffixation. For example, this is the case of verbs in group 9 *(Verbs of Putting),* namely subgroup 9.9 *(Butter verbs)* and 9.10 *(Pocket verbs).* E.g.:

  9.9: *Lora buttered the toast*
  9.10: *Lydia pocketed the change*

  It is obvious that the basic structure of these derived verbs and that of the non-derived form (namely, the remaining subgroups in section 9) is different, and that syntactic structures are unable to relate the derived and non-derived forms. However, verbs of groups 9.9 and 9.10 admit/reject the same alternations, and thus, they should not be considered as 'syntactically' distinct; however, they are distinguished in Levin's system.
- Certain semantic groups do not display any alternations (for instance, group 52 *Avoid verbs*, and the subgroup 54.2 *Cost verbs* in section 54). This, according to Levin's methodology, would imply that verbs that display such structures do not accept any alternations, and hence, we would have to conclude that they do not form any group.

— She uses the term alternation in various senses. As was mentioned above, Levin considers alternations the pair of structures that certain verbs admit and that share certain common semantic property. Nevertheless, this is not always so. In fact, there are several alternations where only one structure is described (namely

those in 7.4.,[4] 7.5 and 8.4), and others that admit two structures, but where one of them is illegitimate (for instance 7.6.1, 7.6.2, 7.7, 8.1, 8.2, 8.3, 8.5 and 8.6). Moreover, we also find differences in those alternations that admit two legitimate structures: sometimes, one syntactic component drops in one structure; others, one component is added, and finally, sometimes, there is no component that is dropped, but the syntactic realization of such components changes.

— To finish up, for each semantic group, she does not specify the source of the structure that is considered as basic within alternations. It seems that the basic structure is already delimited (or it looks that she considers it to be so), and that based on this, she then lists the various alternations which are accepted in each case. Thus, there seems to be a gap in the methodology or theory that she proposes.

3.2. Our choice

Considering the problems in the previous section, rather than taking into account the semantic groups that she suggests for English, we decided to analyze 100 verbs in Basque by employing certain syntactic resources and by making use of the Corpus[5] that is available to us. When specifying our resources, we have taken into account how useful the selected resources may be for our computational tools. On the one hand, it is from these resources that we will retrieve useful data for our manual analyses, and on the other, those resources constitute the onset for our future research. However, we found it interesting to consider the alternations that we may find in Basque compared to those found in Levin (1993). The fact that we may find parallel alternations in Basque and English provides generality to the structures, and moreover, it may be relevant from a comparative perspective. Thus, we have considered the research that was developed within the IXA group (Aldezabal et al. 2002) as a basis, which includes a comparison of the alternations proposed by Levin with Basque.

As for the computational analyses in our research group, we have mentioned that the current computational tools analyze the phrases in sentences that appear in a corpus. These analyses provide as a result morphosyntactic information of phrases —namely information on number, definiteness and Case—. Additionally, our tools can easily provide us with the correct auxiliary that corresponds to the verb in each instance. These are ample resources that are available in the research. Next, we will describe the details of our line of research.

3.2.1. *Features considered in verbal analyses*

In order to complete the information pertaining to verbs, we have made use of two particular surface syntactic features when analising the sentences in the corpus.

— The type of auxiliary, by using the following typical means of expressing types of verbs: DA (purely intransitive), DU (transitive), DIO (ditransitive) and ZAIO (involving two arguments, one in dative and one in absolutive).

---

[4] The numbers in the alternations in the text strictly follow the ones in Levin (1993).

[5] The available corpora refer to the electronic samples of the daily *Euskaldunon Egunkaria* between January 1999 and May 2000.

— Case: we determine which cases verbs accept. However, we have only considered those Case markings that display a meaningful degree of presence in the corpus, specifically, only eight:

absolutive (41,79%), completive *-ela* (2,70%),
ergative (36,37%), instrumental (1,77%),
inessive (6,38%), sociative (1,52%),
dative (3,61%), ablative (1,34%)
adlative (1,28%).[6]

The remaining Cases have a percentage of presence lower than 1. This is the way we have analyzed Cases:

— On the one hand, Cases that outstand in frequency, namely those that are semantically closely related to the verb (or more specifically, those that we consider to be related to the verb), will be marked exceptionally. Thus, we will call these Cases 'outstanding Cases'.

— On the other hand, and in order to help distinguish between alternations and non-alternations pertaining to verbs, we have attempted to consider the constraints on the simultaneous appearances of Cases. In other words, we have analyzed Cases in terms of the restrictions that they impose on the realization of other Cases.

3.2.2. *Verbal values: syntactic/semantic values (ssv)*

The features described in the previous section will be assigned based on the different values that correspond to verbs. This is, indeed, the most complicated task. As it was mentioned above, the theory proposed by Levin suggests that, by virtue of their innate ability, speakers are able to determine the existing (and non-exiting) alternations pertaining to a verb. The underlying idea is that alternations share some semantic component. Hence, the crucial task is to determine which is/are the component(s) that alternations share. In fact, the semantic nature of such components (their semantic relation with verbs) determines how outstanding Cases are. Thus, our goal has been to determine those semantic components by analyzing 100 verbs in depth, and moreover, we have intended to identify the syntactic structures that are involved in alternations, i.e., to identify alternations, and those which are not. Thus, we have described several values for each verb, which are specified by their meaningful semantic components and by their syntactic Case realization. As a result, we have considered the values of verbs as semantic/syntactic values (*ssv*).[7]

---

[6] Note two important facts regarding Case: first, we have employed the term Case in a very general sense by including all declension Cases, both simple and complex (the later involving various words) as well as subordinating conjunctions (also simple and complex ones); second, works involving automatic retrieval of information regarding subcategorization consider all partitives as absolutives. This is the reason why, in manual analyses, we do not distinguish between appearances of partitive and absolutive.

[7] Note that the *ssv*-s that we have defined do not necessarily correspond to the verbal entries that are defined in dictionaries.

In addition, we have not distinguished between the two variants that belong to the alternation(s) of a verb (namely, the *ssv*-s that are related by some semantic component), nor the *ssv*-s that are not related to each other (namely, those that do not take part in alternations). Thus, various *ssv*-s are suitable for each verb (regardless of the existence of alternations among them). Note that we have not described *ssv*-s that are not in the corpus, although we acknowledge that there may be some.

We have designed a database in order to keep the information pertaining to *ssv*-s in a structures manner, and we have selected a marking-system to codify the information. Nevertheless, we have disregarded several topics to avoid the analysis from becoming too complex. For present purposes, we will only present the basics of this subject, and for more information, see Aldezabal et al. 2001 (forthcoming). First, we will present the topics that we have excluded from the research, and next, we will describe the marking-system that we have employed.

3.2.3. *Excluded topics*

3.2.3.1. *Impersonal, passive and antipassives*

Along the lines of Levin, in the task of marking different *ssv-s,* we have tried to solely resort to lexical values. From this perspective, it is well known that impersonal, passive and antipassive constructions are structures that are derived in the sense that they emerge as a result of applying some lexical operation to lexical structures. We have accepted this claim, and thus, when we have come across verbs that involve such constructions, we have not marked them as distinct in terms of *ssv*-s. Thus, when verbs appear in such constructions in the corpus, we have merely marked them as involving the values that they would have in non-impersonal and active sentences.

3.2.3.2. *Phrases without case*

Certain phrases are not formed by Case. These are adverbial phrases. We have not considered them because they do not display any Case.[8]

3.2.3.3. *The same case only once in each:* ssv, *except absolutive*

In the *ssv*-s, we will not mark the same Case more than once. If necessary, the Case will specify the possible semantic values that we have determined for the *ssv* in each instance. In other words, rather than distinguishing Cases we will distinguish semantic values. For example, we will not mark the two well-attested values of the ablative (source and path —or prosecutive value, Azkarate and Altuna 2001: 128) with two ablative markings, but rather, we will consider them as two legitimate values of the ablative in the same *ssv.* Nevertheless, we will make an exception; specifically we will accept two absolutives in the same *ssv.* Arguably, adjectives and nouns can form nominal predications that are formed with the absolutive (mostly

[8] This implies that certain legitimate values of verbs will be left out. Notice that adverbial phrases are sometimes necessary in the *ssv* definition of a verb.

with indefinite absolutive forms).[9] Yet, we will consider them as if they were first level nominal predicates, namely, only when the component in the *ssv* is most relevant.

3.2.3.4. *Lexicalized units*

The fact that we have considered Case does not imply that we have considered every phrase that contains some Case. It is well known that many of the Cases that we have selected display tendencies for lexicalization when they appear attached to other lemmas, either in the form of single words or in the shape of various forms. (e.g., *orduan* ('then'), *sekulan* ('ever'), *patxadan* ('relaxed'), *marmarrean* ('muttering'), *azken batean* ('after all'), *hitz batean* ('in a nutshell'), *gogotik* ('willingly'), *horratik* ('nevertheless'), *aspalditik* ('for a long time'), *inondik ere* ('absolutely not'), *gora* ('upwards'), *ahoz behera* ('face down'), *hankaz gora* ('upside down'), *adibidez* ('for example'), *negarrez* ('crying'), *beldurrez* ('in fear of'), etc.).

They also participate in various compounds (e.g. ate*z* ate ('from door to door'), mendi*z* mendi ('form mountain to mountain'), etc.), and in complex declension Cases (e.g. -i buru*z* ('about something'), -tik at ('out of'), -n zehar ('through'), etc.). It may also be part of units that are composed of several phrases (e.g. *bostetik bi* ('two out of five'), *zazpitik lau* ('four out of seven'), *lurretik bost metrora* (five meters form the ground'), *egunetik egunera* ('day by day'), *goitik behera* ('thoroughly').

However, we need to mention that it is not easy to decide on the degree of lexicalization of such items. In fact, in our view, the fact that many such forms are in the process of lexicalization is related to the growing loss of the values that Cases have with respect to verbs in general. For example, based on what we have seen in our analyses, ablative Case involves values related to departure location, path and static setting of the entity, and adlative Case involves values related to goal. However, occasionally, ablative and adlatives receive other values too. For instance 'manner': *gogotik* ('willingly'); *hautura* ('at someone's discretion'). When this phenomenon happens lexicalization appears. Regarding units composed by more that one phrase, one of the reasons for considering them as units is that phrases in isolation do not make sense with respect to a particular *ssv* of a verb. In other words, what gets the value is the element resulting from the union of two phrases in the *ssv*. For example, in *Goitik behera busti zuten* ('they soaked him all over'), the unit shows 'manner'; in *Leihotik behera bota zuten* ('they threw him/her out of the window') the unit refers to the direction (and not to the departure and target points). In our view, this is precisely the reason why these strings should be considered as a complex declension Cases.[10] Thus, all the forms described above should be considered as lexicalized forms or units, and we

---

[9] Zabala analyzes predication relations in depth in her 1993 thesis entitled *Predikazioaren teoriak Gramatika Sortzailean (Euskararen kasua)*, where she includes several proposals for the elements that realize such predication relations.

[10] Let us mention that we have taken steps in analyzing units that contain various words (what we call Multi-word Lexical Units (MWLU)) (Aduriz et al., 1996). Moreover, there is current doctoral research on this topic in our group (Urizar, R.: *Kolokazioak euskaraz*). In addition, some research has been done in analyzing structures that contain various phrases from a semantic and pragmatic perspective (Garai & Ibarretxe 2002).

should analyze the syntactic and semantic values that they take as a whole. Since, automatically we have only analyzed units as postpositions (and since these forms are not among postpositions), we have decided not to analyze them.

3.2.3.5. *Cases that may have temporal reference*

It is well known that verbs may usually take phrases that contain temporal reference, and that temporal reference may be expressed by various types of Cases, such as inessives, ablatives, adlatives, instrumentals, sociatives, and also, absolutives —of course, only if one considers such forms as absolutives— (*gauean* ('at night'), *igandetik* ('since Monday'), *igandera* ('til Monday'), *arratsaldez* ('in the afternoon'), *igandearekin* ('with Sunday'), *bi egun* ('two days'), etc.). We know that, apart form setting the action denoted by the verb in temporal reference, these temporal references do not usually provide special information about the verb, and that most verbs accept such Cases.

Thus, when marking Cases, we have decided not to consider instances that contain temporal reference.

3.2.4. *The database and the marking-system*

3.2.4.1. *The database*[11]

The database contains five charts. There is one main chart, where we mark the type of auxiliary that corresponds to each *ssv* of the verb. Each of the remaining four charts corresponds to types of auxiliaries, and they contain a specification of Cases that will be analyzed in each chart. A small square beside the Case signals whether the case is accepted or not, and the Cases that we have determined as outstanding contain an additional domain that specifies their semantic value. The charts that correspond to auxiliaries have room for explanations, examples, and comments. Thus, after marking the type of auxiliary in the main chart, we fill the chart that corresponds to the auxiliary that we have marked.

3.2.4.2. *The marking-system*

We have employed three specific symbols in the marking-system, namely √, – and +. We have marked '√' the auxiliary and the outstanding Cases that are used in each *ssv*. Concerning cases, this symbol signals the following: 'it may appear, and it is outstanding'. In other terms, regardless of its presence/absence in the corpus, we consider that the Case has the ability to surface in the *ssv* under consideration, and it is typically outstanding. Assigning '√' to auxiliaries means that the verb under consideration takes the auxiliary in that *ssv*, although it may not appear conjugated. We employ the symbol '–' to express that a Case is unacceptable in a combination. Finally, we may find that, although a given Case is accepted, it is not closely related to the verb, namely, it is not an outstanding Case. Such Cases are marked with symbol +.

---

[11] For the moment being, the content and shape of this database is not available to the public. However, we are planning to include it in our webpage so that anyone can consult it.

Concerning Cases, it is well known that absolutive, ergative and datives are exceptional in displaying agreement with the auxiliary. As such, symbols '√', '–', and '+' on them contain a more specific meaning:

Absolutive and Ergative

Symbol '√' on absolutive and ergatives automatically implies that the verb requires them, but that the Cases may be absent due to ellipsis. In other words, these Cases 'may' appear, as is well known: when they are absent, it means that they are absent for elliptical reasons (the phenomenon of pro-drop). In contrast, instances where these Cases are absent for other reasons will be marked with '–' (namely as a distinct *ssv*) to signal they must be absent.

As for auxiliaries, we will mark the auxiliary type that the verb takes in the SSV.

Dative

We will mark the dative with '√':

— If a verb accepts the dative, where the dative is not a mere addition in instances that involve no dative. E.g.:

Pello adiskideen izenez ***ahaztu*** *da* → **Pello adiskideen izenez* ***ahaztu*** *zaizkio Anderri*

(Lit: Pello friends-of names-post forget is → *Pello friends-of names-post forget Aux(ABS-**DAT**) Ander-DAT)
Meaning: 'Pello forgot the name of his friends' → *'Pello forgot the names of his friends to Ander'.

(Correct structure: *Pellori adiskideen izenak* ***ahaztu*** *zaizkio,*

Lit.: Pello-DAT friends-post. names forget Aux (ABS-DAT)
Meaning: 'Pello forgot the names of his friends.'

— When solely the dative is accepted. E.g.:

Ekin *genion lasterrari* (e.g. from Sarasola 1996) → ****ekin*** *genuen lasterra*

(Lit.: Start Aux (ERG-DAT) run-DAT → *start Aux (ERG-ABS))
Meaning: 'We engaged in the task of running (i.e., we started running)'

— Finally, where the dative is mere addition, but appears very frequently. E.g.:

Lehen saria ***eman*** *zioten* (from Sarasola 1996)

Lit.: first prize give (ABS-DAT-ERG)
Meaning: 'They gave him/her the first prize.'

For these later instances, we will check whether the dative is very frequent in the corpus, and if so, we will mark it as outstanding. Where the dative is a mere addition

and is not frequent will be marked '+'. Of course, when the dative is not accepted we will mark it with '–'. Marking the dative does not imply that it will be reflected in the Auxiliary. Specifically, although the dative is marked with '+' —namely, when it is a mere attachment that is not frequent—, the auxiliary will be marked as either DA (ABS) or DU (ABS-ERG) (we will do the same, or course, when the dative is not accepted). Otherwise, the auxiliary will be marked as ZAIO (ABS-DAT) or DIO (ABS-DAT-ERG).

To summarize, these are the marking options that arise in the auxiliaries and the agreements.

	ABS	ERG	DAT
DA	√/–	–	√/+/–
ZAIO	√/–	–	√
DU	√/–	√/–	√/+/–
DIO	√/–	√	√

This means that, only instances that involve dropping of ergative and absolutive Cases will be considered as variants of an alternation, i.e., as separate *ssv-s* (the remaining Cases have the 'may appear' value signaled by '√').

3.2.5. *Alternations attested in both English and Basque*

As it was mentioned above, Aldezabal et al. (2002) analyze which alternations that have been proposed for English appear in Basque and which are absent. For present purposes, and without entering into details, among the ones that are accepted in Basque, we have selected instances that involve the Cases which were mentioned above as well as those forms that we have considered as lexical. Below is the list of the attested alternations illustrated by examples in English and Basque. The types of alternations are numbered according to the numbers in Levin's work. All these alternations have been marked according to the marking-system that we have suggested above. Since Levin considers the components that take part in the alternations as arguments (and she explicitly signals the ones that are not), we have marked the Case that such components show with '√'.

Here is the list:

Causative/Inchoative alternation; Levin's 1.1.2.1.

Eng. Janet broke the cup/The cup broke

Basq. *Janetek katilua puskatu zuen/Katilua puskatu egin zen*

Substance/Source alternation; Levin's 1.1.3.

Eng. Heat radiates from the sun/The sun radiates heat

Basq. *Beroa eguzkitik irradiatzen da/Eguzkiak beroa irradiatzen du*

Unspecified Object alternation; Levin's 1.2.1.
Eng. Mike ate the cake/Mike ate
Basq. *Mikek opila jan zuen/Mikek jan zuen*
Understood Reciprocal Object alternation; Levin's 1.2.4.
Eng. Anne met Cathy/Anne and Cathy met
Basq. *Annek Cathy topatu zuen/Anne eta Cathy topatu ziren*
Characteristic Property of Agent alternation; Levin's 1.2.6.1.
Eng. That dog bites people/That dog bites
Basq. *Zakur horrek jendeari hozka egiten dio/Zakur horrek hozka egiten du*
Characteristic Property of Instrument alternation; Levin's 1.2.6.2.
Eng. This knife cut the bread/This knife doesn't cut
Basq. *?Labana honek ogia mozten du/Labana honek ez du mozten*
Conative alternations; Levin's 1.3.
Eng. Paula hit at the fence/Paula hit the fence
Basq. *Paulak hesian/-ren kontra jo zuen/Paulak hesia jo zuen*
Locative Preposition drop alternation; Levin's 1.4.1.
Eng. Martha climbed up the mountain/Martha climbed the mountain
Basq. *Paula mendira igo zen/Paulak mendia igo zuen*
With preposition drop alternation; Levin's 1.4.2.
Eng. Jill met with Sarah/Jill met Sarah
Basq. *Jill Sarahekin topatu zen/Jillek Sarah topatu zuen*
Spray/load alternation; Levin's 2.3.1.
Eng. Jack sprayed paint on the wall/Jack sprayed the wall of paint
Basq. **Jackek horman pintura ihinztatu zuen/Jackek horma pinturaz ihinztatu zuen*
Simple Reciprocal alternation (Transitive); Levin's 2.5.1.
Eng. I separated the yolk from the white/I separated the yolk and the white
Basq. *Gorringoa zuringotik bereizi nuen/Gorringoa eta zuringoa bereizi nituen*
Simple Reciprocal alternation (Intransitive);[12] Levin's 2.5.4.
Eng. The oil separated from the vinegar/The oil and vinegar separated
Basq. *Olioa ozpinetik banandu zen/Olioa eta ozpina banandu egin ziren*
Body-Part possessor Ascension alternation; Levin's 2.12.
Eng. Selina touched the horse on the back/Selina touched the horse's back
Basq. (Lit.) *Selinak zaldia ukitu zuen bizkarrean*; (Meaning) *Selinak zaldiari bizkarra ukitu zion/Selinak zaldiaren bizkarra ukitu zuen*
Possessor object; Levin's 2.5.5.
Eng. I admired his courage/I admired him for his courage
Basq. *Bere kemena miresten nuen/Bere kemenagatik miresten nuen*

[12] We need to mention that we are unable to distinguish some alternations according to our marking-system, and hence, we have not marked them as distinct *ssv*-s. This applies to Simple Reciprocal alternation transitive 'and 'Simple Reciprocal alternation intransitive'. Thus, we have listed them as accepted alternations, but keeping in mind that one variant of the alternation will not be considered as a separate *ssv*.

Attribute Object; Levin's 2.13.1.

Eng. I admired his honesty/I admired the honesty in him

Basq. *Bere zintzotasuna miresten nuen/Beregan zintzotasuna miresten nuen*

Possessor and Attribute alternation; Levin's 2.13.3.

Eng. I admired him for his honesty/I admired the honesty in him

Basq. *Bere zintzotasunagatik miresten nuen/Beregan zintzotasuna miresten nuen/Bere zintzotasuna miresten nuen*

Possessor subject (transitive); Levin's 2.13.4.

Eng. The clown amused the children with his antics/The clown's antics amused the children

Basq. *Pailazoak bere bihurrikeriekin haurrak entretenitu zituen/Pailazoaren bihurrikeriek haurrak entretenitu zituzten*

Time Subject alternation; Levin's 3.1.

Eng. The world saw the beginning of a new era in 1492/1492 saw the beginning of a new era

Basq. *Munduak aro berri baten hasiera ikusi zuen 1492an/1492k aro berri baten hasiera ikusi zuen*

Abstract Cause Subject alternation; Levin's 3.4.

Eng. He established his innocence with the letter/The letter established his innocence

Basq. *Bere inozentzia gutunaren bidez frogatu zuen/Gutunak bere inozentzia frogatu zuen*

Cognate Object construction; Levin's 7.1.

Eng. Sarah sang/Sarah sang a ballad/Sarah sang a song

Basq. *Sarah-k abestu egin zuen/Sarah-k balada bat abestu zuen/Sarah-k abesti bat abestu zuen*

3.2.6. *Selecting verbs*

The first task in analyzing verbs involves a selection of a set of verbs. For this purpose, we have made use of the Statistical Corpus of the XX. Century (i.e., *XX. mendeko euskararen corpus estatistikoa*). After selecting a sample of 22.000 words from the corpus, we have listed verbs according to degree of frequency in which they appear (overall 622 verbs), and, from this list, we have finally selected 100 verbs. We first present the criteria that we have followed for excluding verbs.

Excluding verbs that involve a clear derivational process

The list of selected verbs includes no verb involving clear and productive derivational processes. In section 3.1 of this article, where we described the proposal by Levin, we have argued that there are syntactic structural differences between a derived verb and its non-derived counterpart, where both contain parallel semantics. It is clear that they are syntactically distinct, and hence, along the lines of Levin's methodology, they are not syntactically comparable. We also mentioned that, in our view, Levin is not consistent in using her own methodology (among others, in cases where derivations are involved). However, this does not imply that we have initially discarded her methodology. Thus, we have excluded verbs that involve derivational processes, albeit

acknowledging the systematic process in them. We have preferred to analyze the general structure of verbs that involve no derivational process, and we leave the analysis of derived verbs based on general structures for future research. In fact, although these derivational processes are systematic, we believe that there is underlying complexity in the system (for instance, considering predicates such as *sartu* ('to put in') and *poltsikoratu* ('to pocket'), predicate *sartu* accepts the ablative Case —specially when expressing path— in addition to the adlative; in contrast, predicate *poltsikoratu* hardly accepts the ablative case). Thus, these are topics that require deeper research.[13]

Moreover, derivational processes are not sometimes very explicit; often, it is difficult to detect the components that take part in the composition of the verb, probably, because their birth is long back in history. For this reason, we have decided to exclude the following from our research. On the one hand, the clear and systematic derivational cases that we found in Basque in analyzing the verbal classes suggested by Levin, namely the forms composed of the following morphemes: *-etsi* (as in *onetsi* ('to accept'), *handietsi* ('to praise'), *-ztatu* (as in *ureztatu* ('to water'), *irineztatu* ('to flour'), *-ratu* (as in *poltsikoratu* ('to pocket'), *botilaratu* ('to bottle'), *-katu* (as in *mailukatu* ('to nail'), *kolpekatu* ('to hit'), and *-gabetu* (as in *hezurgabetu* ('to unbone'), *gazgabetu* ('to unsalt'). On the other hand, we have left out most of the derived semantic values that are attributed to suffix *-tu* (some of them also attested in the above analysis) in Gràcia et al. (2000). Specifically, these authors propose 6 interpretations for this suffix:

— Change in state/quality (-tu1, -tu2, -tu3, -tu8, -tu9): *gizondu* ('to become a man'), *izoztu* ('to ice'), *beldurtu* ('to (be) *frighten(ed)*'), *lotsatu* ('to (be) *embarrass* (ed)'), *zatitu* ('to divide'), *puskatu* ('to break'), *lasaitu* ('to calm'), *garbitu* ('to clean'), *mailakatu* ('to classify'), *lerrokatu* ('to align'), etc.
— Removal (-tu4): *larrutu* ('to skin'), *lumatu* ('to pluck feathers')
— Transmission (-tu7): *babestu* ('to protect'), *zigortu* ('to punish'), *aholkatu* ('to give advice'), etc.
— Change of Location (-tu6, -tu11, -tu12): *baztertu* ('to put aside'), *saihestu* ('to move sideways'), *alboratu* ('to approach'), *kaiolaratu* ('to cage'), *beruneztatu* ('to cover with lead'), *ureztatu* ('to water'), etc.
— Repetition (with some instrument) (-tu10): *mailukatu* ('to nail'), *mokokatu* ('to peck'), etc.
— Location (involving realization of the locus) (-tu5): *lumatu* ('to grow feather'), *hostatu* ('to become covered by leaves'), *loratu* ('to flower')

For our purposes, we have decided to only taken the first values into account.

Excluding verbs that are composed of more than one component

In our process of selection, we have excluded verbs that contain more than one component (e.g., *lo egin* ('to sleep'), *zain egon* ('to wait'), *axola izan* ('to matter'), *ari izan* ('be doing'), *barre egin* ('to laugh'), *bat egin* ('to unite'), *gogora ekarri* '(to remind'), *merezi izan* ('to be worth', etc.). In these cases, the component that appears together

[13] In this book, Odriozola (2003) makes a proposal on the regularities regarding verb derivation in Basque.

with the verb displays a close relation with it, which suggests that the verb and the accompanying component form a semantic unit. However, with respect to our project, the fact that they behave as a single unit produces syntactic structures that usually do not surface when the verb appears in isolation (for example, unlike the verb *ekarri* ('to bring'), the phrase *gogora ekarri* ('remind') accepts subordinate clauses of the *-ela* type. In addition, the element that accompanies the verb is not often the type of element that the verb would take in isolation. For example, in the phrase *hegaz egin* (literally 'wing-with do', meaning 'to fly'), the Case in the accompanying element is instrumental Case. However, as noted by Rodríguez and García Murga (2003), predicate *egin* in isolation does not include the instrumental Case as one of its outstanding Cases. These are some of the reasons that we have taken into account when determining whether a phrase should be considered as a unit or not. Nevertheless, there are units that involve several components where the accompanying element displays a syntactic structure that is compatible with the structure that the verb would take in isolation. In such cases, we have considered such complements as valuable elements of the verb, and the semantic value resulting from the composition must be expressed elsewhere (namely, by considering it as a single unit in the dictionary; this is parallel to the instances of lexicalized units that were described in section 3.2.3.4). However, there is much research that needs to be done on these complex units. It is a hard task to decide what elements belong to the verb itself or to the unit as a whole. We hope that our results serve for future research on this topic.[14]

After applying the above criteria for excluding verbs, let us next present the criteria that we have followed for selecting verbs.

— Frequency. We have selected verbs that display more than 1% frequency in the corpus: *izan*[15] ('to be', 'to have') (20,72%), *egin* ('to do') (6,98%), *egon* ('to be/stay') (4,44%), *esan* ('to say') (2,40%), *ikusi ('to see')* (1,75%), *eman* ('to give') (1,61%), *joan* ('to go') (1,49%), *jarri* ('to place/sit') (1,29%), *aritu* ('to be doing') (1,16%), *hartu* ('to take') (1,12%).

— Verbs that are interesting for our procedure: Among the verbs that display frequency rates lower than %1, we have selected verbs that are interesting for their subcategorization properties as well as for the Cases that they accept. Considering the criteria listed above, we have selected the following 100 verbs as our object of study.

— *abestu* ('to sing')
— *adierazi* ('to express')
— *afaldu* ('to have diner')
— *agertu* ('to appear')
— *ahaztu* ('to forget')
— *aldatu* ('to change')
— *amaitu* ('to finish')
— *argitu* ('to clarify')
— *aritu* ('to be doing')
— *asmatu* ('to figure out')
— *atera* ('to take out')
— *aurkitu* ('to find')
— *baieztatu* ('to confirm')
— *banandu* ('to separate')
— *barkatu* ('to forgive')
— *bazkaldu* ('to lunch')
— *besarkatu* (to embrace')
— *bete* ('to fill')

[14] Zabala (2002) has studied complex predicates. Her claims will be a good point of departure to work on this phenomenon.

[15] *Ukan* ('to have') also displays high frequency *(ukan* 6,34%), but we have subsumed it under *izan* ('to be'). Thus, we have added the frequency rate of *ukan* to the frequency of *izan.*

— *bilakatu* ('to become')
— *bisitatu* ('to visit')
— *dedikatu* ('to dedicate')
— *deitu* ('to call')
— *edan* ('to drink')
— *egin* ('to do')
— *egokitu* ('to adapt')
— *egon* ('to stay')
— *ehizatu* ('to hunt')
— *ekarri* ('to bring')
— *elkartu* ('to unite')
— *eman* ('to give')
— *entzun* ('to listen')
— *erabili* ('to use')
— *eragin* ('to cause')
— *eraman* ('to take')
— *erantzun*('to answer')
— *erre* ('to burn/smoke')
— *erreparatu* ('to notice')
— *esan* ('to say')
— *eskaini* ('to offer')
— *eskatu* ('to ask for')
— *etorri* ('to come')
— *eutsi* ('to hold')
— *existitu* ('to exist')
— *ezkondu* ('to marry')
— *flotatu* ('to float')
— *gainditu* ('to overcome')
— *gertatu* ('to happen')
— *gosaldu* ('to have breakfast')
— *grabatu* ('to tape')
— *hartu* ('to take')
— *haserretu* ('to get angry')
— *hasi* ('to start')
— *hautatu* ('to choose')
— *hautsi* ('to break')
— *hazi* ('to grow')
— *hil* ('to die')
— *hornitu* ('to supply')
— *hustu* ('to empty')
— *igo* ('to raise')
— *ikasi* ('to learn')
— *ikusi* ('to see')
— *irakin* ('to boil')
— *irakurri* ('to read')
— *iraun* ('to last')
— *iritsi* ('to arrive')
— *isildu* ('to quiet')
— *isuri* ('to pour')
— *izan* ('to be')
— *jaitsi* ('to descend')
— *jan* ('to eat')
— *jarri* ('to put')
— *jasan* ('to endure')
— *jaso* ('to raise')
— *jo* ('to hit')
— *joan* ('to go')
— *jokatu* ('to bet')
— *jolastu* ('to play')
— *kezkatu* ('to worry')
— *kokatu* ('to place')
— *konparatu*('to compare')
— *konturatu* ('to realize')
— *landatu* ('to plant')
— *landu* ('to elaborate')
— *laztandu* ('to caress')
— *loratu* ('to flower')
— *lortu* ('to achieve')
— *mintzatu* ('to speak')
— *moztu* ('to cut')
— *mugitu* ('to move')
— *nahastu* ('to mess')
— *onartu* ('to accept')
— *oroitu* ('to remember')
— *otu* ('to occur')
— *pasatu* ('to pass')
— *sartu* ('to enter')
— *topatu* ('to meet')
— *ukitu* ('to touch')
— *ulertu* ('to understand')
— *zeharkatu* ('to cross')
— *zintzilikatu* ('to hang')

4. Conclusions drawn from the analysis of verbs

We have drawn many conclusions after analyzing the 100 verbs in detail. In fact, because the different nature of the verbs —some are semantically heavy, and other are lighter— we have found various relevant phenomena.[16] For present purposes, we will mention three relevant phenomena: first, we will present the difficulties that we encountered in determining which are syntactic variants in a given alternation among the existing *ssv*-s of each verb, and which are not. We will further explain the decisions that we made in such instances. Next, we will briefly present and explain the semantic components that we have employed for distinguishing the *ssv*-s. Finally, we will clarify what we understand by subcategorization, and we will explain the difficulties and phenomena related to the realization of subcategorized elements in sentences.

[16] Further details on the results of the analysis are included in the dissertation research that will be available shortly (Aldezabal, forthcoming).

4.1. Distinguishing between syntactic variants and non-variants in an alternation

Our analysis reveals that some verbs are semantically heavier than others. Typically, semantically loaded verbs tend to have few semantic values, and the *ssv*-s that we have marked involve alternations of the same semantic value. In addition, most of the times they do not allow for alternations. We have found 21 verbs that lack alternations and involve a single semantic value, and 44 verbs that have been assigned more than one *ssv* and contain a single semantic value. Thus, out of 100 verbs, 65 involve a single semantic value. The remaining predicates have the ability to express more than one semantic value, and sometimes we find alternations within those semantic values.

It has not been an easy task to decide on the above facts. In fact, we have been forced to make certain decisions when we have encountered such problems.

This section describes the general problems that we have encountered.

— In the general meaning of some predicates (or better, the meaning that is most frequently attested in the corpus) certain Cases that do not appear to be relevant —usually the inessive— refer to the element in the absolutive, where the later specifies the particular location (versus the location of the event denoted by the verb). Sometimes, this phenomenon becomes relevant to the extent that it seems to induce a new different semantic value. Moreover, the element in the absolutive is different from the usual value of the verb (more specifically, for example, in the usual value of the verb the absolutive element is usually animate, and yet, in the new arising value of the verb, it involves a definite or abstract entity). We have considered these two phenomena (the fact that an element may take force and the fact that the absolutive has different value from the usual verb value) for marking a separate *ssv*. E.g. *etorri-3* ('to include'):

> *Bigarren liburu honetan* ***badatoz****, gainera, aurrekoaren zuzenketak*
>
> Lit.: 'Second book this-in come-they in addition, previous-det-gen corrections'
>
> Meaning: 'This second book includes the corrections of the previous one'.

Elsewhere, in cases where the absolutive is not different from the usual value of verb we have not distinguished a separate *ssv*. For example, *erabili-0*:

> *Ez nuen aspaldian argazkirik poltsikoan* ***erabiltzen***
>
> Lit.: 'not did-I for a long time pictures-partitive pocket-in use-Nominalization-Inn
>
> Meaning: 'I had not used pictures in my pocket for a long time.'

— Sometimes, the presence of certain Cases depends on the object or absolutive element that the verb takes. In such instances, some Case that, for a given verb has previously been considered as unacceptable becomes acceptable. Conversely, a Case that has been acceptable may become unacceptable. E.g.:

Egin ('to do'): adlative and ablative

... eta Artikotik Tropikora bidaia ***egin*** *zuen*

Lit.: 'and Artic-ADL Tropic-ABL trip made did'
Meaning: '...and he made the trip from the Artic to the Tropic.'

Here, the dative and adlative Cases, which are not commonly accepted by this verb are acceptable. Moreover, the dative Case, which is commonly accepted by this verb (with the value goal) is not acceptable. Thus, in such we have not accepted these adlative and ablative Cases, because, they arise as a result of some constraint on the element that is selected by the verb rather than by some constraint on the general value of the verb.

— We found that the semantic value may also be altered by the noun heading the phrase, but without altering other Cases. E.g., in the two examples with the verb *topatu (meet/encounter)* below:

Eskolan gazteleraz irakurtzean hitz arrotz asko ***topatzen*** *genituen.* Value: ENCOUNTER

Meaning: 'At school, we used to come across many unknown words when we were reading.'

Festibalak ***topatu*** *ditu estatu batuar aitabitxiak.* Value: INTENTIONALLY LOOK FOR AND FIND

Meaning: 'The godfather in the USA has found festivals.'

The following may also happen: the semantics of a verb may change according to context —often due to pragmatics— even in cases involving the same item.

Arazoen gainetik irtenbidea ***asmatzeko*** *eskatzen dizue, hala ere, gizarteak, urratsak egitea alegia.*

Meaning: 'However, despite the problems, society demands that a solution be sought.'

Ez da ikerketa sakonik egin eta horrelakoetan beti gertatzen da gauza bera, jendeak ***asmatu*** *egiten dituela gauzak.*

Meaning: 'No serious research has been done, and in such cases, people typically make things up.'

In the above two instance, we know that *irtenbidea* ('solution') and *gauzak* ('things') are usually sought/made up. However, these meanings are provided by context; without context, they would have merely meant 'figure out'. Such differences cannot be expressed by the resources that we have selected. Moreover, they are often determined by pragmatic factors. Thus, they involve further semantic specifications, and hence, we have not considered them as distinct *ssv*-s.

— We have mentioned that some verbs do not have much semantic load, i.e., they contain very little or general semantic information. In such instances, their sem-

antic value in each sentence is provided by the nature of the elements that they take in syntax. When faced with such cases, we have had to make certain decisions. First, we will present casuistry, and next we will specify what we have decided in each instance.

- Various semantic values may sometimes be realized with the same combination of Cases, and the differences are set in the head of the phrase, i.e. in *aldatu-3* ('change'):

  Oñatiko ur-hoditeria Urretxuko ur-biltegitik saihesbidera ***aldatzeko*** *proiektua eta lehendabiziko fasearen egite-lanak enkante bidez kontratatzeko baldintza.* VALUE OF CHANGE OF LOCATION

  Meaning: 'the project to change Oñati's water-pipes from Urretxu's water tanks to the by-pass and the condition to contract the first phase of the works through auction.'

  Izan ere, autonomi edo probintzia-mailara ***aldatu*** *nahi baditugu, zatikaturiko inkestak ez dira lehen bezain adierazgarriak.* VALUE OF CHANGE OF STATE

  Meaning: 'In fact, if we wish to change them into autonomy or a province, the divided surveys are not as meaningful as they were before.'

- Sometimes, the nature of the head of the phrase requires the Case combination to be fixed and syntactically explicit. For instance, *joan-2*:

  Urdailetik irteerara ***doan*** *zentimetroko hodia*

  Meaning: 'The one-centimeter duct that goes from the stomach to the exit.'

In this example, the phrase *zentimetroko hodia* expresses the path, and hence, rather than involving some meaning of movement it refers to its location. For this, it seems that the presence of the ablative or the adlative is necessary.

- Other times, different semantic values are expressed by various Case/value combinations. E.g.: *izan-1, izan-2, izan-4:*

  izan-1: - Leopoldo, zu idazlea zara, baina zure familian idazle ugari ***izan*** *dira, horrek zuregan eraginik izan du?*

  Meaning: 'Leopoldo, you are a writer, but there have been several writers in your family, did this have any influence on you?

  izan-2 Hitzarmena ***da*** *bidea*

  Meaning: 'A treaty is the (only) way/solution.'

  izan-4: Ezer ez dute erraza ***izan*** *ezta* ***izanen*** *ere*

  Meaning: 'Nothing was easy for them, nor it will be.'

- What changes (or specifies) the semantics of certain verbs is not the noun head of a phrase, but the presence of the phrase itself. For example, *bilakatu-1* and *egokitu-4:*

  Lianak suge ***bilakatu*** *ziren*

  Meaning: 'Whyps became snakes.'

  Gizartearen baloreak ***bilakatuz*** *doaz gizarte horren kontzeptuekin batera*

  Meaning: 'The values of society are developing parallel to the concept of society.'

  Lehen gazteek beraiek ***egokitzen*** *zituzten euren arauak unean uneko egoerara*

  Meaning: 'In the past, young people would determine their rules according to the situations.

  Betaurrekoak ***egokitu*** *zituen*

  Meaning: 'He/she adjusted his/her glasses.'

Considering the casuistry described above, we have decided the following: Those that display the same combination of Cases but change the semantic value according to the head will be included in the same *ssv*. Those that display the same fixed and syntactically explicit combination of Cases will be treated as different *ssv*-s. Those that show different values through different Case-combinations will be considered as different *ssv*-s. Finally, when the presence of a phrase changes/specifies the semantics, the case(s) that belong to the same *ssv* will be marked as optional and outstanding. However, the optionality will be specified in the explanations that will be provided for verbs, not in the marking-system.

This is a generalization of the phenomena that we have found. Yet, in most cases, the problems must be dealt separately in each verb.

4.2. The semantic specifications we have employed in defining the components of the *ssv*-s

We have made use of certain semantic specifications in order to define the most relevant features of each *ssv*. In fact, one of our goals in the onset was to determine such specifications. We may view such semantic specifications as thematic roles, since, in our view, thematic roles are semantic features of verbs, and therefore, they refer to the semantics of verbs rather than to positions and functions of arguments as is usually suggested. Moreover, in Basque, we need to consider that positions are not stable and that they are usually determined by the so-called *Topic*-structure. In addition, the specification of thematic roles has typically been decided in reference to typical or general values of verbs. However, we suggest that a thorough analysis of verbs requires defining various values of verbs, and in order to distinguish between different values, we need to consider additional features. Thus, in view of the procedure that is typically employed in defining thematic roles, we have preferred the term 'semantic specifications' rather than thematic roles.

We have noted that certain semantic specifications are only understood in relation to other semantic specifications. In other words, there is some dependency between certain semantic specifications. For example, if one component of a verb is an *affected_theme* or a *displaced_theme*, the remaining component (of course, in cases where the verb accept the latter) must be *cause*; when one component is *created_theme*, the other will be *producer*; when one is a *container* the other will be *content*. Where there is a *point of departure* there will be a *goal* —or at least it may appear—, and conversely. In contrast, other specifications such as the *experiencer*, the *theme*, and the *activity* do not show any implications.

Thus, it may happen that one element, say the *producer*, may additionally behave as *point of departure* because the sentence may contain some *goal* (when the set of its relevant specifications does not include *point of departure*). Alternatively, it may behave as a *goal* when the sentence includes a *point of departure* (when the set of its relevant specifications does not include any *goal*). After all, depending on the element of the sentence that we choose as target relation, we accept the fact that one component may have more than one semantic specification (the relation with *goal* is *point of departure*, and the relation to *created_theme*, instead, the *producer*).

However, note that these semantic specifications are not directly related to the so-called selectional restrictions. Thus, the semantic specification *cause* does not invariably refer to inanimate entities (in contrast to the definition given for thematic roles, where agents must be animate), or the specification *experiencer* does not imply *affected_object*. The semantic specifications that we have defined are related to the type of event denoted by the verb. Thus, when there is a change of state, we suggest that there is at least a *cause* and an *affected_object* regardless of their animacy. In general, when a predicate is an activity, we have taken the entity involved in the event as being an *experiencer*; it turns out that, in such cases, the entity involved in the event is not only animate but also human. Hence, the specification and assignment of semantic features depends on the way we view the semantics of the verb. Of course, we may view the semantics of verbs in various ways. As for our position, we have considered various viewpoints, and we have created a list of specifications that best fit the resources we have been considering. Only after we have analyzed the 100 verbs have we been able to define the set of specifications, and we have achieved it by basing on the semantics of the 100 verbs —and sometimes the alternations contained in them.[17]

The list of semantic specifications is provided below. However, note that we do not consider the list to be closed, in the sense that other demands may arise when we analyze other verbs in the future. We believe that we have provided an account of the overall casuistry of verbs. At present, the list contains 24 semantic specifications:

—created theme	—target location	—agent	—container
—displaced theme	—target state	—cause	—content
—affected theme	—departure location	—producer	—feature
—theme	—path	—experiencer	—activity
—state	—point of departure	—cause/experiencer	—measure
—location	—goal	—duration	—attitude

[17] For further details see Aldezabal (forthcoming).

We have also been able to specify certain selectional restrictions in some cases, because, in principle, semantic specifications do not have any implications with regards to selectional restrictions. Here is the list we have defined:

— [±biz] (+/-animate) [+giz] (+human)
— [±konkr] (+/-definite) [+lek] (+location)

Actually, we have selected further semantic specifications for defining entities when analyzing the 100 verbs. However, when defining the *ssv*-s in an abstract way, we have restricted to the list provided above.

Here is the list of the types of verbs that we have created based on those semantic specifications:

Verbs of change of state
Verbs of change of location
Verbs that indicate some change
Verbs that involve movement
Verbs that indicate change of psychological state
Verbs that indicate reaction
Verbs that indicate activity
Verbs that involve creation processes
Verbs involving interchange
Existentials, verbs of happening
Verbs that involve a stative location
Verbs that involve description
Verbs that indicate the passing of the time
Verbs that indicate possession
Verbs that indicate attitude
Verbs that indicate assignment of a feature
Opinion verbs

In the above list, certain verbs contain a richer, and hence, more specific information than others (for example, *verbs that indicate some change* vs. *verbs of change of state*; *verbs of change of state* vs. *verbs that indicate change of psychological state*). In fact, verbs that contain a general sense may obtain more specific values. For this task, we need to determine the relation existing between all the elements of the sentence. This is the reason why it is hard to define semantic sets coherently solely based on syntactic structure. In addition, alternations that are general provide a means of grouping verbs coherently and more abstractly (i.e. causative/inchoative alternation: change verbs). However, there are some verbs that contain the semantics carried out by sharing alternations, and nevertheless, do not display such alternation. Finally, there are some semantically similar verbs that do not display any alternations. Thus, there are various ways or parameters for grouping verb: those that share the semantics, those that contain the same number of relevant components, those that employ the same syntactic realization of such components, or those that share the same alternations. These parameters are not exclusive from each other.

We do not consider that Levin's proposal for classifying verbs may provide us with a coherent classification of verbs. Hence, the study of alternations is not enough to develop the decomposition or the internal composition of verbal items.

4.3. Conclusion: subcategorization from our viewpoint

This content of the article thus far shows that there are many difficulties in binding the internal semantic of verbs and their final meaning in sentences. By now, it is obvious that, in order to analyze the semantic value of verbs in sentences, we need to analyze in depth the internal structure of the verbs as well as the interrelation of the elements that make up sentences. This is even more obvious in verbs that are considered as primitives, such as *izan* ('to be'), *egon* ('to be/stay'), *mugitu* ('to move'), *bilakatu* ('to become'), *aldatu* ('to change'), etc.

Hence, we have not proposed specific groups of verbs. Instead, what we have done is to present the *ssv*-s of the 100 verbs we have analyzed (Case/value-combinations, including alternations), and determine the components that are outstanding in our view as well as the semantic specifications of such components. Thus, we will consider that, verbal subcategorization includes all those *ssv*-s, as well as the outstanding Cases of each *ssv*. In fact, Case specifications of components suggest what the syntactic realization of those components will be. However, this does not imply that all the elements that are included in the subcategorization must have a realization in the sentence. Hence, the fact that some element is semantically necessary and the fact that it may not appear syntactically are reflections of distinct phenomena. The next section presents such cases in detail.

4.3.1. *The presence of semantically categorized components in the sentence: unspecification and ellipsis. Dependency between cases*

It is clear that, apart from the Cases that show agreement in the Auxiliary, other elements (inessives, adlatives, ablatives, sociatives, instrumentals and those containing the suffix *-ela*) have also been taken as part of subcategorization in accordance with the semantics of verbs. However, the later, in contrast to the former, do not display agreement in the auxiliary. This hardens the task of determining their presence in the sentence. The next sections describe phenomena related to this issue.

4.3.1.1. *Unspecification and ellipsis*

Sometimes, the reason why a component is not present in the sentence is ellipsis. This is related to the phenomenon of pro-drop, whereby ergative, absolutive and dative elements may be absent in the sentence. However, even if these phrases may be absent, coreference with a previous argument rescues the interpretation that we need.

In contrast, sometimes we face the problem of unspecification. In other words, it is impossible to recover the element that is absent through ellipsis. More specifically, an element that is typical (in Levin's terms) or general (in terms of Vázquez et al.) in a verb, is not syntactically present with the purpose of reinforcing the event denoted by the verb. Sometimes, this object is attached to the lexical item and appears as a *cognate*. This is, in fact, what we find in the Unspecified Object Alternation and in Cognate Object Constructions.

This phenomenon has been widely analyzed in cases where the element is the semantic and syntactic object of the verb (mostly because, despite the presence of agreement suffixes in the auxiliary, there is no phrase in the sentence that may corefer

with such agreement). However, some authors (among others, Vázquez et al.) analyze unspecified cases that express *target location* or *departure location* in verbs such as those expressing displacement (or change of location). In addition, they also analyze cases of ellipsis involving *target* or *departure location* that are recoverable through coreference or some other devices (like deixis). After all, they pose cases parallel to the ones involving semantic and syntactic objects.[18]

The careful analysis of verbs has also revealed that, apart form the unspecification related to typical elements of a verb, there is unspecification that is based on pragmatic knowledge. In such instances, rather than a typical component, what is being unspecified is a specific element that we take as obvious based on our knowledge about the world, and yet, it does appear in the context. E.g.:

> *Lanestosako Herri Eskolan ere ikasle gehienek D ereduan* ***ikasten*** *dute [batxillergoa], izan ere 15 ikasleetatik 11 eredu honetan daude*
>
> Meaning: 'Similarly, in the town school of Lanestosa most students study in the D model [their secondary studies], in fact, out of 15 students, 11 belong to that model.'
>
> *Urduritasunik gabe* ***erre*** *zuen ordea [tabakoa]*
>
> Meaning: 'He/she smoked [cigarettes] with no nervousness.'

Thus, we may assume that we are facing such instances when the elements that we have considered as part of semantic subcategorization are not overtly realized. This is not easy to determine, however, since most of the times we do not know whether we have general unspecification, unspecification due to pragmatic factors, or whether unspecification results from the fact that the unspecified object is attached to the lexical element.

For example, in the case of the verb *konparatu* ('compare'), if the absolutive shows plural number, and if there is no sociative element in the sentence, it seems that, by default, we understand that the action of comparing involves reciprocity; hence, it seems that the lexical item includes this information, and that the sociative has the ability to specify it. The *ssv*-s of *bete-1-3* that belong to the verb *bete* ('to fill'), we know that something becomes full by filling something into it. However, the object that is used for filling may be absent, probably because the information is understood (for instance, a *sack* will be filled by some element that appears in the context, and similarly with objects such as *bottles; questionnaires* will be filled by answers, etc.). In *jarri-1* and *jarri 3 ssv*-s of the verb *jarri* ('to put'), although the outstanding Case is the inessive, sometimes, it is not explicit in the sentence because of the presence of a dative. However, in such cases, we understand that there is an element that is not specified and makes reference to some part of the body, and that the part belongs to the entity in the dative Case. For example, in the example *txapela jarri zion* (literally 'he put the beret'), we understand *buruan* ('in his head') as the locus of where he put the beret, since it is

[18] In addition, note that these authors consider unspecification of elements that denote *departure location* and *target location* as major alternations in what they call 'Trajectory verbs'. They locate verbs that express displacement (or change of location) within this concept of 'trajectory' or 'path'.

customary to put the beret into one's head. Similarly, in the *ssv*-s of a*ldatu-1-2-3-4* ('to change'), in the absence of ablative and adlative Cases, we assume that the change involves some change of state, unless the context forces some other reading. The *ssv jo-4* ('to keep on') typically implies an ablative and adlative, but the later does not usually surface, and where it does, it must be *aurrera* ('on/forward'). In this case, it looks like the item *aurrera* is sometimes included in the verb itself, and in others, it may surface syntactically.

To summarize, in all these cases we need to assume that the understood information is somehow included in the verb, and hence, it should be included and coded in the lexicon.

4.3.1.2. *Dependencies between cases*

In contrast to the examples in the previous section, not all elements that have been considered as involving outstanding Cases can appear in the text as freely. In other words, sometimes it seems that the presence of some Cases depends on the existence of other Cases. For instance, in the *ssv*-s, *pasatu-1-5* ('to pass'), when the ablative expresses the departure location or state of the source, the presence of the target location or state of the source must be explicit. E.g.:

> *...bata, lehen esan bezala, gaztelaniadunen ghettotik gure gizarte katalanera* ***pasatuko*** *direla pertsona batzuk, gazteak bereziki*
>
> Meaning: '...one, as was mentioned before, that several people, specially the young ones, will **pass** from Spanish-speaking ghettoes to our Catalan society.'

The converse does not hold, however. E.g.:

> *Erran diot juristak errandakoa, eta berak oso argi utzi nahi izan dit ni 3. gradura* ***pasatzeko*** *fax-a heltzeko denbora materialik ez dela izan*
>
> Meaning: 'I told him what the jurist said, and he wanted to make it clear to me that there has not been time for the fax that would allow my **passing** to the 3rd grade.

Similarly, in the case of *joan-2,* when the ablative expresses the departure location, the target location must be present, but here, the presence of the adlative forces the presence of the ablative. Consider the following example:

> *Urdailetik irteerara* ***doan*** *zentimetroko hodia*
>
> Meaning: 'The one-centimetre duct that goes from the stomach to the exit.'

However, in these instances of *joan,* we already mentioned that the head of the absolutive phrase has influence on the appearance of the ablative and the adlative.[19]

[19] This kind of dependency phenomena is analized in Boons (1987), within the "dependent point of departure" concept.

We need to conclude that much research needs to be done in the domains of contextual ellipsis, pragmatic ellipsis, and unspecification. In turn, this confirms that we need to take into account many complex phenomena when linking the internal structure of lexical items and their syntactic realization.

5. Summary and general conclusion

This work has presented the following. First, it has shown the complex phenomena that are involved in verbal subcategorization. Second, it has presented the line of work that we have developed in our field, i.e., in computational linguistics. It is clear that specifying the subcategorization of each verb is a difficult task due to the following reasons: first, because distinguishing the semantic values and the alternations in each verb is problematic, and second, because of the presence of phenomena such as ellipsis, unspecification (of general and specific elements), and dependencies between Cases.

After the research has been completed, we have defined what we have considered as subcategorization, namely, all the semantic/syntactic value(s) that we have defined for each verb *(ssv)*, the set of outstanding elements in each *ssv*, their semantic specifications, and their Case realizations. We have employed various resources in order to define the components that make up subcategorization, and we have tried to provide a coherent proposal based on our resources.

In addition, considering all the phenomena that we have encountered, and along the lines of semantic decomposition, it is clear that we need to consider many features in order to determine the semantic value of predicates in specific contexts as well as to account for the different alternations. In order to complete this task, we would have to look at complex lexicons such as the one suggested by Pustejovsky (1995), and, apart from decomposition, we would have to specify the rules and features that serve in the composition of elements that make up verbs.

We need to point out that there is a big gap between what the current computational approach offers and the demands required by the conclusions of manual analyses. In other words, there is still much work left if we want the computational analyses to achieve the specifications achieved by manual analyses. However, the automatic resources will serve enormously in confirming the conclusions that we have obtained in the areas of combination of Cases, in the nature of the head of the phrase that bears Case, and with regards to outstanding Cases that are not present in the text.

To conclude, our main task has been to explore all these difficulties and to suggest subcategorizations for the initially selected 100 verbs. As we mentioned above, future research will include the confirmation by automatic tools, and at the same time, the analysis of more verbs based on the data we have provided; all these, by applying semi-automatic methods.

References

Aduriz, I., 2000, *EUSMG: Morfologiatik sintaxira Murriztapen Gramatika erabiliz.* Doctoral Dissertation, Basque Philology Section, University of the Basque Country.

—, I. Aldezabal, X. Artola, N. Ezeiza & R. Urizar, 1996, "Multiword Lexical Units in EUSLEM: A Lemmatiser/Tagger for Basque". In *COMPLEX'96 Forth Conference on Computational Lexicography and Text Research.* Budapest, Hungary.

—, J.M. Arriola, X. Artola, A. Díaz de Ilarraza, K. Gojenola, M. Maritxalar, 1997, "Morphosyntactic disambiguation for Basque based on the Constraint Grammar formalism". *Proceedings of RANLP.*

Aldezabal, I., (forthcoming), *Aditzaren azpikategorizazioa aplikazio konputazionalari begira.* PhD dissertation. University of the Basque Country.

—, O. Ansa, X., Artola, A. Ezeiza, K. Gojenola, J.M. Insausti & Lersundi, M., 1999, (= Aldezabal et al., 1999a). *Euskararen Datu-Base Lexikala (EDBL): eskema berriaren proposamena.* Internal report. Department of Computer Sciences. University of the Basque Country.

—, B. Arrieta, X. Artola, A. Ezeiza, G. Hernández & M. Lersundi, 2001, (= Aldezabal et al. 2001a), "EDBL: a General Lexical Basis for the Automatic Processing of Basque". *IRCS Workshop on Linguistic Databases.* Philadelphia.

—, K. Gojenola & M. Oronoz, 1999, (= Aldezabal et al., 1999b), "Combining Chart-Parsing and Finite State Parsing". *Proceedings of the European Summer School in Logic, Language and Information (ESSLLI).* Student Session. Utrecht, the Netherlands.

—, K. Gojenola & K. Sarasola, 2000, "A Bootstrapping approach to parser development". *International Workshop on Parsing Technologies (IWPT 2000).* Trento, Italy.

—, M. Aranzabe, A. Atutxa, K. Gojenola, M. Oronoz & K. Sarasola, 2001, (= Aldezabal et al. 2001b), "Applications of Finite State Transducers to the Acquisition of Verb Subcategorization Information". *Finite State Methods in Natural Language Processing.* ESSLLI Workshop. Helsinki.

—, P. Goenaga, K. Gojenola & K. Sarasola, 2001, (= Aldezabal et al., 2001c). "Extracción masiva de información sobre subcategorización verbal vasca a partir de corpus". *SEPLN'2001.* September 12-14. Jaén.

—, K. Gojenola & K. Sarasola, 2003, "Baterakuntzan oinarritutako euskararen analizatzailea". *Euskalgintza XXI. Mendeari buruz. Euskaltzaindiaren XV. Biltzarra,* September 17-21. Bilbo and Baiona. (forthcoming).

—, M. Aranzabe, A. Atutxa, K. Gojenola & K. Sarasola, 2002, "Learning Argument/Adjunct Distinction for Basque". *ACL'2002 SigLex Workshop on Unsupervised Lexical Acquisition.* July 8-13. Philadelphia.

Alegria, I., 1995, *Euskal morfologiaren tratamendu automatikorako tresnak.* Doctoral Dissertation. Language and Computational Systems Section, University of the Basque Country.

Arriola, J.M., 2000, Hauta-Lanerako Euskal Hiztegi-*ko informazio lexikalaren erauzketa erdi-automatikoa eta bere integrazioa sistema konputazional batean.* Doctoral Dissertation, Basque Philology Section, University of the Basque Country.

Azkarate, M. & P. Altuna, 2001, *Euskal morfologiaren historia.* Elkarlanean.

Boons, J.P., 1987, "La notion sémantique de déplacement dans une classification syntaxique des verbs locatifs", *Langue Française* 76: 5-40.

Chomsky, N., 1965, *Aspect of the Theory of Syntax.* Cambridge, Massachusetts: MIT Press.

—, 1981, *Lectures on Government and Binding. The Pisa Lectures.* Berlin, New York: Mouton de Gruyter.

Dowty, D., 1991, "Thematic protoroles and argument selection", *Language* 67: 547-619.

Euskaltzaindia, 1993, *Euskal gramatika laburra. Perpaus bakuna.* Bilbo: Euskaltzaindia.

Garai, K., I. Ibarretxe, 2002, "'...*-tik* ...*-ra*' egituraren azterketa". Euskararen Semantika eta Pragmatika Jardunaldiak (ILCLI).

Gojenola, K., 2000, *Euskararen sintaxi konputazionalerantz.* Doctoral Dissertation. Language and Computational Systems Section, University of the Basque Country.

Gràcia, M.T.; Azkárate, M.; Cabré, M.T.; Varela, S. *et al.,* 2000, *Configuración morfológica y estructura argumental: léxico y diccionario.* Resultados del proyecto de investigación DGICYT, PB93-0546-C04. University of the Basque Country.

Grimshaw, J., 1990, *Argument Structure.* Cambridge, Massachusetts: MIT Press.

Gruber, J., 1965, *Studies in Lexical Relations.* Bloomington: Indiana University Linguistics Club.

Hale, K.L. & S.J. Keyser, 1987, "A view form the middle", *Lexicon Project Working Papers 10.* Center for Cognitive Science. Cambridge, Massachusetts: MIT Press.

Jackendoff, R.S., 1972, *Semantic Interpretation in Generative Grammar.* Cambridge, Massachusetts: MIT Press.

—, 1990, *Semantic Structures.* Cambridge, Massachusetts: MIT Press.

Jones D., R. Berwick, F. Cho, Z. Khan, K. Kohl, N. Nomura, A. Radhakrishnan, U. Sauerland & B. Ulicny, 1994, *Verb Classes and Alternations in Bangla, German, English, and Korean.* Massachusetts Institute of Technology center for Biological and Computational Learning and the Artificial Intelligence Laboratory.

Karlsson, F., A. Voutilainen, J. Heikkila & A. Anttila, 1995, *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text.* Mouton de Gruyter, Berlin.

Koskenniemi K., 1983, *Two-level Morphology: A general Computational Model for Word-Form Recognition and Production.* Doctoral Dissertation. University of Helsinki.

Levin, B., 1993, *English Verb Classes and Alternations. A preliminary Investigation.* Chicago and London. The University of Chicago Press.

Odriozola, J.C., 2003, "Verb-deriving processes in Basque", this volume, 185-222.

Pustejovsky, J., 1995, *The Generative Lexicon.* Cambridge, London: MIT Press.

Rodríguez, S. & F. García Murga, 2003, "Predicados 'sustantivo + *egin*' en euskara". In J.M. Makazaga & B. Oyharçabal (eds.), *Euskal gramatikari eta literaturari buruzko ikerketak XXI. mendearen atarian. Gramatika gaiak,* Iker-14 (I), Euskaltzaindia, Bilbao, 417-436.

Rothstein, S., 1992, "Case and NP Licensing", *Natural Language and Linguistic Theory* 10, 119-139.

Saint Dizier, P., 1995, "A semantic classification of French verbs based on B. Levin's approach". Research report. IRIT.

Sarasola, I., 1996, *Euskal hiztegia.* Kutxa Fundazioa. Donostia.

Shieber, S.M., 1986, *An Introduction to Unification-Based Approaches to Grammar.* CSLI Lecture Notes 4, Stanford.

Speas, M., 1990, *Phrase Structure in Natural Language.* Kruwer, Dordrescht.

Talmy, L., 1985, "Lexicalization patterns: semantic structure in lexical forms". In Shopen T., ed., *Language Typology and Syntactic Description.* Cambridge University Press: 57-149.

Tapanainen, P., 1996, *The Constraint Grammar Parser CG-2.* Publications of the Department of General Linguistics, 27. University of Helsinki.

Taulé, M., 1995, *Representación verbal en una base de conocimiento léxico.* Doctoral Dissertation, Barcelona.

Urkia, M., 1997, *Euskal morfologiaren tratamendu automatikorantz.* Doctoral Dissertation, Basque Philology Section, University of the Basque Country.

Van Valin, R.D., 1993, *Advances in Role and Reference Grammar.* Amsterdam: John Benjamins Publishers.

Vázquez, G., A. Fernández & M.A. Martí, 2000, *Clasificación verbal. Alternancias de diátesis.* Quaderns de Sintagma 3. Edicions de la Universitat de Lleida.

Zabala, I., 1993, *Predikazioaren teoriak Gramatika Sortzailean (Euskararen kasua).* Doctoral Dissertation. Basque Philology Section, University of the Basque Country.

—, 2002, "Predicados complejos vascos". Manuscript.

AUTOMATIC EXTRACTION OF VERB PATERNS FROM *HAUTA-LANERAKO EUSKAL HIZTEGIA*

Jose Mari Arriola, Xabier Artola, Aitor Soroa

Abstract

This paper presents some of the results obtained by means of the method we developed for the study of verb usage examples, emphasizing as we do so that the primary aim was the development of a method rather than the results per se, and dwelling on the importante of shallow syntactic patterns in obtaining the patterns of the verbs studied. We are concerned with the extraction of verb patterns from the verb entries examples of an ordinary dictionary in machine readable version. The corpus of verb usage examples that we have analysed is composed of 13.089 examples. A shallow analysis allowed us to detect the verb chains and phrasal units that appear with the verb under study. The use of an SGML (Standard Generalized Mark-up Language) data structure to represent the analysed verb entry examples facilitates the extraction of the information contained in this data structure. We present an evaluation of the basic subcategorization patterns found and the principal problems encountered in the automatic extraction of them.

1. Motivation: Why analyse verb examples?

The investigation reported in this article was motivated by two considerations: (1) the use of existing lexical resources in order to contribute to the design of more complete lexical entries for the Lexical Database for Basque (Agirre *et al.* 1995, Aldezabal *et al.* 2001); and (2) the acquisition of a basic subcategorization information of verbs to support our parsing tools. The practical goal of our work is to enrich the information in verb entries with their corresponding basic subcategorization patterns. In that sense we think that our effort could be useful to increase the lexicographer's productivity and to help solving the problem of identifying predicate-argument structures of verbs.

It is widely recognised that verb subcategorization represents one of the most important elements of grammatical/lexical knowledge for efficient and reliable parsing. Researchers in NLP have increasingly felt the need to construct computational lexicons dynamically from text corpora, rather than relying on existing 'static' lexical databases (Pustejovsky and Boguraev 1994). Because of the lack of accurate verb subcategorization information causing half of the parse failures (Briscoe and Carroll 1993), attempts have been made to construct, from empirical data, lexicons that encode

information about predicate subcategorization that capture the valences of the verb and its structural collocations (cf. Brent 1991, Manning 1993, Briscoe and Carroll 1997).

In our project we extract information from a machine readable dictionary (MRD) as a starting point to guide the lexical acquisition from corpora. We think that dictionaries and corpora can and should be combined in the acquisition of this kind of information. The main reasons for deciding to use the verb examples in particular were these:

— More controlled analyses: the dictionary contains, together with other information about each verb, a statement of what type of auxiliary it takes, as well as certitude that the verb will be there.
— Comparison with the main corpus: as we said above, the examples may be considered a kind of specialized corpus because they have been taken from the general corpus. We can thus study low-frequency verbs by obtaining basic information about them from the examples, without needing to resort to much larger corpora.

In view of these reasons, the initial assumption, as stated earlier, is that the examples in the dictionary will be of use in determining the basic subcategorization of verbs.

2. Previous work: from the MRD to a LDB

We considered the *Euskal Hiztegia* (EH) dictionary (Sarasola 1996) an adequate source because it is a general purpose monolingual dictionary, and it covers standard Basque. The content of one entry of the EH dictionary is: headword; date; variants; part of speech; abbreviations (style and usage labels, field labels, etc.); definition; relations; scientific names; examples; subentries and grammatical information. All this information is given implicitly or explicitly in the hierarchical structures of dictionary articles, which are quite complex. The structural complexity presents some problems that must be treated in the analysis and interpretation of the articles. It contains 33.111 entries and 41.699 senses.

The previous work dealt with the conversion of EH (MRD version) into a labelled structure (for more details, see Arriola & Soroa 1996). The MRD version was intended for human rather than machine interpretation. The lexicographer used a text-processor (Word Perfect, Word) to type the entries, so we had to face a text file in which the only available codes were of typographic and lexicographic nature. In order to generate a structured representation of the information contained in the MRD the following three main tasks were carried out: (1) the parsing of the internal structure of the articles; (2) the definition of a grammar of entries that covered the general structure of the dictionary (as a Definite Clause Grammar (DCG) in Prolog) and (3) the conversion of the labelled structure which was encoded automatically following the Text Encoding Initiative (TEI) guidelines (Sperberg-McQueen *et al.* 1994). The TEI guidelines have been applied to the dictionary with considerable ease.

As a result of this conversion process we recognised the structure of the 98,49% of the entries with all the information contained in them, being the error rate of 3% (evaluation based on a sample). There were some errors referred to the date or some

grammatical codes, but the part of speech, definition, examples and so on were correctly recognised.

Through the work of adaptation we have taken a first step to facilitate the study of dictionary examples. It also provides an opportunity to take note of the problems and weaknesses of the lexicographer's approach for building the dictionary. The work of preparation for subsequent automatic analysis makes manifest the dictionary's structure; this is seen particularly in the parsing grammar. This is the grammar that the lexicographer had in mind when producing the dictionary.

3. Corpus of vers usage examples

The corpus of verb examples that we have been able to analyse in the previous work is composed of 13.089 examples. These examples were extracted by the lexicographer when writing the dictionary from a very large corpus in order to show the actual usage of the verbs. So we can consider it a specialised corpus.

The average of words per example is 6,44. This implies that sentences are not too complex and we expected this made them appropriate for the subcategorization extraction process. However, sometimes we had to reject some examples as material for automatic subcategorization, when these consist of incomplete sentences containing syntactic structures that are not pertinent to the verb under consideration. Consider for example *Zaldiak alhatzen diren soroa* 'The field where horses graze'. Here a relative clause is used as an example to indicate the usage of the verb *alhatu* 'to graze'. A shallow parse would correctly detect the absolutive subject, *zaldiak* 'horses', but the other noun phrase, *soroa* 'field', has no argument function vis-à-vis the verb *alhatu*. There is no criterion for deciding between a subject or object function for *soroa*, without specifying another verb outside the relative clause, which is not provided in the example. Since only the relative part of the sentence is given, no choice is possible. Information extracted from such examples will therefore show a higher proportion of error.

4. A methodology for the analysis of verb usage examples

In this section we describe the steps followed for the analysis of verb usage examples (Arriola *et al.* 1999). The main bases in the analysis of the examples are the morphological analyser and the disambiguation grammar.

4.1. Morphological analysis of example sentences

The two-level morphological analyser (Alegria *et al.* 1996) attaches to each input word-form all possible interpretations and its associated information. The result is the set of possible analyses of a word, where each morpheme is associated with its corresponding features in the lexicon: category, subcategory, declension case, number and definiteness, as well as the lexical level syntactic functions and some semantic features. The full output of the morphological analysis constitutes the input for the processes of context-based morphological disambiguation and syntactic function assignment.

4.2. Morphological disambiguation and assignment of syntactic functions

We chose the Constraint Grammar (CG) formalism (Karlsson *et al.* 1995) to disambiguate and analyse the examples syntactically. CG is based not on context-free grammars but on rules encoded in finite state automata. The fact that is morphology-based makes it attractive in our case because of Basque's morphological complexity. Moreover, the fact that it is aimed to process real texts and implemented through automata makes it a robust and efficient tool. For these reasons a decision was made in favour of CG for the writing of a general Basque parser (Aduriz *et al.* 2000). We also believe it to be an adequate solution for the purpose of analysing the verb examples in EH. As Abney (1997) points out, shallow parsers have been used, among other things, for extracting subcategorization patterns. Therefore we developed a shallow syntax, a constraint grammar for Basque or EUSMG, following CG formalism.

```
/<lemma          ausiki, ausikitzen>/
      /<Category          verb. >/
      /<Type _of _Auxiliary  DU>/
      /<Example>/
      "<$.>"
      PUNT-PUNT
      "<Basurdeek>"
      "basurde" NOUN COMMON ERG PL DEFINITE @SUBJ
      "<ausikiko>"
      "ausiki" V SIMPLE PART PERFECTIVE DU @-FMAINVERB
      "ausiki" V SIMPLE PART S DEFINITE GEL ABS UNDEFINITE DU @<NCOMP
   @NCOMP> @ADVERBIAL @OBJ @SUBJ @PRED          "ausiki" V SIMPLE PART
   DEFINITE GEL S DEFINITE DU @<NCOMP @NCOMP> @ADVERBIAL
      "ausiki" NOUN COMMON S DEFINITE GEL ABS UNDEFINITE IWLP @<NCOMP
   @NCOMP>
      "ausiki" NOUN COMMON S DEFINITE GEL IWLP @<NCOMP @NCOMP>
      "<gaituzte>"
      "*edun" AUXV PRESENT_OF_INDICATIVE TRANSITIVE 1stPER_PL
   3rdPER_PL@+FAUXVERB
      "*edun" SYNTHETICV PRESENT_OF_INDICATIVE TRANSITIVE 1stPER_PL
   3rdPER_PL @+FMAINVERB
      "<gutxien>"
      "gutxi" ADJ GEN PL DEFINITE ABS UNDEFINITE @<NCOMP @NCOMP> @OBJ
   @SUBJ @PRED
      "gutxi" ADJ GEN PL DEFINITE GEN DEFINITE @<NCOMP @NCOMP>
      "gutxi" ADJ SUPERLATIVE ABS UNDEFINITE @OBJ @SUBJ @PRED
      "gutxi" ADJ SUPERLATIVE
      "gutxi" DET ABS UNDEFINITE @OBJ @SUBJ @PRED
      "gutxi" DET UNDEFINITE
      "<ustean>"
      "uste" NOUN COMMON S DEFINITE INESIVE @ADVERBIAL
      "<$.>"
```

Example 1. Example before the analysis process: *Basurdeek ausikiko gaituzte gutxien ustean* 'The wild boars will bite us when we least expect it'

The Basque Constraint Grammar that currently contains 1.100 rules works on a text where all the possible interpretations have been assigned to each word-form by the morphological analyser. The rules are applied by means of the CG-2 rule compiler developed and licensed by Pasi Tapanainen (1996). On the basis of eliminative linguistic rules or constraints, contextually illegitimate alternative analyses are discarded. As a result we get almost fully disambiguated sentences, with one interpretation per word-form and one syntactic label. But there are word-forms that are still morphologically and syntactically ambiguous. At this point we are aware that there can also be analysis errors and, consequently, due to the remaining ambiguity and the errors, the results of the extraction process must be manually checked.

In order to improve the disambiguation process performed by the grammar, apart from the information of the output of the morphological analyser we use the information contained in the dictionary itself. We add in the morphological reading of the verb entries the tag corresponding to the type of auxiliary[1] that appears in the dictionary. This tag is useful to discard some interpretations that do not agree with the type of auxiliary.

Apart from that, a new tag is added for us as a result of the assumption that those readings of the verb under study which do not have the verb category in their interpretation have less probabilities to occur in an example: the tag IWLP (interpretation with less probabilities). This tag is only used by the disambiguation grammar in the case we have not enough linguistic information to discard this interpretation. In the example 1 we can see a verb entry example in which we have added the above mentioned tags[2] to the verb entry interpretation before the analysis process.

4.3. Analysis of verb chains and phrasal units

At this stage we have the corpus syntactically analysed following the CG syntax which stamps each word in the input sentence with a surface syntactic tag. In this syntactic representation there are not phrase units. But on the basis of this representation, the identification of various kinds of phrase units such as verb chains and noun phrases is reasonably straightforward. For that purpose we base on the syntactic function tags designed for Basque (Aduriz *et al.* 1997). We can divide these tags into three types: main function syntactic tags, modifier function syntactic tags and verb function tags. The last ones are used to detect verb chains. This distinction of the syntactic functions is essential for the subgrammars that have been developed apart from the general grammar. These subgrammars are CG-style grammars that contain mapping rules.

4.3.1. *Subgrammar for verb chains*

We use the verb function tags like as for example: @+FAUXVERB, @-FAUXVERB, @-FMAINVERB, @+FMAINVERB, etc.; and some particles: the negation particle and

---

[1] The verb in Basque is split up into two components: the main verb and the auxiliary. The lexical meaning and aspectual information is encoded in the main verb, while tense and mood are encoded in the auxiliary. Moreover, the auxiliary can exhibit up to three agreement morphemes corresponding to the absolutive, dative and ergative cases.

[2] The syntactic function tags designed for Basque are based on the Constraint Grammar formalism. The set of categories, syntactic functions and abbreviations used in the article are explained in Appendix A.

the modal particles, in order to detect verb chains. Based on these elements we are able to make explicit the continuous verb chains as well as those that are not continuous. The tags attached to mark-up the continuous verb chains are the following:

— %VCH: this tag is attached to a verb chain composed only by one element.
— %VCHI: this tag is attached to words with verb syntactic function tags that are linked to other words with verb syntactic function tags and constitute the initial element of a complex verb chain.
— %VCHE: this tag is attached to words with verb syntactic function tags that are linked to other words with verb syntactic function tags and constitute the final element of a complex verb chain.

The tags used to mark up the non-continuous verb chains are:

— %NCVCHI: this tag is attached to the initial element of a non-continuous verb chain.
— %NCVCHC: this tag is attached to the second element of a non-continuous verb chain.
— %NCVCHE: this tag is attached to the final element of a non-continuous verb chain.

As we can see in Example 2 the maximum length of a non-continuous verb chain is of three elements.

```
"<$.>"
   PUNT-PUNT
"<Euriak>"
   "euri" NOUN COMMON ERG S DEFINITE @SUBJ %PHR
"<ez>"
   "ez" PARTICLE CERTAINTY @PRT %NCVCHI
"<du>"
   "*edun" AUXV PRESENT_OF_INDICATIVE TRANSITIVE 3rdPER_ABSS
3rdPER_ERGS @+FAUXVERB %NCVCHC
"<ia>"
   "ia" ADVERB COMMON @ADVERBIAL %PHR
"<kalea>"
   "kale" NOUN COMMON ABS S DEFINITE @OBJ %PHR
"<busti>"
   "busti" V SIMPLE PART PERFECTIVE DU @-FMAINVERB %NCVCHE
"<$.>"
```

Example 2. A non-continuous verb chain and its corresponding syntagmatic units: *Euriak ez du ia kalea busti* 'The rain has scarcely wetted the street'

4.3.2. *Subgrammar for noun phrases and prepositional phrases*

Our assumption is that any word having a modifier function tag is linked to some word with a main syntactic function tag. And a word with a main syntactic function tag can by itself constitute a phrase unit. Taking into account this assumption we establish three tags to mark up this kind of phrase units:

— %PHR: noun phrases or prepositional phrases; this tag is attached to words with main syntactic function tags that constitute a phrase unit by themselves.
— %PHRI: this tag is attached to words with main syntactic function tags that are linked to other words with modifier syntactic function tags and constitute the initial element of a phrase unit.
— %PHRE: this tag is attached to words with main syntactic function tags that are linked to other words with modifier syntactic function tags and constitute the end of a phrase unit.

The aim of this subgrammar is to attach to each word-form one of those three tags in order to delimit the noun phrases and prepositional phrases. They make explicit the linking relations expressed by the syntactic functions and facilitate the recognition of phrase units. In Example 3 some examples of the analyses got after applying the above mentioned subgrammars are shown:

```
"<$.>"
    PUNT-PUNT
"<Harria>"
    "harria" NOUN COMMON ABS S DEFINITE @OBJ %PHR
"<zortzi>"
    "zortzi" DET PL ABS @ID> %PHRI
"<aldiz>"
    "aldiz" NOUN COMMON INS UNDEFINITE %PHRE
    "aldiz" LOT LOK @LOK
"<jaso>"
    "jaso" V SIMPLE PART PERFECTIVE DU @-FMAINVERB %VCHI
"<du>"
    "*edun" AUXV PRESENT_OF_INDICATIVE TRANSITIVE 3rdPER_ABSS 3rdPER_ERGS
@+FAUXVERB %VCHE
"<minutu>"
    "minutu" NOUN COMMON @CASE_MARKER_MOD> %PHRI
"<batean>"
    "bat" DET INE S DEFINITE @ADVERBIAL %PHRE
"<$.>"
```

Example 3. A continuous verb chain and the corresponding syntagmatic units detected: *Harria zortzi aldiz jaso du minutu batean* 'He picked the stone up eight times within a minute'

4.4. An SGML data structure for the exploitation of the results

As a result of the steps described in the previous points, the corpus of verb examples contains very rich information. In order to exploit this information we designed an SGML data structure in which we recover the verb usage examples classified by sense code and the type of auxiliary tag that appears in the MRD. We organise verb examples taking into account the sense code and the tag corresponding to the auxiliary type since we think it is interesting to study the impact of these factors in the argument structure. Figure 1 shows how the examples are organised.

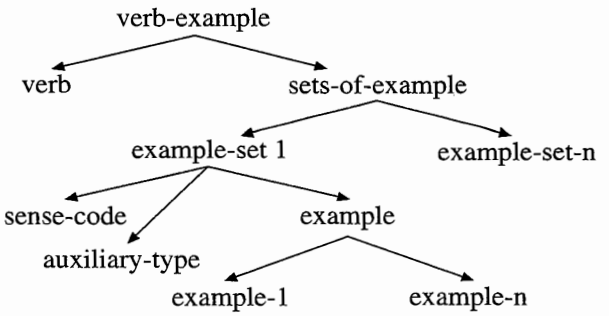


Figure 1. Outline of the organisation of examples

We adopt the SGML mark-up language format for all the corpus of verb examples. From this corpus we extract some pieces of information that we consider more important for verb argument extraction. We choose the verb entry that is object of study with the following information:

— The sense code and the type of auxiliary tag that appear on the MRD.
— The set of examples and the different phrase units that have been detected by means of the above described subgrammars.
— For the verb chains that have been detected, we distinguish between the verb chains that correspond to the verb entry and the other verb chains that can be associated or not with this verb entry. Anyway, for both kinds of verb chains the following information is offered: verb chain, type of auxiliary, syntactic function, person, aspect, modality, mood and time, and the subordinate relation.
— For phrase units we get this kind of information: the phrase unit chain, syntactic function, case, number, definiteness, and subcategorization in the case of nouns. This information is extracted from the last element of the phrase unit.

Apart from these features for each chain or phrase unit of the example, we know its position in the sentence. This is an important factor in order to study the relationship between the verb entry under study and the position in which the different phrase units appear. Those phrase units that are not close to the studied verb entry have fewer possibilities to be considered as arguments. Below we can see the verb usage example we shown in Example 3 represented in this way:

```
<Verb-Chain-Example>
  <Verb> jaso, jasotzen. </Verb>
  <Set-of-Examples>
    <Example-Set>
    <Sense-Code>A1.</Sense-Code>
    <Type-of-Auxiliary>DU</Type-of-Auxiliary>
    <Examples>
      <Example>
        <Example-Sentence>Harria zortzi aldiz jaso du minutu batean.</
Example-Sentence>
        <Verb-Entry-Chain>
          <Chain>jaso du</Chain>
```

```
            <Position>3</Position>
            <Auxiliary-Verb>
               <Base>*edun</Base>
               <Syntactic-Function>@+FAUXVERB</Syntactic-Function>
               <Chain>nuke</Chain>
            </Auxiliary-Verb>
            <Person>
               <PER_ABS>3rdPER_ABSS</PER_ABS>
               <PER_ERG>3rdPER_ERGS</PER_ERG>
            </Person>
            <Mood-Time>Present_of_Indicative</Mood-Time>
            <Main-Verb>
               <Chain>jaso</Chain>
               <Syntactic-Function>@-FMAINVERB</Syntactic-Function>
            </Main-Verb>
         </Verb-Entry-Chain>
         <Phrases>
            <Phrase>
               <Chain>Harria</Chain>
               <Position>1</Position>
               <Part-Of-Speech>NOUN</Part-Of-Speech>
               <Syntactic-Function>@OBJ</Syntactic-Function>
               <Case>ABS</Case>
               <Number>S</Number>
               <Definiteness>DEFINITE</Definiteness>
            </Phrase>
            <Phrase>
               <Chain>zortzi aldiz</Chain>
               <Position>2</Position>
               <Part-Of-Speech>NOUN</Part-Of-Speech>
               <Syntactic Function>@ADVERBIAL</Syntactic-Function>
               <Case>INS</Case>
               <Definiteness>UNDEFINITE</Definiteness>
            </Phrase>
            <Phrase>
            <Chain>minutu batean</Chain>
            <Position>4</Position>
            <Part-Of-Speech>DET</Part-Of-Speech>
            <Syntactic Function>@ADVERBIAL</Syntactic-Function>
            <Case>INE</Case>
            <Number>S</Number>
            <Definiteness>DEFINITE</Definiteness>
            </Phrase>
            </Phrases>
         </Example>
      </Examples>
   </Example-Set>
</Set-Of-Examples>
</Verb-Chain-Example>
```

Example 4. The verb usage example seen in example 3 represented in SGML

5. Evaluation of the analysis

The results of the analysis are referred to the above mentioned subgrammars applied to the output of the disambiguation grammar.

5.1. Evaluation of the verb chains and the phrasal units stablished

After marking verb chains and phrasal units, a random sample of 400 examples was taken out of the total of 13.089 examples. We checked this sample manually, looking at two points in particular:

1) Whether the chain labels were assigned correctly.
2) Whether any elements that should have had a label lacked one. Elements that should have a chain label are those forming part of phrasal units and verb chains discussed in the preceding section.

With regard to the first point, 84 of the examples contained a phrasal unit or verb chain that escaped correct detection. Thus 79% were labelled properly. Wrong labelling occurred chiefly for the following reasons:

— Ambiguity remaining in the examples. Since the chunk marking strategy is based on syntactic functions, ambiguity of syntactic function is a source of problems. But not all ambiguities affect the chunk marking phase. There will be problematic ambiguity when a single word contains both a major syntactic function and a minor one. This kind of ambiguity is of low frequency; it does not reach 2%.
— Disambiguation errors. In this section we include the consequences of incorrect assignments of syntactic function, which affect the identification of chunks.
— Unknown words. These are words for which there is no entry in the Lexical Database for Basque. The words also get analysed by lexicon-independent lemmatisation, but in such cases it is more difficult to get a correct analysis.
— Coordinate phrases. The rules for such structures need to be refined and improved.
— Postpositional structures. We have incorporated some postpositions, but the coverage is incomplete and many are not recognised; these are important for studying verb behaviour.
— Unpredicted structures in parsing label chains. For instance, modifications are necessary in the label set used for parsing structures such as *-ik ena*, as in *Arbolarik ederrena* (English gloss: 'the prettiest tree').
— Other errors. This category includes, *inter alia*, errors inherited from previous phases, such as one case in which a verb's category had been wrongly read as an example due to a mistake occurring in dictionary preparation.

Concerning the second point, elements that should have a chain label are those forming part of phrasal units or verb chains discussed in section 4.3. Therefore we do not take into account for this evaluation certain elements lacking labels, where we have not given rules for them to be labelled as parts of a chunk so they cannot be evaluated. Elements falling outside the labelling rules given include, among others, linkers,

conjunctions, relative clauses, multiple-word lexical units, etc. The chains recognised, with the exception of discontinuous verb chains, are all continuous.

5.2. Evaluation of the assignment of syntactic functions to phrasal units

To measure the accuracy of assignment of syntactic functions to the phrasal units detected, we created a random sample mirroring the characteristics of the whole set of examples, and performed a manual assignment of functions to each phrase. After the manual analysis, we compared this with that obtained automatically. This sample contained 1.211 examples, of which we only checked those containing a single verb, numbering 646.

The following criteria were used:

— We checked for the following functions: subject, object, indirect object and adverbial.
— We checked whether the functions assigned by manual and automatic means agreed. Disagreement, or error, might consist of incorrect marking or failure to mark.

The following table shows the results of the evaluation:

PHRASES	TOTAL	CORRECT	WRONG
MARKED AS SUBJECT	177	126	51
MARKED AS OBJECT	358	251	107
MARKED AS INDIRECT OBJECT	21	20	1
MARKED AS ADVERBIAL	220	213	7

Table 1. Results of the evaluation of the assignment of functions of phrases

As the table shows, indirect object and adverbial function assignment was successful. The weak point is assignment of subject and object functions. Nevertheless we consider the results obtained quite good, since %70 were correctly labelled and our syntactic disambiguation grammar is still under development.

With regard to subject and object assignment, some errors resulted from the difficulty of assigning these functions to arguments of verbs in non-finite form. In such cases, although there is only one verb, we lack the help given by finite auxiliaries whose agreement with subjects and objects facilitates the assignment of syntactic function. There are further difficulties with verbs for which the auxiliary-type specification in the dictionary is not helpful, as with the specification DA-DU (which indicates that the verb may be either intransitive or transitive). Even though such sentences may look simple, with the available resources there is no way to determine, in such examples, the function of every phrase associated with a non-finite verb. To do this, the lexicon needs to contain subcategorization information. For example: *Lana banatu* 'Distribute work'. To determine that *lana* 'work' is the object, the lexicon would have to specify what kind of objects the

verb *banatu* can take. Here there would be a specification of the thematic role of the object. We could then differentiate object from subject: the lexicon would need to state that this verb's agent is animate, whereas its object is inanimate. Thus it is very important for the thematic roles of verbs to be specified, to know what features make it possible for such an element to be either the subject or the object, where it might potentially be either.

Apart from the results shown in the table, the number of phrasal units recognised in the automatic analysis disagrees with that obtained manually (see 5.1, and remember that 79% were correctly detected), and consequently, the number of phrases marked for a given function may be larger of smaller in the automatically marked sample. The automatically marked sample shows 40 more phrasal units than the manually analysed one. On detecting the phrases belonging to a verb and their syntactic function and case, the shallow pattern that emerges is therefore distorted. For example, in *Meza azkendu zen arte* ('Until the mass was finished'), two 'subjects' are found: *meza* (a noun) and *arte* (a subordinating conjunction that happens to be homonymous with a noun), and the result would be to classify this as a verb taking two subjects.

6. Criteria for verb classification

As mentioned earlier, we obtained the analysis of each example through shallow parsing, and proceeded to extract from that analysis features that might be relevant for work on subcategorization. Given the wealth of data, examples may be classified in numerous ways, but in the present case we chose to focus on case and syntactic function. We based our classification of the syntactic structures obtained on the syntactic functions/cases @SUBJ_ERG, @SUBJ_ABS, @OBJ_ABS and @ZOBJ_DAT. With a classification based upon these functions and cases, we examined the lexically realized items that carried these markers in the dictionary examples. Given that it is extremely common in Basque that items related by agreement to the verb are not overtly realized, we should remark that such elided items are not included in our classification.

Of the examples of finite verbs studied, in 500 out of 2.700 there is neither an ergative subject, an absolutive subject, an absolutive object nor a dative indirect object. It is also common in other cases for one or another of these functions to undergo elision; the type of argument most commonly elided is the ergative subject. This fact is significant, and suggests that other cases appearing in shallow structure, cases not included in our shallow patterns, ought to be considered when studying subcategorization. Probably some cases/functions falling outside our analysis of syntactic structure should be included for consideration when determining whether or not they participate in argument structure. Thus for example local cases participate in the argument structure of certain verbs. Here are a few verbs that appeared in classes lacking any ergative subject, absolutive object or indirect object (ZERO-@SUBJ_ERG-@OBJ_ABS-@ZOBJ_DAT) and the cases that occur with each:

atera 'go/take out': 8 examples with local cases: ABL and INE (out of 32 total)
igo 'go up': 4 times ALA and 1 INE (22 total)
iritsi 'arrive, reach': 2 ALA, 2 INS, 1 INE, 1 ABL (17 total)
itzuli 'return': 5 ALA and 1 ABL (32 total)
hurbildu 'approach': 2 ALA and 1 INE (14 total)
dudatu 'doubt': 3 INS (6 total)

In these verbs, which are mainly verbs of motion, the cases that chiefly appear overtly are local cases. With some other verbs the instrumental occurs, such as, in our examples, *aldatu* 'change', *baliatu* 'use', *begiratu* 'look after', and *burlatu* 'make fun (of)'. The cases mentioned are frequently excluded from studies of argument structure, but as we have shown, they probably ought to be considered.

Our reason for not having taken these into account is that they are not the most common cases or functions to participate in argument structure. Since, overall, they rarely appear in a verb's specification for argument structure, they were not made a criterion for establishing the classes. However, more directed analyses can be carried out using the query system,[3] in order to look at examples of verbs taking local cases/functions, for instance. We have extracted the complete analysis of such examples and consequently dispose of information about the cases and functions of phrasal units associated with a given verb. We know what examples are given for each verb, with examples classified according to the sense of the verb and subcategory. This information is preceded by an indication of the verb's participle, the verb's sense, its subcategory and an example number; in this way examples are uniquely indexed. Each index is followed by a shallow parse, first showing the auxiliary type pertaining to the verb according to the dictionary entry, and then pairs of syntactic function and case.[4] If any other verb complexes occur in the same example, this is indicated by the sign MP (for 'subordinate clause') accompanied by + for subordinate or - for non-subordinate.

Thus for example the following patterns are listed for the verb *bultzatu* 'push, press':

```
bultzatu, bultza, bultzatzen.

bultzatu-A0.-DU-1    DU.@SUBJ_ERG-@OBJ_ABS.
bultzatu-A0.-DU-2    DU.@OBJ_ABS.MP+
bultzatu-A0.-DU-3    DU.@ADLG.
bultzatu-A0.-DU-4    DU.@SUBJ_ABS-@OBJ_ABS @PRED_ABS.MP-
bultzatu-A0.-DU-5    DU.@OBJ_ABS-@OBJ_ABS-@ADLG_ABZ-@OBJ_ABS-@OBJ_ABS-@ADLG.MP+
bultzatu-N1.-DU-1    DU.@SUBJ_ERG.MP+
bultzatu-N1.-DU-2    DU.@OBJ_ABS.
bultzatu-N1.-DU-3    DU.@SUBJ_ERG-@ADLG_ALA.
bultzatu-N1.-DU-4    DU.@SUBJ_ERG-@OBJ_ABS.MP-MP+
bultzatu-N1.-DU-5    DU.@ADLG_ABZ-@OBJ_ABS.
bultzatu-N1.-DU-6    DU.@OBJ_ABS.
```

Example 5. Basic verb patterns for the verb *bultzatu* 'push, press'

The shallow pattern class of each verb was obtained automatically and we defined a code identifying the verb examples occurring in each of those patterns. An example will

[3] The query-system as a tool to manipulate the full range of information contained in the examples, in order to derive the most reasonable argument structure (Arriola *et al.* 1999).

[4] Syntactic function and case are linked by an underline character. A hyphen separates function/case pairs.

serve to show what kind of information the code contains. The example is *bultzatu-A0.-DU-2*:

— the participle (used as the verb's citation form), in this case *bultzatu*.
— sense index: specifies the sense, subsense or nuance of the verb in this example, e.g. *A0*.
— auxiliary type: the type of auxiliary indicated in the dictionary (DA, DU, DIO, ZAIO, or DA-DU). In this case, *DU*.
— example number: the examples for each verb are numbered, e.g. *2*.

The appendix of the thesis (Arriola 2000) lists all the verb examples classified by verb, in such a way as to show what shallow syntactic structures show up with what verbs. However, when classifying verbs in the next section, we shall only take function and case into consideration. The appendix shows all examples, but below we will select a few for illustrative purposes, following the above-mentioned criterion.

It needs to be noted too that the set of syntactic functions (Arriola 2000) that were defined affects the range of structures that can be recognised. The shallow structures that are detected correspond, of course, to those defined in our set of syntactic functions. Now these functions are adequate from the point of view of the parser, but when applied to the examples some of the functional distinctions turn out to be undesirable. The distinctions in question are very difficult to decide upon automatically, and consequently incorrect syntactic structures will sometimes be assigned. For example, distinguishing the nominal predicate function @PRED usually led to incorrect identification of structures. In principle we consider it necessary for subcategorization to distinguish the @PRED function; the trouble is that accurate detection of this function is hard to achieve, precisely because the lexicon lacks information about subcategorization. Therefore, it was thought advisable to proceed in our initial analysis without distinction of the function in question.

False recognition of patterns was also caused by the specification, where a subordinate clause was involved, of its function within the main clause. Even though inclusion of such distinctions in the set of syntactic functions is justified on linguistic grounds, this is not appropriate for the purpose of the method we developed. If for example, a verb has associated with it a non-finite subordinate clause, we may detect the subordinate clause but be unable to determine what the non-finite clause's role is vis-à-vis the main clause. To do this requires assistance from subcategorization information. In practice, then, more detailed syntactic functions hinder the disambiguation process and make it more likely for errors to occur in the information that is extracted.

Thus with regard to the set of syntactic tags, it may be concluded from our experiment that specification of the function of subordinate clauses in relation to a main clause, as part of the set of syntactic functions, ought to wait until subcategorization has been described. Likewise, the function of nominal predicate, @PRED, should be specified once there is a working subcategorization. At that point we would have the option of specifying what kind of subordinate clause each verb can take and the functions of the subordinate clauses.

7. The set of shallow patterns detected

In this section we present the shallow patterns that were extracted. The following diagramme shows what patterns were found:

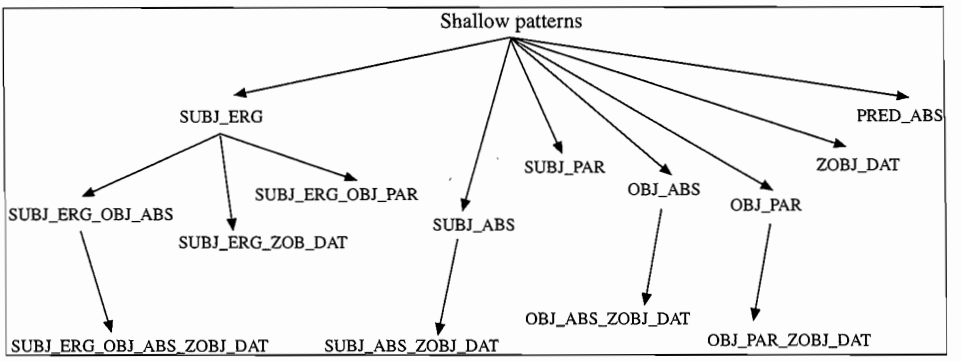


Figure 2. Surface patterns[5] in the examples

As we said before, we consider syntactic functions and cases when classifying examples. In this way, different verbs will be grouped together according to the shallow syntactic functions and cases with which they occur. Although verbs coincide in taking those functions and cases, criteria clearly need to be developed for a finer classification. The present classification is merely a modest first step. Work could begin on thematic roles on the basis of this material, among other sources.

These patterns merely show what structures each verb accepts. As we have pointed out, it takes a deeper analysis to determine what the obligatory arguments of these verbs are. Some authors argue that semantics should come under consideration here, in addition to other factors; Levin (1993) claims that the semantics of a verb determines its syntactic behaviour. In order to facilitate such analyses, we have decided to include information about which sense a verb is used in for each example. However, this task, among others, is for the future.

8. Automatically derived shallow patterns: difficulties and evaluation

In this section we will discuss the main difficulties encountered for classifying verbs on the basis of the methods developed and the reliability of the resulting classification. With regard to the difficulties, we will talk about the limitations of shallow syntax, the limited usefulness of position, and certain features of these verb examples. Following this we evaluate the classification, using measures of reliability for each pattern on the basis of an analysis of a sample.

---

[5] The shallow patterns that are detected correspond, of course, to those defined in our set of syntactic functions.

8.1. Limitations of shallow syntax

In developing the shallow syntax section we took an important step towards verb classification, labelling explicitly the phrasal units and verb complexes associated with a given verb with chunk marker tags (4.3). Thus we must take into account what we are able and unable to detect, i.e. what kinds of phrase (4.3). We furthermore evaluated the phase of phrase detection at the end of the section 5, noting the kinds of problem or error occurring with those phrases that could be detected. We find that of the phrases recognised, 79% were tagged correctly, that is, 79% of the chunks are correctly parsed. It is also necessary to consider the reliability of function and case identification in correctly marked chunks (5.2).

Considering what was said in the sections mentioned, it should be noted that the shallow syntax also fails to specify the relations between main and subordinate clauses. Thus we cannot use data from examples containing more than one verb for classification purposes. For example:

> *Liburu askoz **baliatu** dira idazlan hori prestatzeko.* 'They have **used** a lot of books to prepare that study.'
>
> The lexicographer is illustrating the use of *baliatu* 'use'. But our method is incapable of distinguishing whether *idazlan hori* 'that study' is the direct object of *baliatu* 'use' or of *prestatu* 'prepare'. Thus we cannot be sure of getting a correct analysis, which would be as follows:
>
> *Liburu askoz **baliatu** dira* [*idazlan hori prestatzeko.*] 'They have **used** a lot of books [to prepare that study].'

For a deeper analysis of such sentences, subcategorization data would need to be specified in the lexicon. But of course that information was not available when we started developing the parser.

With our resources it is very difficult to use the parser we developed to determine automatically which verb each argument (or potential argument) belongs to in multiple verb sentences. The information extracted would contain more mistakes if these were included, since the parser has no way of dealing with this problem. Such results would then require much manual work to determine whether automatically produced patterns were right. We preferred for the information extracted automatically to be more reliable and require less manual checking. This led us to study one-verb sentences, but we used some multiple-verb sentences to study the usefulness of position.

8.2. The use of position

We used position to help determine, in examples with more than one verb, which phrases (or subordinate clauses) go with which verb. We attached a number to each phrasal unit and verb complex detected, to indicate the order in which they occur. The order does not determine what function arguments have, except for focalisation, focused elements being placed immediately before the verb. But our hypothesis is that potential arguments and verb complexes do not appear just anywhere, but will normally occur in the vicinity of the verb in whose subcategorization they are included. On this ass-

umption, examples containing more than one verb were truncated according to the following criteria:

— When the verb under primary consideration precedes another verb complex, items following the second verb complex are ignored.
— Conversely, if another verb complex precedes the verb complex we are interested in, items preceding the first verb are ignored.

In the former case, where a second verb complex occurs later than the verb under consideration, then, it was decided not to count phrasal units occurring after the second verb. The example is truncated at that point; however, the second verb complex itself *is* counted, since it is possible that this might be part of the subcategorization of the verb we are considering. For example:

— Original example (the first two verbs in the example are underlined; the verb whose subcategorization is being analysed is in bold): *Zure okerrak **tapatu** nahirik egin dituzu pausuak, zer enganio egin didazun jakitun daude auzoak.*
— The same example after applying the criterion of position, i.e. truncated: *Zure okerrak **tapatu** nahirik egin dituzu...*

What we have done is to truncate the example appearing in the dictionary in order to limit our analysis to the part that remains after truncation. The rationale for this is that pertinent information about the verb being considered is located in the part of the example remaining after truncation, whereas the part of the original example that has been removed does not contain information relevant to the verb under consideration. However, this truncation criterion can give erroneous results, as for example when the two verbs are related by coordination. In such cases the two verbs may share the same arguments, but these will fail to get included in the analysis. For example:

— Original example: *Edanak eragiten ditu eta **erasaten** gauza lotsagarriak* 'Drink brings about, and **causes to be said**, shameful things'
—Truncated example: *...eragiten ditu eta **erasaten** gauza lotsagarriak* '...brings about, and causes to be said, shameful things'

Here our criterion leads us to exclude *edanak* 'drink' from the analysis, even though this is in fact the subject of *erasaten* 'causes to be said'.

Despite our awareness of the complexity of these issues, in our development of a shallow syntax we considered position a useful criterion and applied the truncation principle. To enhance the usefulness of this approach, it would be preferable to be able to take into account conjunctions, linkers and punctuation, assigning position to these and referring to them in the course of the truncation process. But recourse to these elements fell outside the scope of this study.

8.3. Evaluation of the patterns

It is important to evaluate the shallow patterns yielded by the verb classification in order to measure the patterns' reliability. We did this on the basis of section 5.2, checking for each pattern, on the basis of the criteria presented there, how often right or wrong syntactic functions and cases have been assigned. The evaluation was done

over a sample, which contains 1.211 examples of which 646 have a single verb. The 406 examples with more than one verb and the 159 examples in which none of the syntactic functions and cases that we have considered for verb classification occur are omitted.

The evaluation results represent comparisons between automatic and manual classifications. For each pattern, the functions and cases taken into account to classify verbs were checked. As we have said, we looked at whether or not the right functions and cases were assigned. We also remark on functions not appearing in the manual analysis of the sample but marked in the automatic analysis. The results show that when there is only an absolutive subject or object in a pattern, accuracy is lower than when these co-occur with other functions. For instance, the results for pattern OBJ_ABS are not as good as those for patterns OBJ_ABS-ZOBJ_DAT and SUBJ_ERG-OBJ_ABS. Indeed, labelling these functions correctly is the biggest problem. Nonetheless the results for pattern SUBJ_ERG are fairly good. Patterns SUBJ_ABS-ZOBJ_DAT and OBJ_PAR are not very reliable, while the most reliable are OBJ_ABS-ZOBJ_DAT and ZOBJ_DAT.

9. Conclusions

Despite the difficulties we encountered in the preceding section, and although the information obtained is *shallow*, we believe that the information may be useful not only as progress in syntactic analysis but also for methodological development. This requires integrating the information obtained into the lexicon for application in parsers. It will take deeper analysis to decide how to incorporate the extracted subcategorization data into the lexicon or parser in such a way as to be useful for parsing.

We also claim to have helped in the aim of facilitating the study of subcategorization in Basque. In that sense we think that the classification achieved provides valuable material for further analysis.

We initially expected the dictionary examples to provide a good source of material for the study of verb behaviour, and as a consequence of the work we have performed on them, that expectation is now even stronger, since the examples have been tagged syntactically and the basic chunks identified. Moreover, the materials have now been converted from plain text to a richer format using SGML, so that all this information will be the more accessible. Use of this encoding also facilitates the development of a query system; new methods and opportunities for research have thus been created (Arriola *et al.* 1999). Through the identification of numerous features, the material can now be employed to study various aspects of verb behaviour. In our own study we have used case and syntactic function, as was seen in section 7, to classify verbs.

We have developed a shallow syntax, with recognition of verb complexes and associated phrasal units, in order to extract a verb classification. If, however, we wish to go beyond the parsing of those units, deeper parsing is required. Specification of the subcategorization of verbs makes it possible to move forward from the analysis of phrases and verb complexes to the analysis of more complex sentences. To develop deeper parsing, of course, we will need to have information on subcategorization that should be specified in the lexicon. In our case, however, we set out with no such information, our goal being to discover which phrases and verb complexes occur in

association with individual verbs, inasmuch as that was possible. There is something of a vicious circle here. On the one hand we perceive the need to strengthen the syntax component in order to obtain information about subcategorization, and on the other, subcategorization information is essential for parser improvement. Notwithstanding, we believe the shallow analysis achieved is a valuable aid for further work on Basque subcategorization.

10. References

Aduriz, I., Alegria, I., Arriola, J.M., Artola, X., Díaz de Illarraza, A., Gojenola, K., Maritxalar, M., 1997, "Morphosyntactic Disambiguation for Basque Based on the Constraint Grammar Formalism", *Proc. of RANLP'97*, Tzigov Charch, Bulgaria.

Agirre, E., Arregi, X., Arriola, J.M., Artola, X., Díaz de Illarraza, A., Insausti, J.M., Sarasola, K., 1995, "Different issues in the design of a general-purpose Lexical Database for Basque". *Proc. of NLDB'95*, 299-313, Versailles, France.

Aldezabal, I., Ansa, O., Arrieta, B., Artola X., Ezeiza, A., Hernández, G. & Lersundi, M., 2001, "EDBL: a General Lexical Basis for The Automatic Processing of Basque". *IRCS Workshop on Linguistic Databases*, Philadelphia, USA.

Alegria, I., Artola, X., Sarasola, K., Urkia, M., 1996, "Automatic morphological analysis for Basque", *Literary & Linguistic Computing* 11.4, 193-203, Oxford University Press. Oxford.

Arriola, J.M., 2000, *Euskal hiztegiaren azterketa eta egituratzea ezagutza lexikalaren eskuratze automatikoari begira. Aditz-adibideen Analisia Murriztapen-gramatika Baliatuz Azpikategorizazioaren Bidean*, Ph. D. dissertation, Gasteiz.

— and Soroa, A., 1996, "Lexical Information Extraction for Basque". In *Proc. of CLIM'96*, Montreal.

—, Artola, X., Maritxalar, A., Soroa, A., 1999, "A methodology for the analysis of verb usage examples in a context of lexical knowledge acquisition from dictionary entries", *Proc. of EACL'99*, Bergen, Norway.

Brent, M., 1991, "Automatic acquisition of subcategorization frames from untagged text". In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics.* Berkeley, C.A., 193-200.

Briscoe, T., and Carroll, J., 1993, "Generalized probabilistic LR parsing for unification-based grammars", *Computational Linguistics* 19, 25-60.

— and —, 1997, "Automatic extraction of subcategorization from corpora", *Proceedings of ACL, SIGDAT Workshop on very Large Corpora.* Copenhagen.

Karlsson, F., Voutilainen A., Heikkila J., Anttila A., (eds.), 1995, *Constraint Grammar: A Language-independent System for Parsing Unrestricted Text.* Berlin and New York: Mouton de Gruyter.

Manning, C., 1993, "Automatic acquisition of a large subcategorization dictionary from corpora", *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, Columbus, Ohio, 235-242.

Pustejovsky, J. and Boguraev, B., 1994, "Lexical knowledge representation and natural language processing". In Pereira, F. and Gross, B. (eds.), *Natural Language Processing*, Massachusetts: The MIT Press, 193-223.

Sarasola, I., 1996, *Euskal Hiztegia.* Kutxa Fundazioa: Donostia.

Sperberg-McQueen, C.M., Burnard, L., 1994, *Guidelines for Electronic Text Encoding and Interchange.* Chicago & Oxford.

Tapanainen, P., 1996, *The Constraint Grammar Parser CG-2* University of Helsinki. Publications n. 27.

Appendix A

@+FAUXVERB: finite auxiliary verb.
@+FMAINVERB: finite main verb.
@<NCOMP: postposed adjectival.
@ADVERBIAL: adverbial.
@CASE_MARKER_MOD>: modifier of case bearing item.
@-FAUXVERB: non-finite auxiliary verb.
@-FMAINVERB: non-finite main verb.
@LOK: linker.
@NCOMP>: preposed adjectival.
@OBJ: object.
@PRED: predicative.
@SUBJ: subject.
@SUBJ_ERG: ergative subject (in this pattern we find transitive verbs with no object).
@SUBJ_ERG-@OBJ_ABS: ergative subject and absolutive object (transitive verbs with an object).
@SUBJ_ABS: absolutive subject (this pattern occurs with intransitive verbs).
@SUBJ_ABS-@ZOBJ_DAT: absolutive subject and dative indirect object.
@OBJ_ABS: absolutive object.
@OBJ_PAR: partitive object.
@ZOBJ_DAT: dative indirect object.
@OBJ_ABS-@ZOBJ_DAT: absolutive object and dative indirect object.
@ZOBJ: indirect object.
1stPER_PL: first person of plural.
3rdPER_ABS: third person of singular (absolutive).
3rdPER_ERG: third person of singular (ergative).
3rdPER_PL: third person of plural.
ABS: absolutive on nominals.
ABZ: ablative of direction.
ADJ: adjective.
ADVERB: adverb.
ALA: alative.
AUXV: auxiliary verb.
CERTAINTY: certainty.
COMMON: common.
DA: intransitive auxiliary.
DAT: dative.
DEFINITE: definite.
DET: determiner.
DIO: transitive auxiliary (with dative object).
DU: transitive auxiliary.
ERG: ergative.
GEL: genitive of location.
GEN: genitive of possesion.
INS: instrumental.
IWLP: interpretation with less probabilities.
LOK: link particle.
LOT: link particle.
MP: subordinative clause.
NOUN: noun.
PART: participle.
PL: plural.
S: singular.
SIMPLE: simple.
SUPERLATIVE: superlative.
SYNTHETICV: synthetic verb.
TRANSITIVE: transitive.
UNDEFINITE: undefinite.
V: verb.
ZAIO: intransitive auxiliary (with dative object).
ZERO-@SUBJ_ERG-@OBJ_ABS-@ZOBJ_DAT: verbs that appeared in classes lacking any ergative subject, absolutive object or dative indirect object.

THE CASE OF AN ENLIGHTENING, PROVOKING AND ADMIRABLE BASQUE DERIVATIONAL SUFFIX WITH IMPLICATIONS FOR THE THEORY OF ARGUMENT STRUCTURE*

Xabier Artiagoitia

Abstract

This article analyzes morphological evidence from Basque to support one basic claim: that subjects of the object-experiencer (i.e. frighten-*type) psych verbs are internal arguments. The derivational suffix* -garri *provides the relevant evidence. This suffix is traditionally characterized as forming adjectives from verbs and, disputably, from nouns or adjectives and as having both an active and a passive meaning. I first establish on several grounds that* -garri *is basically a deverbal suffix which forms adjectives productively. Secondly, I show that the so-called* passive *value of the suffix is restricted to diadic transitive verbs: the internal argument of the verb becomes the external one of the adjective. The so-called* active *value of the suffix is restricted to psych verbs with experiencer objects: the surface subject of the verb becomes the external argument of the adjective. Thirdly, a unified characterization of the* -garri *suffixation is proposed along the lines of Grimshaw (1990): if subjects of the* frighten *class are internal arguments, there is one single rule of* -garri *suffixation which adds an R(eferential) argument to bind the first internal argument of a diadic verb; the original external argument, if there is one, is supressed. Verbs whose surface subject bear an instrumental* θ*-role also admit the suffix* -garri *(the subject becomes the external argument of the adjective); this fact suggests that instrumental subjects count as internal arguments in Basque. The mere existence of* -garri *supports the claim that the* frighten *verb class lacks an external argument; its existence can be also taken as a fair prediction of Belleti and Rizzi's unaccusative analysis of the* frighten *class, but runs counter to both Pesetsky's analysis (1995) and a purely transitive-causative analysis of object experiencer verbs (cf. Arad 1999a-b).*

---

* This is a shortened and slightly modified version of Artiagoitia (1997), published as a (1995) article; the insistence of some friends and colleagues that I do an English version of that article and the realization that I often find myself in the need to explain the contents of that article to non-Basque scholars have finally convinced me that an English version is necessary. This update is one way or another in debt with Lisa Galvin, Maider Huarte, Cecile McKee, Juan C. Odriozola, Antxon Olarrea, Javier Ormazabal, Beñat Oyharçabal and, above all, Koldo Zuazo, to whom I owe a lot of examples and grammatical judgements. My students in the 1997, 1999 and 2003 morphology classes were also a big and stimulating help. The remaining errors are only mine. This research is funded by the research projects 9/UPV 00027.130-13587/2001 and Euresco BFF 2002-10379. The article uses the following abbreviations: art = article, aux = auxiliary verb, D = dative, E = ergative, gen = genitive, lit = literally, loc = locative.

0. Introduction

The purpose of this article is twofold: to show that the Basque morpheme *-garri* is indeed one and only deverbal suffix thus dismantling the classical view that there is a passive *-garri* and an active *-garri*; and secondly, to strengthen and support Belleti and Rizzi's original insight that object experiencer verbs (their *preoccupare* class) have two internal arguments. The contribution I make here can be understood from two angles: from the bascologist's point of view, this article sets out to settle an old issue in Basque morphology using the tools of modern generative grammar; from the generative grammarian's point of view, this article tries to show that, once Belleti and Rizzi's treatment of *preoccupare* and *piacere* verbs is assumed to be correct, the Basque derivational suffix *-garri* is precisely the kind of morpheme which they predict will exist, for it treats the internal argument of simple transitive verbs and the deepest internal argument of the *preoccupare* and *piacere* psych verbs alike.

The article is structured as follows: in section 1 I present the relevant data and the previous accounts of the suffix *-garri*. Section 2 establishes on several grounds that *-garri* is indeed a *deverbal suffix* which forms adjectives productively. The third part studies the verb classes that may be the base for *-garri* suffixation and shows that all these classes share fundamentally the same argument structure: they are all biargumental verbs; given the correctness of Belleti and Rizzi's treatment of *preoccupare* verbs, the rule of *-garri* suffixation affects the same kind of argument, namely the deep direct object of the verb. Part four is devoted to showing that the class of verbs that give rise to the so called *active* value of *-garri* corresponds to Belleti and Rizzi's *preoccupare* class. Finally, section 5 reviews four different proposals regarding the morphology/argument structure and psychological verb/argument structure connection; Grimshaw's approach seems to be the most adequate for a unified analysis of *-garri*.

I warn the reader that this article doesn't intend to argue for a specific theory of derivational morphology and its interaction with argument structure or lexical semantics; on the contrary, it strives to present a piece of (hopefully) interesting data that may help clarify theoretical issues to others.

1. Presenting *-garri*

The Basque suffix *-garri* is traditionally characterized as: (a) forming adjectives and nouns from verbs and, disputably, from nouns/adjectives; and (b) as having both an *active* and a *passive* meaning. The data in (1-2) illustrate the first point:

(1) undisputable verbal base

verb	derived word
a. *erakarri* "attract"	*erakargarri* "attractive"
b. *bete* "fill"	*betegarri* "filling" (n/adj), "filler" (n)
b. *gehitu* "add"	*gehigarri* "additive" (n), "additional" (adj)
c. *ikusi* "see"	*ikusgarri* "spectacular, visible"
d. *kontsolatu* "console"	*kontsolagarri* "consoling/consolable"
e. *ulertu* "understand"	*ulergarri* "understandable"

(2) disputable verbal base

	noun/adjective	verb	derived word
a.	*argi* "bright", "light"	*argi(tu)* "shine, clarify"	*argigarri* "clarifying"
b.	*eder* "beautiful"	*eder(tu)* "embellish"	*edergarri* "embellishing, embellisher"
c.	*harri* "stone"	*harri(tu)* "astonish"	*harrigarri* "astonishing"
d.	*ikara* "fear"	*ikara(tu)* "scare, frighten"	*ikaragarri* "scaring, frigthening"
e.	*lagun* "friend"	*lagun(du)* "help"	*lagungarri* "helping"

A look at a contrast between (1) and (2) shows that, although the base of *-garri* is generally taken to be a verbal root (Azkue 1925, Villasante 1974, Azkarate 1990), there are some cases where that claim seems questionable: there are many noun-verb or adjective-verb pairs, where the verb is derived by zero suffixation. The citation form of a Basque verb is the participle (i.e. the *-tu* form or *-i* form),[1] so the verb root is homophonous with the corresponding noun or adjective; hence, the verb roots in the examples in (2) are best characterized as *argi, eder, harri, ikara* and *lagun*, respectively. The fact that Azkue (1925) and Villasante (1974) paraphrase one meaning of *-garri* as "producer of" gives the impression that in many cases *-garri* must indeed attach to noun roots. Azkarate (1990), perhaps led by this confusing paraphrase, proposes to derive words like *ikaragarri* (= 2c) "scaring, frightening" from verbal expressions like *ikara sortu, eman, eragin* "create, produce, give fright", implicitly assuming that the base may well be a noun in the relevant cases.

The data in (3-4) illustrate the second traditional claim about the value of the suffix:

active value

(3) ***Berri horrek*** *Jon hunkitu du* → ***Berri hori*** *oso hunkigarria da*
new that.E move aux new that very moving is
"The news deeply impressed John" "That piece of news was very impressive"

passive value

(4) *Jonek* ***film hori*** *ikusi du* → ***Film hori*** *ikusgarria da*
Jon.E film that see aux film this spectacular is
"John saw that movie" "That movie is worth seeing"

The *active/passive* terms simply reflect the fact, pointed out most clearly by Azkarate (1990), that the subject of the derived adjective may correspond to the original subject of the verb in the the so called active use of *-garri* as is the case in example (3), while it corresponds to the original direct object of the verb in the passive use of *-garri*, as in (4).[2]

Another unsettled question is the category of the suffix *-garri*: although Azkue (1925) clearly mentions that it can be both an adjective and a noun, Villasante (1974) and

[1] *-tu* is the only productive participle ending in modern Basque; other endings are *-i* (cf. *ikusi* above), *-n* or simply *ø*, in cases where the participle and the verbal root is homophonous.

[2] The passive value becomes more prominent when giving the relevant translation: *ikusgarri* "property of something that deserves being seen"; interestingly, it is debatable whether Basque has a true passive construction.

Azkarate (1990) only talk about its adjectival value. In the next section I shall argue that *-garri* is uniformly attached to verbal roots and that it is only productive as adjectival suffix.

2. The suffix *-garri* only attaches to verbs and is an adjective

There are three basic arguments to claim that the base of *-garri* is also a verb even in the disputed cases like those in (2). I summarize them below.

2.1. The regularity in the meaning of the derived word

The meaning of the derived word is best understood taking the verb as the base. Take the following examples:

(5) a. *aberasgarri* "enriching"	(cf. *aberats* "rich"; *aberastu* "become rich, enrich")
b. *argigarri* "clarifying"	(cf. *argi* "bright, light"; *argitu* "clear, clarify")
c. *harrigarri* "surprising"	(cf. *harri* "stone"; *harritu* "get astonished/ astonish")
d. *lagungarri* "helping, helpful"	(cf. *lagun* "friend"; *lagundu* "help")
e. *lazgarri* "impressive"	(cf. *latz* "rough"; *laztu* "roughen/impress")

The meaning of these is crystal clear if paraphrased with the corresponding verb:

(6) a. *aberasgarri = (bat edo bat) aberasten duena*	"that it enriches (someone)"
b. *argigarri = (zer edo zer) argitzen duena*	"that it clarifies (something)"
c. *harrigarri = (bat edo bat) harritzen duena*	"that it surprises (someone)"
d. *lagungarri = (bat edo bat) laguntzen duena*	"that helps (someone)"
e. *lazgarri = (bat edo bat) lazten duena*	"that it impresses (someone)"

There is a regular and coherent meaning relation between the verb and the derived adjective.[3] If we take the noun/adjective as the base, the regularity is lost. For the active use, the meaning traditionally associated with the suffix is "producer of"; nonetheless, it is pretty obvious that *lagungarri* doesn't mean "producer of friends" but "helpful"; the meaning of *aberasgarri* is not "producer of rich (people?)" but "enriching"; the meaning of *argigarri* is "clarifying", but nothing implies that light or brightness is produced; *harrigarri* simply means "astonishing, surprising" given that the original connection between *harri* "stone" and *harritu* "astonish" or "get astonished" (literally "stone") is presently lost. Similar considerations apply to *lazgarri*, derived from *laztu* "impress", which can also have the literal meaning "roughen" (cf. *latz* "rough"): the meaning of the derived adjective stems from the meaning of the verb, not the actual meaning of the original adjective.

---

[3] The examples in (5-6) all correspond to the active value of *-garri*, the one that is more problematic for noun/adjective-verb pairs. There are fewer disputed cases with the passive value:

(i) a. *zenbat* "how much/many" / *zenbatu* "count"
b. *zenbakarri* = countable
(ii) a. *txalo* "applause" / *txalotu* "applaud"
b. *txalogarri* = "that it deserves being applauded by someone, worth applauding"

In both cases it is clear that the meaning must be paraphrased departing from the verb.

We safely conclude then that, even in the alleged doubtful cases, *-garri* attaches to verbs and there is no need to duplicate the range of possible bases.

2.2. The morphophonological argument

There exist many pairs of noun/adjective-verb alternations where the verb stem undergoes a phonological change:

(7) verbs that change/e, o, u/ → [a]

noun/adjective	verb
a. *aipu* "citation"	*aipatu* "cite"
b. *akuilu* "spur"	*akuilatu* "incite, spur on"
c. *amorru* "irritation"	*amorratu* "become irritated, irritate"
d. *arbuio* "despise"	*arbuiatu* "despise"
e. *deitore* "regret"	*deitoratu* "regret"
f. *errespetu* "respect"	*errespetatu* "respect"
g. *fresko* "fresh"	*freskatu* "become fresh, refresh"
h. *gomendio* "advice"	*gomendatu* "advise"
i. *gorroto* "hatred"	*gorrotatu* "hate"
j. *gozo* "sweet"	*gozatu* "sweeten"
k. *kutsu* "taint"	*kutsatu* "taint, contaminate"
l. *laudo* "praise"	*laudatu* "praise"
m. *luze* "long"	*luzatu* "lengthen, delay"
n. *oso* "complete"	*osatu* "complete"
o. *susmo* "suspicion"	*susmatu* "suspect"
p. *zoro* "crazy"	*zoratu* "get crazy, drive crazy"

Not surprisingly, the adjectives derived with *-garri* all display the corresponding vowel change, as if derived from the verb and not from a noun/adjective root:[4]

(8) a. *aipagarri* /**aipugarri* "mentionable" b. *akuilagarri* /**akuilugarri* "incentive"
c. *amorragarri*/**amorrugarri* "irritating" d. *arbuiagarri* /**arbuiogarri* "despicable"
e. *deitoragarri* /**deitoregarri* "regrettable" f. *errespetagarri*/ **errespetugarri* "respectable"
g. *freskagarri* /**freskogarri* "refreshing" h. ...

We thus have additional evidence that *-garri* invariably attaches to verbal roots.

---

[4] A possible objection to this argument: the vowel change might be induced by the noun/adjective itself, a phenomenon well-known in Basque noun roots:

(i) a. *asto* "donkey" b. *astakeria* "donkey-nonsense" (cf. also *astokeria*)

Nonetheless, as pointed out in Artiagoitia (1995), these alternations have a lexicalized flavor and need not be respected in modern Basque as long as the suffix is still productive; hence, *astokeria* is also possible. If the vowel change in (8) were induced by the noun/adjective itself, we woud expect pairs like *aipagarri* / *aipugarri*, contrary to fact. Indeed, the few cases where we find an alternation (e.g. *tristegarri* / *tristagarri* "saddening") it is because two forms of the verb (i.e. *tristetu*/*tristatu* "sadden") also exist.

2.3. The modern use of *-garri*

A third and crucial argument comes from the coinage of new words with *-garri*. The spontaneous formation of new words and the judgements of speakers when confronted with possible words show that in order for a new *-garri* word to be derived there must be a verbal root. Below is a list of possible words formed taking nouns with no corresponding verb root as the base:

(9)	**noun/adj.**	**verb**	**derived word**	
a.	*depresio* "depression"	**depresiotu*	**depresiogarri*	cf. *tristagarri* "saddening"
b.	*fede* "faith"	**fedetu*	* *fedegarri*	cf. *susmagarri* "suspect"
c.	*gezur* "lie"	**gezurtu*	**gezurgarri*	cf. *engainagarri* "deceiving"
d.	*ilusio* "illusion"	**ilusiotu*	**ilusiogarri*	cf. *liluragarri* "fascinating"
e.	*lo* "sleep (n)"	**lotu*	**logarri*	cf. *nekagarri* "tiring"
f.	*izpiritu* "spirit"	**izpiritutu*	**izpiritugarri*	cf. *penagarri* "distressing"
g.	*gogo* "mind, will"	**gogotu*	**gogogarri*	cf. *kezkagarri* "worrying"
h.	*hots* "sound"	**hostu*	**hoskarri*	cf. *deigarri* "calling, requiring attention"

The result is that none of these words is acceptable. The reader should be aware that there exist grammatical examples which are apparently similar to the ones rejected by the speakers in that the base *might* be taken to be a noun or an adjective very much like the ones in the first column. But this similarity dissolves very quickly because all of the good examples have a corresponding verbal root:

(9)	**noun/adj.**	**verb**	**derived word**
i.	*triste* "sad"	*tristatu* "get sad, sadden"	*tristagarri* "saddening"
j.	*susmo* "suspicion"	*susmatu* "suspect"	*susmagarri* "suspicious"
k.	*engainu* "lie"	*engainatu* "deceive"	*engainagarri* "deceiving"
l.	*neke* "tiredness"	*nekatu* "get tired, tire"	*nekagarri* "tiring"
m.	*pena* "distress"	*penatu* "distress"	*penagarri* "distressing"
n.	*kezka* "worry"	*kezkatu* "get worried, worry"	*kezkagarri* "worrying"
o.	*dei* "call"	*deitu* "call"	*deigarri* "calling"

Thus, whether a new *-garri* word is possible is a function of the existence of an appropiate verb, not a noun or an adjective. We reach the same conclusion by looking at loanwords. Below I give a list of verbs, many loanwords from Spanish, of the type that generally accept *-garri*; not surprisingly, Basque speakers accept these words. I also provide a list of the corresponding possible noun root, when there is one:

(10)	**noun**	**verb**		**derived word**
a.	*aluzine* "hallucination"	*aluzinatu* "freak out"	→	*aluzinagarri* "hallucinatory"
b.	*depresio* "depression"	*deprimitu* "depress"	→	*deprimigarri* "depressing"

c. *dibersio* "diversion"	*dibertitu* "divert"	→	*dibertigarri* "diverting"
d. *entretenimendu*	*entretenitu* "entertain"	→	*entretenigarri* "entertaining"
e. *estres* "stress"	*estresatu* "stress"	→	*estresagarri* "stressing"
f. *gezur* "lie"	*gezurtatu* "deny"	→	*gezurtagarri* "deniable"
g. *gobernu* "government"	*gobernatu* "govern"	→	*gobernagarri* "governable"
h. --	*kantsatu* "get tired, tire"	→	*kantsagarri* "tiring, tiresome"
i. *konbikzio* "conviction"	*konbentzitu* "convince"	→	*konbentzigarri* "convincing"
j. *sedukzio* "seduction"	*seduzitu* "seduce"	→	*seduzigarri* "seductive"
k. *sorgin* "witch, wizard"	*sorgindu* "bewitch"	→	*sorgingarri* "bewitching"
l. *tolerantzia* "tolerance"	*toleratu* "tolerate"	→	*toleragarri* "tolerable"

It is clear that newly coined *-garri* word always chooses the verbal root.

In view of all the concurring evidence, then, one is led to claim that *-garri* is indeed a deverbal suffix and that there is no evidence to postulate both a verb *and* a noun or adjective subcategorization frame. Hence:

(11) *-garri*, N/Adj, [+V ___]

There is however a small set of exceptions where the root is necessarily a noun. Below I provide a short list of these exceptions:

(12)	**noun/adjective**	**verb**	**derived word**
a.	*azpi* "bed (for animals)"		*azpigarri* "material for making a bed"
b.	*eredu* "model"		*eredugarri* "model"
c.	*interes* "interest"	*interesatu* "interest"	*interesgarri* "interesting"
d.	*onura* "profit"		*onuragarri* "profitable"
e.	*xarmant* "charming"		*xarmantgarri* "charming"
f.	*zirrara* "impact, shock"		*zirraragarri* "shocking"
g.	*barre* "laugh (n)"	*barre egin* "laugh"	*barregarri* "derisive"
h.	*irri* "smile (n)"	*irri egin* "smile"	*irrigarri* "derisive"
i.	*negar* "cry (n)"	*negar egin* "cry"	*negargarri* "deplorable"

There is little doubt about many of the examples in (12): *azpi, eredu, interes, onura,* and *zirrara* are all nouns, and no corresponding verb exists as the possible source for the *-garri* word. The case of the French loanword *xarmant* < *charmant* "charming" is different since the adjective *xarmant* already exists in Basque as a direct loanward; hence *xarmantgarri* is just a pleonastic form, with two suffixes, French *-nt* and Basque *-garri.*

The last three exceptions are set apart: they all involve three nouns that form unergative verbs of the type [noun + *egin* "do"], well known in the literature on Basque (cf. Levin 1983, Laka 1993). For the time being (but see note 15), I will just note that the absence of *egin* in the derived word seems to indicate that the adjectives are derived directly from the noun, the inner argument of the complex verb.

2.4. Only the adjective value is alive

Finally, I turn to consider the categorial status of the suffix: it seems unquestionable that some words are nouns, others are adjectives, and others can be both:

(13) a. *Euskarri sendoa behar du mahai honek* (noun)
holder strong.art need aux table this.E
"This table needs a strong *holder*"

b. *Jokabide lotsagarria izan da zurea* (adjective)
behavior embarrassing.art be aux yours
"Yours has been an embarrasing behaviour"

c1. *Musika lasaigarria maite dut, ez rokanrola* (adjective)
music relaxing.art love aux not rock-n-roll
"I like relaxing music, not rock'n'roll"

c2. *Lorik egin ezean, hobe duzu lasaigarri gogor horietako bat hartzea* (noun)
sleep.prt do if-not better aux tranquilizer strong those.of one take
"If you can't sleep, you'd better take one of those strong tranquilizers"

In short, *euskarri* "holder" is only a noun; *lotsagarri* "embarrassing" only an adjective; and *lasaigarri* can be both a noun "tranquilizer" or an adjective "tranquilizing, relaxing".

There is evidence to claim that the noun value of *-garri* is no longer productive, while the adjective value is pretty much productive. By productive, "we understand the possibility for language users to coin, unintentionally, a number of formations which are in principle uncountable" (Schultink 1961, cited in Lieber 1992: 3). Thus, as the little experiment in (10) indicates, Basque speakers have no trouble in coining new adjectives based on *-garri*; many of the examples in (10) were drawn from articles in the opinion section of the newspaper *Egunkaria*, and most of them are not found in Basque normative dictionaries. Furthermore, if a speaker of Western Basque hears the Eastern adjective *akigarri*, she will have no problem in figuring out what it means if we give her the corresponding Western synomym of *akitu* (i.e. *nekatu* "tire, get tired").

The situation with nominal *-garri* is somewhat different: many speakers don't know what object *berogarri* is ("coat, overall jacket" for those who use it), even though the verb *berotu* "heat" is common to all Basque dialects; similar considerations apply to other possible nouns like *aipagarri* (literally "citer") or *aurkigarri* ("finder"); in fact, speakers' reaction is that they don't know these nouns because they were never confronted with them. In other words, nouns with the suffix *-garri* have to be learned on a one-by-one basis, so it is a very hard task to persuade anyone that an *aipagarri* might be, for example, a CD-rom that carries all kinds of citations by writers and artists.[5] Similarly, the meaning of many *-garri* nouns is often highly idiosyncratic and that is hardly the case with *-garri* adjectives: *pasagarri* (lit. "passer, thing that helps to pass"), as a noun, means "footbridge" in some places but "amount of food needed to

[5] *aurkigarri* is used with the meaning "index" by Etxeberri of Ciboure, a classical Basque writer of the 17th century. Speakers need to memorize this meaning, or else they will think it is an adjective.

survive/pass the winter" in others; according to Azkue (1905), *begiragarri* (lit. "protector") is the word for "preservative" in some dialects, not just protector in general; *luzagarri* "lengthener" is also the word for "excuse, apology" for some speakers, and so on and so forth.

In sum: there is a clear sense that one could in principle gather all the nouns derived with the suffix *-garri*, but there seems to be little sense in doing so with the corresponding adjectives; these are, in principle, uncountable. Having determined that *-garri* is a deverbal suffix which forms adjectives in a productive way, we now turn to the next question: which verbs accept *-garri?* And, what do they have in common?

3. Verbs that can take *-garri*

For ease of exposition, I will start out by separating the so called passive and active values of the suffix just to later show that both values are amenable to a unified analysis.

3.1. Verbs that give rise to the so called passive *-garri*

It is no secret that this passive *-garri* is similar, though not totally equivalent, to its English counterpart *-able*, at least as described in Williams (1981). The verbs that take *-garri* with this sense are simple transitive verbs, usually with an agent-theme argument structure and the corresponding ergative-absolutive pattern. I will call this type the *aipatu* class:

(14) the ***aipatu*** class: simple transitive verbs

aipatu "cite", *agurtu* "greet", *aldatu* "change", *aztertu* "examine", *barkatu* "forgive", *begiratu* "look", *edan* "drink", *egin* "do", *erabili* "use", *entzun* "hear", *eskuratu* "retrieve", *ezagutu* "know", *eztabaidatu* "dispute", *garbitu* "clean", *gogoratu* "remember", *gomendatu* "recommend", *ikusi* "see", *irakurri* "read", *jakin* "know", *jan* "eat", *kontatu* "count, tell", *laudatu* "praise", *onartu* "accept", *oroitu* "remember", *sinetsi* "believe", *susmatu* "suspect", *tolestu* "fold", *txalotu* "applaud", *ukitu* "touch", *ulertu* "understand", *zenbatu* "count"

(15) *Jonek liburua aipatu du* → *Liburua aipagarria da*
.E book.art cite aux book.art citable is
"John mentioned the book" "The book is worth mentioning"

Similarly, pysch verbs of the *subject experiencer* type, the *temere* class of Belleti and Rizzi (1988), may also undergo *-garri* suffixation:

(16) the ***gorrotatu*** class: transitive psych verbs (experiencer = subject)

adoratu "adore", *auhendatu* "regret", *arbuiatu* "despise", *deitoratu* "regret", *desiratu* "desire", *eraman* "suffer", *erdeinatu* "despise", *estimatu* "esteem", *gaitzetsi* "refuse, detest", *gorrotatu* "hate", *higuindu* "abhor", *jasan* "bear", *maitatu* "love", *miretsi* "admire", *pairatu* "suffer"

In Basque, these verbs behave as a regular transitive verbs, with an ergative-absolutive pattern:[6]

(17) *Mirenek John Coltrane miresten du* → *John Coltrane miresgarria zen*
Mary.E admire aux admirable was
"Mary admires John Coltrane" "John Coltrane was admirable

A third class of verbs includes those verbs whose subject is absolutive and are not considered transitive; instead they take an instrumental or dative object:

(18) the ***fidatu*** class: absolutive-quirky case verbs

baliatu "make use of", *fidatu* "trust", *errukitu* "take pity on", *ohartu* "realize, be aware of", *urrikaldu* "take pity on"

(19) a. *Jon Mirenez fidatzen da* → *Miren fidagarria da*
Mary.inst trust aux trustworthy is
"John trust Mary" "Mary is trustworthy"

b. *Jon arazo horri ohartu zaio* → *Arazo hori ohargarria da*
problem that.D realize problem that noticeable is
"John became aware of that problem" "That problem is noticeable"

In a sense, given that the object (not direct object proper) becomes the subject of the derived adjective, I assimilate this use of *-garri* to a subcase of the *passive* value.

3.2. Verbs give rise to the so called active value of *-garri*

The range of verbs that give rise to the active value of the suffix reduces to two blocks: psych verbs of the experiencer-object type (cf. English adjectival *-ing* studied by Brekke 1988), and verbs whose subject may bear an instrumental, non-agentive, θ-role.

The first class corresponds basically to Belleti and Rizzi's *preoccupare* class, which I term with the similar Basque verb *kezkatu*:

---

[6] Some of these subject experiencer verbs, most of them mental states, have a variant with a predicate noun and the verb *ukan* "have" (cf. Etxepare 2001):

(i) *Mirenek Joni {gorroto, beldur, higuina} dio*
Mary.E John.D hatred fear disgust has
Lit: "Mary has hatred/fear/disgust for John"

Other mental states are sometimes expressed with the verb *izan* "be":

(ii) *Miren {beldur, haserre} da*
Mary fear angry is
"Mary is {afraid, angry}"

(20) pysch verbs of the ***kezkatu*** class (experiencer object)

arduratu "become concerned, concern", *aspertu* "get bored, bore", *beldurtu* "become frightened, frighten", *debeiatu* "become/make impatient", *engainatu* "deceive", *entretenitu* "get entertained, entertain", *erakarri* "attract", *ernatu* "get excited, excite", *ernegatu* "get/make desesperate", *erotu* "get/drive crazy", *espantatu* "scare", *estutu* "become distressed, distress", *etsi* "get/make desesperate", *galdu* "harm, corrupt", *gogaitu* "get bored, bore", *gogobete* "satisfy", *harritu* "get astonished, astonish", *harrotu* "become/make arrogant", *haserretu* "get angry, anger", *hunkitu* "get impressed, impress", *ikaratu* "get scared, scare", *iraindu* "insult", *izutu* "get scared, scare", *kaltetu* "harm", *kezkatu* "get worried, worry", *kilikatu* "excite", *laidotu* "offend", *larritu* "become/make anguished", *lasaitu* "get calmed, calm", *laztu* "impress", *liluratu* "become fascinated, fascinate", *lotsatu* "get embarrassed, embarrass", *mindu* "hurt", *nahastu* "get confused, confuse", *nazkatu* "get disgusted, disgust", *nekatu* "get tired, tire", *okaztatu* "disgust", *penatu* "get distressed/distress", *poztu* "become/make happy", *samindu* "afflict", *sortu* "get sedated, sedate", *sumindu* "get enraged, enrage", *sutu* "infuriate", *tentatu* "tempt", *tristatu* "become sad, sadden", *txunditu* "surprise", *zoratu* "get/drive crazy", *zuzpertu* "recover"

(21) *Film horrek Jon aspertu du* → *Filma aspergarria da*
film that.E bore film.art boring is
"That movie bored John" "the movie is boring"

As the reader may see from the example (21), the theme or stimulus argument is marked ergative, the case usually born by the subject in transitive structures in Basque, and the experiencer is marked absolutive.

There is a second, much smaller, class of psych verbs that mirrors the corresponding *piacere* class of Belleti and Rizzi's; here the experiencer is marked dative, and the other argument bears absolutive case:

(22) the ***gustatu*** class (absolutive-dative verbs; dative = experiencer)[7]

damutu "regret", *gustatu, laket* "like"

(23) *Mireni Janis Joplin gustatzen zaio* → *Janis Joplin oso gustagarria zen*
Mary.D like aux very likable was
"Janis Joplin appeals to Miren" "Janis Joplin was very appealing/likable"

It is indeed very much debatable whether the absolutive argument is the subject in the *gustatu* class; nonetheless, given that the tendency among Basque scholars is to

[7] *komeni* "to be convenient" is similar, though it is not considered a psych verb:

(i) *Joni irakurtzea komeni zaio* → *irakurtzea komenigarria da*
.dat reading convene aux reading convenient is
"Reading is convenient for Jon" "Reading is advisable"

consider that the absolutive is the subject, I assimilate this class to the active use of *-garri*, but nothing hinges upon this choice.[8]

The second block of verbs which allow the active value of *-garri* are verbs whose subject bears what appears to be an instrumental role; the verb list below might be amenable to a more subtle analysis. I term these verbs the *erakutsi* "show" class:

(24) the ***erakutsi*** class (transitive verbs with instrumental subjects)[9]

agertu "display", *akuilatu* "spur", *altxatu* "lift", *apaindu* "decorate", *arindu* "thin/lessen", *astundu* "get/make something heavy", *atzeratu* "delay", *aurreratu* "go forward", *azkartu* "speed", *babestu* "protect", *balakatu* "adulate", *begiratu* "free, save", *behaztopatu* "block", *berandutu* "delay", *beratu* "soften", *berotu* "heat", *bete* "fill", *biguindu* "soften", *bizitu* "enliven, vitalize", *bultzatu* "push", *deitu* "call", *deklaratu* "declare" *(des)ohoratu* "(dis)honour" *(des)ugertu* "(dis)oxidize", *edertu* "embellish", *eragin* "influence", *eragotzi* "prevent", *erakutsi* "show", *erdiratu* "tear, break", *estali* "cover", *eutsi* "hold", *ezeztatu* "eliminate", *freskatu* "refresh", *gazitu* "salt", *gehitu* "add", *gelditu* "stop", *gogoratu* "remind", *gorde* "hide", *gozatu* "sweeten", *handitu* "enlarge", *hil* "kill", *hoztu* "cool", *isolatu* "isolate", *ito* "asphyxiate", *itsutu* "blind", *jaso* "raise", *kitatu* "remove a debt", *kutsatu* "taint", *labaindu* "slip", *laburtu* "shorten", *lagundu* "help", *lerdendu* "slenderize", *leundu* "soften", *liraindu* "slenderize", *mehatxatu* "threaten", *ondu* "improve", *oroitu* "remind", *osatu* "complete", *piztu* "switch on, revive", *sendatu* "cure", *sendotu* "strengthen", *trabatu* "block", *txikitu* "diminish", *zabaldu* "widen", *zulatu* "perforate"

(25) *Argazki horrek Irakeko egoera erakusten du* →
picture that.E Iraq.of situation show aux
"That picture shows that situation in Iraq"

Argazki hori Irakeko egoeraren erakusgarria da
picture that Iraq.of situation.gen showing is
"That picture is indicative of the situation in Iraq"

It is particularly revealing to show that in order for the suffix *-garri* to be possible with the verbs in (24), the argument that the suffix controls must be a non-agent subject:

(26) a. *Jonek hutsunea bete du* → **Jon izan da hutsunearen betegarria*
. E gap.art fill aux be aux gap.gen filling.art
"John filled the gap" "John was a filler of the gap"

---

[8] Eventually, what is crucial is that the *-garri* adjective corresponds to the direct object (i.e. the theme in B & R's approach) argument of the verb. See Artiagoitia (2000a: 411), where it is hinted that the dative is indeed the subject.

[9] Some verbs in the list have a psych interpretation by extension: *bizitu* "enliven" but also "encourage, cheer up"; *berotu* "heat" but also "anger"; *erdiratu* "tear, break" and *ito* "asphixiate", which both have a physical and psychological meaning; *piztu* "switch on" but also "revive, encourage". Conversely, the verb *estutu* "distress" in (20) also means "narrow, tighten", but the adjective *estugarri* is used with meaning "distressing" than with the meaning "narrowing". As a lexicalized noun, *estugarri* is "rope, string" in some varieties.

b. *Eztiak edonor betetzen du* → *Eztia janari betegarria da*
honey.E anyone fill aux honey food filling.art is
"Honey fills just anyone" "Honey is a filling food"

(27) a. *Kazetariak Irakeko egoera erakutsi du* →
journalist.E Iraq.of situation show aux
"The journalist showed the situation in Iraq"

**Kazetaria egoeraren erakusgarri izan da*
journalist situation.gen showing be aux
"The journalist was indicative of the situation"

b. *Argazki horrek Irakeko egoera erakusten du* →
picture that.E Iraq.of situatin show aux
"That picture shows the situation in Iraq"

Argazkia egoeraren erakusgarri da
picture situation.gen showing is
"The picture is indicative of the situation"

The word *betegarri* "filling, having the property of filling something" in (26) can only be applied to the instrument that fills the gap, not to the agent of filling.[10] Similar considerations apply to *erakusgarri* "showing, indicative"; it is a property of the instrument that shows or displays something, not a property of the agent that shows or displays that same thing.[11]

3.3. Verbs that give rise to both values of *-garri*

There is a number of verbs that admit *-garri* in the two traditional senses:

(28) *gogoratu* "remember, remind", *gomutatu* "remember, remind", *kontsolatu* "console", *luzatu* "lengthen", *oroitu* "remember, remind", *pasatu* "pass", *seinalatu* "signal", *urratu* "tear"...

But this is simply due to the fact that these verbs can have both a regular transitive use and an instrumental-theme interpretation. Consider the following examples:

---

[10] To refer to the agent, we'd need something like *Jon izan da hutsunearen betetzailea* "John was a filler of the gap". The condition on the suffix *-t(z)aile* is that it must control the external argument of the verb (cf. Azkarate 1995).

[11] Nothing prevents an animate argument from being the subject of a *-garri* adjective, as long as that argument is not an agent:

(i) a. *Familia galdu duen {umeak/umearen egoerak} gerraren zentzugabekeria erakusten du*
"{The kid/the situation of the kid} who lost his family shows the nonsense of the war"
b. *Familia galdu duen {umea/umearen egoera} gerraren zentzugabekeriaren erakusgarri da*
"{The kid/the situation of the kid } who lost his family is indicative of the nonsense of the war"

The kid is definitely not a true agent in (ia), but a sort of stimulus/instrument which shows the nonsense of the war.

(29) a. *Hitz horiek kondenatuko zaituzte zu*
word these condemn aux you
"These words will condemn you"

b. *Hitz horiek izango dira (zure) kondenagarriak*
word these be aux your condemning
"Those words will be condemning (of you)/your condemnation"

(30) a. *Jonek hilketa kondenatu zuen*
.E killing condemn aux
"John condemned the killing"

b. **Jon kondenagarria zen*
condemning was
"John was condemning"

(31) a. Jonek *hilketa kondenatu zuen*
.E killing condemn aux
"John condemned the killing"

b. *Hilketa kondenagarria zen*
killing condemnable was
"The killing was condemnable"

As we can see from the contrast between (29) and (30), the active meaning of *kondenagarri* is restricted to the *instrumental* interpretation of the verb *kondenatu* (cf. 29a-b), but it is impossible with the agentive interpretation (= 30b). On the other hand, the suffix *-garri* is amenable to the passive interpretation since the derived adjective can be predicated of the corresponding direct object in cases like (31b).

3.4. What do all verbs that take *-garri* have in common?

A first look at the types of verbs to which the suffix *-garri* can attach is very telling: all of them take two arguments. We may generalize this fact and propose it as an actual rule:

(32) The suffix *-garri* is restricted to verbs that take two arguments[12]

---

[12] It is tempting to say that the suffix is restricted to verbs that take *two DP arguments*, because verbs which also select sentential complements seem to only take *-garri* provided the DP complement is involved:

(i) a. *Jonek proposamena onartu du* → *Proposamena onargarria da*
"John accepted the proposal" "The proposal is acceptable"
b. *Jonek gaixorik dagoela onartu du* → * *Jon gaixorik dagoela onargarria da*
"John accepted that he is sick" "It is acceptable that John is sick"

Adjectives formed with the active value of *-garri* are incompatible with sentential complements, because the verbs that they are derived from don't accept a sentential subject to start with:

(ii) * *Nola mintzatu zaren lasaitu nau* → * *Lasaigarria da nola mintzatu zaren*
"How you talked calmed me down" "It is relaxing how you talked"

To my mind, the only cases where a sentential complement to a *-garri* adjective is allowed are those with the lexicalized adjectives *ikaragarri/izugarri* literally "frightening"; these adjectives often simply mean "incredible":

(iii) *Ikaragarria da Athleticek zelan irabazi duen partida*
"It is incredible how Athletic won the game"

Another datum that supports the lexicalized flavor of *izugarri/ikaragarri* is that they can be used as adjective/verb degree modifiers:

If (32) is truly part of the rule that forms new words from *-garri,* it predicts that typical monoargumental verbs will be excluded; the prediction is certainly fulfilled:

(33) **afalgarri* "dining/dinable", **borrokagarri* "fighting/fightable", **bukagarri* "finishing/finishable", **dimitigarri* "resigning/resignable", **egongarri* "staying/stayable", **erorgarri* "falling/fallable", **etorgarri* "coming/comeable", **funtzionagarri* "functioning/functionable", **hasigarri* "starting/startable", **ibilgarri* "walking/walkable", **irakingarri* "boiling/boilable", **irtengarri* "exiting/exitable", **izangarri* "being/beable", **jaiogarri* "be borning/bornable", **jardungarri* "engaging/engageable", **joangarri* "going/goable", **jokagarri* "playing/playable", **mintzagarri* "speaking/speakable"...

The list in (33) includes both thematically unaccusative (*bukatu* "finish", *egon* "stay", *erori* "fall", *etorri* "come", *funtzionatu* "function", *hasi* "start", *ibili* "walk, march", *irakin* "boil", *irten* "exit", *izan* "be", *jaio* "be born", *joan* "go") and unergative verbs[13] (*afaldu* "dine", *borrokatu* "fight", *dimititu* "resign", *jardun* "engage", *jokatu* "play", *mintzatu* "speak"), and speakers clearly reject them.[14] Furthermore, when a verb has an inchoative/transitive alternation, the *-garri* word always relates to the transitive use of the verb:

(34)	**inchoative**	**transitive**	**derived word**
a.	*altxatu* "rise"	*altxatu* "raise"	*altxagarri* "leaven" (lit. "raiser")
b.	*gelditu* "stop"	*gelditu* "stop someone"	*geldigarri* "stopping someone"
c.	*hil* "die"	*hil* "kill"	*hilgarri* "mortal, killing"
d.	*ito* "drown"	*ito* "asphyxiate, drown (someone)"	*itogarri* "asphyxiating"
e.	*lotsatu* "get embarrased"	*lotsatu* "embarras"	*lotsagarri* "embarrassing"

(iv) a. *Donostia ikaragarri polita da*
"San Sebastian is absolutely beautiful" (lit. "frightening beautiful")
b. *Bilbo ikaragarri aldatu da azken boladan*
"Bilbao has changed a lot lately" (lit. "has changed frightening lately")

Therefore, leaving aside the special case of *izugarri/ikaragarri,* it seems that *-garri* adjectives are possible with verbs that take two arguments, but those arguments need to be DPs, not sentences.

[13] I follow a thematic criterion when classifying verbs: *funtzionatu* "function", *irakin* "boil", and *irten* "get out, exit" select the transitive auxiliary (*irten* only in Western Basque); the verbs *borrokatu* "fight", and *jokatu* "play" may choose the transitive and intransitive auxiliary depending on the dialect.

[14] In the Basque General Dictionary, one can find the following words: *erorgarri* "property of making one fall", *joangarri* "reason to go", *etorgarri* "origin", *irakingarri* "property of making something boil", which are nonetheless rejected by all my informants. All of them are creations of lexicologists, and have rarely been used in the history of written Basque; it is remarkable, however, that all these creations imply a causative/transitive and, hence, biargumental use of the verbs, even though the relevant transitive is non-existent. For example, *erorgarri* means "faller = that makes one fall", *joangarri* "something that makes one go" and so on, but the verbs *erori* and *joan* cannot be used transitively.

A second prediction of (33) is that verbs which typically take three arguments will be at odds with the suffix *-garri*; again, the prediction is borne by the data; speakers find the relevant examples ackward or sharply reject them:

(35) ??*erosgarri* "buyable", ??*salgarri* "sellable", ??*esangarri* "sayable/saying", **eskagarri* "askable", **eskaingarri* "offerable", **ipingarri* "putable", **jasogarri* "receivable/ receiving", ??*kengarri* "removable", **uzgarri* "lendable"...

In view of all this evidence, I conclude that the suffix *-garri* can only attach to verbal roots that take two arguments, and this salient feature is precisely what both uses of *-garri* share, superficially at least.[15]

We now try to go a step forward: is there any other coincidence between the two types of verbs that take *-garri* beyond their taking two arguments? Or are we doomed to propose that there are basically two rules for the suffix? There are good *a priori* reasons to reject this latter possition: first, traditional Basque grammarians always treat *-garri* as one single suffix, not two; the vowel change detected in denonimal verbs that take *-garri*

---

[15] The realization that *-garri* attaches to two-argument verbs can help us do away with at least three of the apparent exceptions in (12), that is to say *barregarri, irrigarri* and *negargarri* (=12g-h-i). These examples are special for two reasons: (a) the base seems to be a noun; (b) even if the verb is taken as the base for the adjective, the interpretation seems to require a causative-like argument apparently absent in the base: something *barregarri* "derisive" would be something that makes one laugh; but that causer is not an argument of *barre egin* "laugh"; I said this much when I analyzed these adjectives as exceptions in Artiagoitia (1995: 367). There is, however, a way to derive these isolated examples from the general pattern of *-garri*. Interestingly, these three verbs which accept *-garri*, i.e. *irri egin, barre egin* and *negar egin*, have something in common; they can all take a dative argument:

(i) a. *Zeri egiten diozu barre?*
what.D do aux laugh
"What are you laughing at?"

b. *Zure ateraldiari irri egin zioten*
your story.D smile do aux
"They laughed at your story"

c1. *Andreek euren bekatuei, eta euren semeenei negar egin behar zieten* (Añibarro)
women.E their sin.D and their son.pl.gen.dat cry do have aux
"Women had to cry for their sins and their sons' (sins)"

c2. *Bazuen zeri negar egin bai, gaixuak*
BA.had what.D cry do yes poor
"Indeed, the poor guy did have what to cry for"

And, as a matter of fact, it is this dative argument the one that the *-garri* adjective is predicated of: something *barregarri* "derisive, laughable" or *irrigarri* "derisive" is such that one laughs or smiles at it (marked dative); something *negargarri* "deplorable" is such that one deplores or cries for it (marked dative). If we assume that the nouns *barre/irri/negar* are part of the complex verb, the verb simply has two arguments: the subject-agent, marked ergative, and the goal-source (?), marked dative. Therefore, these three words turn out to be a subcase of the passive use of *-garri*.

One question remains: why is the verb *egin* "do" absent from the derived adjective? In fact, this is strictly incorrect: there exist less frequent variants of the adjectives with the correspondig verb base: *barre egingarri, irri egingarri* and *negar egingarri*. Nonetheless, as pointed out in Artiagoitia (2000b) and Oyharçabal (2003a), to the extent that [noun + *egin* "do"] unergative verbs participate in derivational processes, the absence of the verb part in the derived form is a constant. In Artiagoitia (2000b), I speculate that *egin* simply fills up an empty V node in the syntax (i.e. computational system); but these unergative verbs may well have their V node empty lexicon-internally. Cf. also Gràcia et al. (2000).

affects both values of the suffix; thirdly, both values of the suffix are equally productive in present Basque.

My position is that there is a stronger coincidence between the two uses of the verb: in fact, a closer look at the verbs that give rise to the active value of *-garri* shows that the *kezkatu* and *gustatu* classes in (20) and (22) correspond almost exactly to Belleti and Rizzi's *preoccupare* and *piacere* classes. Furthermore, these authors analyze these two verb classes as fundamentally *unaccusative* verbs in that they have two internal arguments. This is crucial; if Belleti and Rizzi are on the right track and if the corresponding Basque verbs display the same characteristics as their Italian counterparts, there is indeed a more subtle connection between the two uses of *-garri:* the suffix affects the most internal argument of two-argument verb to make it the subject/ external argument of the derived active. I illustrate this subtle connection between the two uses of *-garri* with the following examples:

(36) passive value of *-garri*: the suffix *affects* the direct object of the base verb

a. *Jonek artikulua aipatu du* → *Artikulua aipagarria da*
John.E article.art cite aux article.art mentionable is
"Jonek cited the article" "The article was mentionable"

b. *Jonek artikulua gorrotatzen du* → *Artikulua gorrotagarria da*
John.E article.art hate aux article.art detestable is
"John hates the article" "The article is detestable"

(37) active value of *-garri*: the suffix *affects* the superficial subject of the verb

a. *Artikuluak Jon hunkitu du* → *Artikulua oso hunkigarria da*
article.E move aux article.art very moving is
"The article touched John" "The article is very touching"

b. *Artikulua Joni gustatu zaio*[16] → *Artikulua gustagarria da*
article.art John.D appeal aux article.art appealing is
"The article appeals to John" "The article is appealing"

In (36) there is little doubt about the connection between the adjective and the verb: the suffix creates a new adjective that is predicated of the internal argument of the verb base. In (37) the suffix affects the superficial subject of the verb; but, crucially, if one adopts Belleti and Rizzi's analysis for verbs like *hunkitu* and *gustatu* we have an initial syntactic derivation like the following:

(38) a. $[_{VP}$ *Jon* ***artikuluak*** *hunkitu*] → ***Artikulua*** *hunkigarria da*
b. $[_{VP}$ *Joni* ***artikulua*** *gustatu*] → ***Artikulua*** *gustagarria da*

---

[16] See note 8 above. Again, personally I don't think the absolutive argument is the surface subject, I just follow the traditional Basque view; if the dative argument is the subject, then the analysis is simpler, for the absolutive is the deep *and* surface direct object.

In (38) the argument affected by the suffix *-garri* is again the deepest or innermost argument of a two-argument verb. They should be compared to the corresponding syntactic representations of the regular transitive vebs that can take *-garri:*

(39) a. [$_{vP}$ *Jonek* [$_{VP}$ ***artikulua*** *aipatu*]] → ***Artikulua*** *aipagarria da*
b. [$_{vP}$ *Jonek* [$_{VP}$ ***artikulua*** *gorrotatu*]] → ***Artikulua*** *gorrotagarria da*

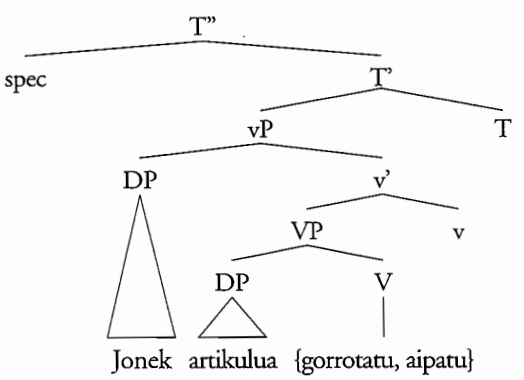
The connection is evident: in both cases, the suffix is making the (most) internal argument (the so called *theme* argument) and making it the external argument of the adjective. Viewed from the other angle, *-garri* is precisely the kind of suffix Belleti and Rizzi's account predicts could exist since it treats the stimulus, i.e. their *theme*, argument of *preoccupare* and *piacere* verbs, indeed the deep direct object in their proposal, in a way similar to the direct object of a regular transitive verb.

In the next section, I will simply make the case to persuade the reader that the verbs of the *kezkatu* and *gustatu* classes do indeed have the same properties that verb classes studied by Belleti and Rizzi have; furthermore, I will also try to show that the *erakutsi* class may be amenable to the same analysis. A note of clarification is in order: many linguists have criticized several aspects of Belleti and Rizzi's account of psych verbs; we can cite, among others, Grimshaw (1990), Ruwet (1993), Pesetsky (1995), Arad (1999a, 1999b). It is not my intention to compare and weigh all theories regarding psych verbs, but rather present a possible line of explanation for the rule of *-garri* suffixation based on the hypothesis that the suffix is one and the same in both its active and passive uses. It will be in section 5 when I shortly review possible accounts of the suffix *-garri;* in any case, my assumption is that there is only one *-garri* and that some version of Belleti and Rizzi's original insight must be correct.

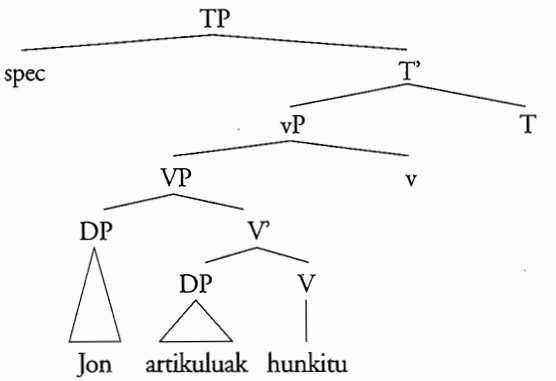
4. The *kezkatu* class equals Rizzi and Belleti's *preoccupare* verbs

Belleti and Rizzi (1988) give a battery of arguments to show that the *preoccupare* verbs are fundamentally unaccusative and lack a true external argument despite their transitive appearence; this is in contrast with the corresponding transitive verbs, among which they include the *temere* class. Adapting the analysis of both classes of verbs to Basque using Chomsky's (1995) clausal structure, we obtain something like the following D-Structure trees for *aipatu* and *gorrotatu*, and for *kezkatu*, respectively:

(40)



(41)



Whereas (40) corresponds to a regular transitive structure, the diagram tree in (41) simply tries to reflect the fact that both the experiencer and the theme argument are internal to VP; according to (and adapting) Belleti and Rizzi's original proposal, the lexical entry of the *kezkatu* class should include something like the specification of the lexical or inherent case of one of the arguments.[17] Hence:

(42) *kezkatu*: experiencer, theme
↕
ERG

This inherent specification prevents the projection of a true external argument, according to Belleti and Rizzi. Now, what kind of evidence is there to prove the correctness of (41)? Belleti and Rizzi give several arguments to show (a) that the subject of *preoccupare* class is a derived subject; (b) that the surface object is not a truly direct object of the verb; and (c) that unusual binding properties follow from a configuration similar to (41). The three kinds of arguments are replicated below for *kezkatu* verbs.

4.1. The subject of *kezkatu* verbs is a derived subject

Belleti and Rizzi provide five arguments to claim that subjects of *preoccupare* are derived subjects. I reproduce or adapt most of them to Basque. Incompatibility with the arbitrary interpretation of third person plural subjects, a property also shared by unaccusative verbs, is one of the clearest tests where *kezkatu* and *preoccupare* verbs go hand in hand:

[17] In Belletti and Rizzi's view, the experiencer is assigned inherent accusative case; from the point of view of the Obligatory Case Parameter accusative case is the marked, non-default case, required to mark a second argument. Basque is an ergative language, so the non-default case is ergative rather than accusative; thus the right adaptation of Belletti and Rizzi's approach to Basque would require that the theme be marked with inherent ergative case.

(43) a. ***pro*** *etxera deitu dute (Jon izan da, bai)*
3pl home call aux be aux yes
"They called home (It was John, yes)"

b. *****pro*** *berandu iritsi dira afaltzera (Jon izan da, bai)*
3pl late arrive aux dinner.for
"They arrived late for dinner (It was John, yes)"

c. *****pro*** *Jon kezkatzen dute (baliteke alaba izatea)*
3pl worry aux could daughter be
"They worry John (It could be his daughter)"

Belleti and Rizzi attribute this property to the theta-marking properties of INF; in current terms, this should attributed to the theta-marking properties of little v, which projects no specifier in (41) at D-Structure.[18] *Kezkatu* verbs also resist the causative construction, a test generally given as a diagnostic for raising verbs:

(44) a. * *Mirenek artikuluari Jon hunkiarazi dio*
Mary.E article.D impress.make aux
"Mary made the article move John"

b. * *Mirenek Olentzerori Jon hunkiarazi dio*
Mary.E SantaClaus.D impress.make aux
"Mary made Santa Claus move John"

As the examples show, a verb of the *kezkatu* class such as *hunkitu* "move, impress" is incompatible with the bound causative verb *arazi*.[19]

A third argument for the lack of external argument of *kezkatu* verbs comes from reflexivization: Basque lacks anaphoric cliticization, ungrammatical with *preoccupare* and raising verbs in general according to Belleti and Rizzi, but has a detransitivization strategy in order to form reflexive structures; in this detransitivization strategy the verb selects the auxiliary verb *izan* "be":

---

[18] Pesetsky (1995: 37ff) calls this third person plural existential interpretation *corporate* interpretation and argues it should be regarded as a test for agentivity. Personally, I don't exactly think his corporate interpretation and Belletti and Rizzi's original interpretation are the same thing. In any case, Basque non-agentive subjects do behave as argued by Belletti and Rizzi:

(i) *Telegrama bitxi bat jaso dute bulegoan; ugazabak berak jaso du*
strange one get aux office.loc boss.E he.E get aux
"They received a strange telegram at the office; it was the boss in person who received it"

The plural arbitrary subject in (i) is a goal argument, and need not be a true plural.

[19] As in Italian, the causative construction is possible under the agentive interpretation:

(i) *Zuzendariak aktoreari beste aktorea hunkiarazi dio*
"T he director made the actor impress the other actor"

But in this case we would have a true external argument as the subject of *hunkitu*.

(45) a. *Jonek ispiluan bere burua ikusi du*
E mirror.loc his head see aux=has
"Johns has seem himself in the mirror"

b. *Jon ispiluan ikusi da*
mirror.loc see aux=is
"John has seen himself in the mirror"

Ortiz de Urbina (1989) accounts for this kind of reflexives by proposing a rule of lexical binding whereby the external θ-role is assigned the φ-features of the internal argument lexicon-internally and, hence, the former need not project in the syntax; only the internal argument is projected. Interestingly, *kezkatu* verbs can't form detransitivized reflexives:

(46) * *Jon kezkatzen da* (= ok under the interpretation "John worries")
"John worries himself"

The ungrammaticality of (46) is expected if *kezkatu* verbs lack an external argument in the first place; there is no argument *higher* than the experiencer argument within the lexical entry that could bind it.[20] This account is very close to Grimshaw's explanation of the Italian data.

Finally, the Basque version of a marginal argument that Belleti and Rizzi mention in a footnote is interesting: *kezkatu* verbs are incompatible with impersonal constructions, just like unaccusatives:

(47) a. *Atzo errakuntza asko egin ziren partidan*
yesterday error many do aux game.loc
"Yesterdary many errors were made during the match"

---

[20] Indeed, the usual reflexive construction with the pronoun *bere burua* (literally "his/her head") is highly marginal for *kezkatu* verbs, just like it is for *preoccupare* verbs:

(i) ?? *Jonek bere burua lotsatzen du*
.E his head.art embarrass aux
"John embarrasses himself"

And only improves when the agentive interpretation is forced:

(ii) *Jonek nahita lotsatzen du bere burua*
.E deliberately embarras aux his head.art
"John embarrasses himself deliberately"

See Belleti and Rizzi (1988) and Grimshaw (1990: 158ff) for two possible explanations of this peculiar behavior.

[21] Unaccusatives accept generic impersonals:

(i) *Hemen berandu iristen da afaltzera*
here late arrive aux dinner.for
"Here one arrives late for dinner"

In Albizu's (in press) recent study of Basque impersonals and detransitivized reflexive structures, the analysis of impersonal constructions of transitive verbs always require an external PRO argument in the specifier of little v. If *kezkatu* verbs have the specifier of v empty, it comes as no surprise that they won't accept either structure.

b. * *Atzo berandu iritsi zen afaltzera*[21]
yesterday late arrive aux dinner.for
"Yesterday it was arrived late for dinner"

c. * *Atzo ardiak beldurtu ziren*
yesterday sheep.art frighten aux
"Yesterday sheep were frightenened"

Example (47c) is only grammatical under the reading "yesterday the sheep got scared", but cannot mean "the sheep were frightened, someone frightened the sheep". Consequently, most if not all,[22] the arguments given by Belleti and Rizzi to show that the subject of *preoccupare* is not a deep subject and should not be treated as an external argument are more or less replicated for the subject of Basque *kezkatu* verbs, one of the verb class that gives rise to the active value of *-garri*. Therefore, it appears that *kezkatu* verbs are typologically similar to *preoccupare* verbs.

4.2. The object of *kezkatu* is a syntactic island

Belleti and Rizzi show that the surface object of *preoccupare* verbs, the experiencer argument, is not a sister to the verb in that it behaves as a syntactic island for extraction. Their examples are relative clauses involving extraction of genitive wh-phrases from the object/experiencer position. Basque doesn't allow extraction/relativization of genitives from DPs, but resumptive pronouns help to improve those examples:

(48) a. *[*Jonek* [e *aita*] *gorrotatzen duen*] *neska*
John.E father hate aux.comp girl
"The girl whose father John hates" lit.: "the girl that John hates father"

b. [*Jonek* [**beraren** *aita*] *gorrotatzen duen*] *neska*
John.E her father hate aux.comp girl
"The girl that John hates her father"

---

[22] Belletti and Rizzi also mention lack of passivization amongst the properties of *preoccupare* verbs, a property challenged by Pesetsky (1995: 21ff), not convincingly in my opinion. It is doubtful whether Basque has a true passive construction. In any case, it seems that *kezkatu* verbs are highly marginal in the alleged passive construction:

(i) a. *Liburu hau Atxagak idatzia da*
book this .E written.art is
"This book is written by Atxaga"

b. *Aresti jende askok miretsia eta gorrotatua izan zen*
people many.E admire.art and hate.art be aux
"Aresti was admired and hated by many people"

c. ?? *Ikusleak Almodovarren azken filmak hunkituak izan ziren*
viewer.E .gen last film.art.E moved.art be aux
"Spectators were moved by Almodovar's last movie"

However, examples are ungrammatical when extraction is out of the object of a verb belonging to the *kezkatu* class, whether there is a resumptive pronoun or not:

(49) a. * [*ekonomiak* [**e** *aita*] *kezkatzen duen*] *neska*
economy.E father worry aux.comp girl
"The girl whose father the economy worries"

b. * [*ekonomiak* [**beraren** *aita*] *kezkatzen duen*] *neska*
economy.E her father worry aux.comp girl
"The girl that the economy worries her father"

Regardless of the proper account of this restriction (L-marking, lexical government, or head-to head feature checking), it seems again that the parallelism between (the surface objects of) *preoccupare* and *kezkatu* verbs is well founded.

There is a second, Basque-specific, piece of data that suggests that the surface object of *kezkatu* is not a true object: the impossibility of existentical interpretation. As observed by Laka (1993) and Artiagoitia (2002), Basque plural DPs allow an existential interpretation equivalent to Romance bare DPs only in internal argument position:

(50) a. ***Irakasleak*** *bileran izan dira*
teacher.art meeting.loc be aux
"(The) teachers were present at the meeting"

b. *Errektoreak* ***irakasleak*** *zigortu ditu*
provost.E teacher.art punish aux
"The provost punished (the) teachers"

c. ***Irakasleek*** *telefonoz deitu dute*
teacher.E phone.by call aux
"The teachers/?? teachers phoned"

d. ***Irakasleak*** *bileran mintzatu dira*
teacher.art meeting.loc talk aux
"The teachers/?? teachers talked at the meeting"

The DP *irakasleak* may have a definite ("the teachers") or existential ("(some) teachers") interpretation in (50a) and (50b) where it is the subject of an unaccusative or the object of a transitive verb, respectively. In (50c), we find the same DP bearing ergative case as the subject of the unergative verb *deitu* "call" and we only obtain a definite interpretation; we have a similar situation in (50d) with the verb *mintzatu* "speak", except that in this case this verb doesn't require ergative case for its subject. The contrast between the two interpretations becomes more evident once the tag "but we don't know which ones" is added, because this forces the existential interpretation:

(51) a. ***Irakasleak*** *pasatu dira baina ez dakigu zeintzuk*
teacher.art go-by aux but not know which
"Teachers went by but we don't know which ones"

b. ?? ***Irakasleak*** *mintzatu dira baina ez dakigu zeintzuk*
teacher.art speak aux but not know which
"Teachers spoke but we don't know which ones"

When we turn to the surface objects of the *kezkatu* class, we see that these don't behave as true direct objects:

(52) a. *Ikasketa-plan berriak* ***irakasleak*** *zoratu ditu* (?? *baina ez dakigu zeintzuk*)
study-plan new.E teacher.art drive-crazy aux
"The new plan has driven all teachers / ??(some) teachers crazy (but we don't know which ones)"

b. *Zurrumurruek* ***irakasleak*** *kezkatzen dituzte* (?? *baina ez dakigu zeintzuk*)
rumours.E teacher.art worry aux
"The rumours worry the teachers / ??(some) teachers (but we don't know which ones)"

In other words, the interpretation of (52a) and (52b) cannot be "there exist teachers that the new plan drove crazy" or "there are some teachers that the rumours worry". Regardless of how this restriction is handled in the grammar (cf. Longobardi 1994, 2000), we've got a case for not taking the superficial object of *kezkatu* as a *bona fide* direct object.

4.3. Binding Theory is peculiar with *kezkatu* verbs

The Basque *kezkatu* verbs replicate the usual peculiar binding data with psych verbs in that a reflexive inside the subject position appears to be bound by the object:

(53) a. *Ni, neure buruak nazkatzen nau*
I my head.art disgust aux
"As for me, my own self disgusts me"

b. *Bere buruaren irudiak Jon asko kezkatzen du*
his head.gen picture.E much worry aux
"The image of himself worries John a lot"

Basque reflexive anaphors mean literally "X's head", so (53a) appears to mean "my head (i.e. *myself*) disgusts me", with the entire reflexive pronoun as the surface subject. As Rebuschi (1993) has shown, however, in Basque reflexives it is the possessor part of the "x's head" expression the one that should really count as truly reflexive. The data are similar with reciprocals: reciprocal subjects cannot be bound by the object, but subjects containing a reciprocal pronoun can:

(54) a. **Elkarrek nazkatzen gaitu*
each other.E disgust aux
"Each other disgusts us"

b. *Elkarren inguruko istorioek nazkatu egin gaituzte*
each other.gen about story.E disgust do aux
"Stories about each other have made us sick"

Object agreement in the verb (prefix *g-*) marks that the absolutive argument is first person plural *gu* "us". Following Belleti and Rizzi's original insight, one can in principle account for these peculiar binding theory facts if the diagram in (41) is correct; at some point of the derivation (e.g. at D-Structure or at LF via reconstruction), the surface object, the absolutive argument, c-commands the anaphor inside the ergative argument.

4.4. Are *erakutsi* verbs similar to *kezkatu/preoccupare* verbs?

The verbs which belong to the *erakutsi* class also give rise to the active value of the suffix *-garri*. It would be very promising if these also displayed some or all of the properties of the *kezkatu* class. In which follows, I offer a preliminary sketch of their properties; this sketch suggests that these verbs may lack an external argument too, although it is not all clear whether the instrumental argument is the closest internal argument of the verb. First of all, the instrumental, non-agent, interpretation of these verbs is incompatible with the third person plural arbitrary interpretation:

(55) a. *Arrazoia duzula erakutsi didate* (**Jaurlaritzaren inkesta izan da*)
reason.art have.comp show aux government.gen survey be aux
"They showed that I am right (it was the government's survey)"

b. *Ingelesa ikastera bultzatu naute* (**azken bidaia izan da*)
English.art learn.to push aux last trip be aux
"They encouraged me to learn English (it was the last trip)"

They are also incompatible with the factive construction:

(56) a. **Ugazabak lantegiko giroari beharginak itoarazi dizkio*
boss.E factory atmosphere.D worker.art asphyxiate.make aux
"The boss made the atmosphere asphyxiate the workers"

b. **Amaiak azken bidaiari Jon ingelesa ikastera bultzarazi dio*
Amaia.E last trip.D English.art learn.to push.make aux
"Amaia made the last trip encourage John to learn English"

Thirdly, *erakutsi* verbs are generally possible in impersonal, detransitivized, structures, but the understood argument cannot be instrumental, the one that the *-garri* adjective is predicated of:

(57) *Horrelako proiektuak beti bultzatu dira sail honetan;*
that.like project.art always push aux department that.loc
"Projects like that have always been supported in this department"

a. *?*Izan ere, dirulaguntza deialdiak bultzatu ditu*
be even fellowship call.E push aux
"As a matter of fact, the fellowship program has supported them"

b. *Izan ere, sailburuak berak bultzatu ditu*
be even director.E he.E push aux
"As a matter of fact, the director himself has supported them"

The impersonal construction can only be followed by the tag that mentions the agent, the external argument, not the instrumental. Finally, the peculiar binding facts typical of psych verbs are also found with the subjects of *kezkatu* verbs:

(58) a. *Horrela segituz gero, neure buruak hilko nau!*
so keep if my head.E kill aux
"If I keep like this, my own self will kill me!"

b. *Estresa dela eta, neure buruak asko itotzen nau azken boladan*
stress is.comp due my head.E much drown aux last occasion.loc
"Due to stress, my own self asphyxiates me a lot lately"

The interpretation of the examples is never agentive: (57a) means something like "my psyche, my worries will kill me", not "I will kill myself"; similarly (57b) means that "my own worries, etc., depress or asphyxiate me", rather than "I asphyxiate myself".

With respect to extraction out of the object of the *erakutsi* class, the data are less clarifying; there is a contrast between the agentive and the non-agentive uses of the verb when relativizing the object:

(59) a. *gaizkileak beraren alaba hil zuen andrea*
bandit.E she.gen daughter kill aux woman.art
"the woman that the bandit killed her daugther"

b. ?*kutsadurak beraren alaba hil zuen andrea*
pollution.E she.gen daughter kill aux daughter.art
"the woman that pollution killed her daughter"

But the contrast is not as robust as in examples (49) above.

In sum, there is some empirical base to conjecture that the *erakutsi* class lacks a truly external argument and has a derived subject, a subject that starts out as (perhaps) the most internal argument of the verb. If correct, this conjecture implies that there is a unifying factor in all uses of the suffix *-garri*: this suffix makes what appears to be the deep direct object of the verb and makes it the external argument of the derived adjective. The existence of *-garri*, then, turns out to be a surprising and independent confirmation of the correctness of Belleti and Rizzi's analysis of the *preoccupare* verb class.

5. Theories that can handle *-garri*

I review several theories that handle the relation between argument structure and morphology (Williams, Grimshaw) or theories that deal with the lexical structure of psych verbs (Pesetsky, Arad).

5.1. Relevance of *-garri* for Williams' thematic constancy

One interesting property of *-garri* is that it appears to be similar to English *-able*, as described in Williams (1981). In that article, Williams defends the view that derivational affixes operate on argument structures, understood as a list of thematic roles; these

affixes either internalize the original external argument of the verb or else introduce a new external argument. English *-able* is, according to him, best defined as a rule of theme-externalization:

(60) a. *read* (agent, theme) → *readable* (agent, theme)
b. *John read the book* → *the book is readable* (book = theme)

The rule, so understood, predicts that unaccusative verbs should take the suffix, a prediction that is fulfilled according to Williams given the existence of words like *perishable*. The Basque suffix *-garri* resembles *-able* in that it does indeed seem to externalize the theme argument of regular transitive verbs, but it fails to attach to unaccusative verbs (cf. examples in (33) above). This failure speaks against the operativity of thematic role labels *per se* in derivational morphology, a point already made in Levin and Rappaport (1986). What is more, the Basque suffix *-garri* also affects what (according to Belleti and Rizzi) appears to be the theme argument of the equivalent of English *frighten* and Italian *preoccupare* verbs, namely *kezkatu* verbs. If thematic roles are constant for the two or three languages, we would expect that *-able* will also affect *frighten* verbs, but it does not:

(61) *frigthen* (experiencer, theme) → **frigthenable*
(cf. *kezkagarri* = frightening)

On the other hand, *-garri* can also affect or externalize non-theme, instrumental, arguments:

(62) a. *ito* (instrumental, theme/experiencer) "asphyxiate" →
itogarri (instrumental...) "asphyxiating"
b. *babestu* (instrumental, theme) "protect" →
babesgarri (instrumental...) "protecting"

In short, claimimg that the rule of *-able/-garri* suffixation affects a particular theta-role (e.g. theme) is too strong in that verbs which appear to have a theme are not affected by the rule in English (cf. 61) or in Basque (cf. 33); it is also too weak, given that theta-roles other than the theme may be affected by the rule. Thus, the mere existence of suffixes like *-garri* helps to clarify that thematic constancy cannot be the right approach to derivational morphology processes.

5.2. Grimshaw's argument structure approach

Although her analysis departs in several aspects,[23] Grimshaw (1990) agrees with Belleti and Rizzi (1988) that verbs of the *preoccupare/kezkatu* and *piacere/gustatu* classes

[23] For Grimshaw, *preoccupare/frigthen* verbs do have D-Structure subjects even though they lack an external argument. Her explanation for the peculiar binding effects is also different (cf. Grimshaw 1990: ch. 5).

fundamentally lack an external argument. For the former class, she proposes that the realization of the theme, more internal, argument of the verb as the subject is a direct consequence of its being more prominent in the aspectual dimension:

(63) *frighten/preoccupare*	(x	(y))	[argument level]
	2	1	[aspectual level]

In other words, *frigthen* is a psychological causative verb, but it is the most internal argument the one that is associated with causation; the more prominent argument, the experiencer, is the one that undergoes the change. As for verbs of the *fear/temere* class, Grimshaw assimilates them to regular transitive verbs; on the other hand, she leaves open the treatment of Belleti and Rizzi's (1988) *piacere* class, but hints that they are perfectly amenable to an unaccusative analysis, given auxiliary choice and crosslinguistic evidence:

(64) a. *fear/temere*	(x	(y))	b. *piacere*	((x	(y)))
	1	2			|
					dative

In sum, except for the *fear/temere* class where experiencer behaves as a true external argument, psych verbs of the *frigthen/preoccupare* and *piacere* classes lack an external argument.

Grimshaw's treatment of psych verbs' argument structures allows for a straight unification of the two uses of *-garri*. The so called passive value would be formulated as follows:

(65) passive *-garri*

a. (x (y))	→	(R<=y> (x-ø (y))
V	→	Adj
b1. *aipatu*	→	*aipagarri* "mentionable, worth mentioning"
b2. *gorrotatu*	→	*gorrotagarri* "detestable"

In other words, and leaving the representation of meaning aside, the suffix introduces an external argument, call it R,[24] that controls the internal argument of a verb that takes two arguments, one of which is external. The original external argument is erased.

The rule of active would look like this:[25]

(66) active *-garri*

a. ((x (y)))	→	(R<=y> (x (y))
V	→	Adj
b1. *kezkatu*	→	*kezkagarri* "worrying, worrisome"
b2. *erakutsi*	→	*erakusgarri* "showing"

---

[24] "R" corresponds to Williams' original *Referential* role of nouns, also present in their predicative use. Grimshaw extends it to predicative adjectives; I follow her practice.

[25] I asssume that *erakutsi* verbs are analized just like *frighten/preoccupare/kezkatu* verbs:

(i) *erakutsi*	((x	(y)))	
	2	1	where x = theme; y = instrumental

That is, the suffix introduces an external argument, call it R, that controls the most internal argument of a verb that takes two arguments; since there is no external argument proper, no argument is eliminated.

The two rules can in fact be united; Grimshaw uses two different notations (p. 28 *vs* 41) for *frighten*-type verbs; in one she uses single brackets but claims that the aspectual dimension is sufficient to ensure that the "x" (i.e. "experiencer") argument doesn't count as external; in the other she uses double brackets to make the point that there is no external argument. If one chooses the first notation, the possibility to unite (65) and (66) above in an elegant way is self-evident:

(67) *-garri* suffixation (final version)
(x (y)) → (R <=y> (x (y)))
where x → ø iff x is external argument

If the other notation is used, the rule should include a bracket to indicate the optionality of an external argument:

(67') *-garri* suffixation (alternative final version)
([]x (y[])) → (R <=y> (x (y)))
where x → ø iff x is external argument

This rule also allows for a clear distinction between English *-able* and Basque *-garri*; in English the "x" argument must be external for the rule to apply (cf. English *-er*, which also needs to refer to external arguments according to Levin and Rappaport 1988);[26] in Basque it need not be so and, hence, verbs with two internal arguments are also affected by the rule.

Grimshaw's approach also predicts that theta-marking of the outer "x" argument will be banned when "x" is external and erased, but not when "x" is internal:

(68) a. **Artikuluak egile-aipagarriak dira*
article.art author-mentionable.art are
"The articles are author-mentionable"

b. **Artikuluak egilearen aipagarriak dira*
article.art author.gen mentionable.art are
"The articles are author's mentionable"

c. *Ikuskizuna begi-mingarria zitzaion*
spectacle.art eye-hurting.art was
"The spectacle was eye-hurting"

d. *Profetaren hitz horiek aukeratuen seduzigarri gertatu dira*
prophet.gen word those elected.gen seducing result aux
"Those words by the prophet turned out to be seducing of the elected ones"

---

[26] I am assuming, *contra* Williams, that unaccusatives verbs take *-able* only marginally.

As examples (68a-b) show, there is no way for the *-garri* adjective to theta-mark the original external argument either via a synthetic compound or a genitive phrase; on the other hand, the "x" argument, if not external, is now more deeply embedded than "y" and thus can be theta-marked both in a synthetic compound and thru a genitive phrase.[27] In sum, Grimshaw's argument structure analysis of psych verbs gives an elegant account the rule of *-garri* suffixation.[28]

5.3. Pesetsky's zero syntax and the suffix *-garri*

Though briefly and schematically, I would like to make a few remarks on Pesetsky's approach to the *frighten* verb class.[29] Pesetsky's analysis is based on two key assumptions: (a) that the theme argument of Belleti and Rizzi masks two or three different theta-roles, *causer* and *subject/target of emotion*, the former being hierarchically superior to experiencers; and (b) that *frigthen-preoccupare* verbs are bimorphemic causative verbs. Simplifying matters, Pesetsky proposes that *frighten* verbs derive from intransitive roots of the type √*frighten*; this verbal root, usually not realized as an actual verb in English, takes two internal arguments, both realized within a PP cascade; the experiencer is projected as the specifier of a PP headed by the empty P CAUSER, and the causer is realized as the complement of the empty P. In the course of the derivation, the empty P incorporates to the verbal root to check its strong morphological feature and gives rise to the actual causative verb:

(69) a. *The ghost frigthened Mary*
b. $[_{VP}$ √frighten $[_{PP}$ Mary $[_{p}$[CAUSER] $[_{DP}$ the ghost]]] →
$[_{VP}$ [√frighten+$_{p}$[CAUSER]] $[_{PP}$ Mary $[t_{p}$ $[_{DP}$ the ghost]]] →
$[_{VP}$ $[_{DP}$ the ghost$]_{i}$ [√frighten+$_{p}$[CAUSER]] $[_{PP}$ Mary $[t_{p}$ $[_{DP}$ t$]_{i}]$

Given that the empty P cannot assign case, the causer argument will move to the [spec, VP] a position where it can get case, just like regular subjects. This is in effect movement to a theta-position, but Pesetsky claims that this is permitted just in case the theta-role assigned compositionally by √frighten and the empty P together doesn't differ from the original theta-role assigned to the causer DP prior to movement.

Leaving aside many of the problematic details raised by this analysis, it is clear that the arguments affected by the suffix *-garri* don't have much in common under Pesetsky's account: in the case of passive *-garri* the argument affected is a true direct object of the verb; in the case of active *-garri*, the argument affected would correspond

---

[27] Grimshaw (1990: 15) claims that examples of the type *a man-frigthening god* are ungrammatical in English because the compound-internal argument (= *man*), the experiencer, is less prominent than the theme argument (= *god*). However, once the external R argument is introduced to form the adjective, this R controls the theme and thus makes it more prominent than the experiencer. Consquently, *a man-frightening god* shouldn't be ungrammatical. As Lisa Galvin (p.c.) tells me, examples like *a man-irritating comment* or *a vegetarian-pleasing menu* are close to right.

[28] The problem mentioned in note 12 remains to be worked out, though. This could be easily handled in Emonds' (1991, 2000) framework. I leave this for future research.

[29] For extensive criticism of Pesetsky's fine-grained semantic analysis see Ruwet (1993: 102-133) and Herschensohn (1999).

to the direct object of an empty P. The reader can check the adaptation of Pesetsky's analysis to Basque in the following two structures:

(70) a. *aipatu/gorrotatu* verbs: $[DP_1\ [DP_2\ V]$

b. *Jonek artikulua {aipatu, gorrotatzen} du*
"John {mentioned, hates} the article"

c. *Artikulua {aipagarria, gorrotagarria} da*
"The article is {mentionable, detestable}"

(71) a. *kezkatu* verbs: $[[_{pS}\ DP_1\ [_{p'}\ DP_2\ P]\ V]\ (\rightarrow [DP_2\ [_{pP}\ DP_1\ [_{p'}\ t_2\ t_P]\ V + P])$

b. *Ekonomiaren egoerak Jon kezkatzen du*
economy.gen situation.E worry aux
"The situation of the economy worries John"

c. *Ekonomiaren egoera kezkagarria da*
economy.gen situation worrying is
"The situation of the economy is worrying"

In other words, the cascade structures that Pesetsky provides for a *frighten*-like verb make the causer argument (the original *theme* in B & R) of that verb and the regular internal argument of a transitive verb look radically different. This difference, in turn, implies that there should be two different rules of *-garri,* contrary to what we have defended so far.[30]

To finish this short note on Pesetsky's treatment of psych verbs, I will mention another point where his zero-affix account is problematic. Pesetsky attributes the lack of nominalization with *frighten* verbs to the fact that the zero-affix doesn't license the attachment of any further suffixes (cf. Myers's generalization). It is interesting to note, in this regard, that many Basque *kezkatu* verbs are derived by zero-suffixation from the corresponding noun or adjective and none of these accept (event) nominalizations:

(72) a. **Our constant annoyance of Mary* (Pesetsky 1995: 74)

b. **Gure Mirenen nahaste etengabea*
our Mary.gen annoy constant.art
"Our constant annoying of Mary"

c. **Bonbardeaketen umeen beldurtzea gelditu behar dugu*
bombardment.gen child.gen frightening.art stop have aux
"We must stop the bombardments' frightening of children"

d. **Albistearen biztanleen harridura*
new.gen citizen.gen surprise
"The news' surprising of the citizens"

[30] I assume that instrumental subjects could be analyzed as *frighten*-verbs in Pesetsky's approach. If this is not so, then we are left with three kinds of arguments affected by *-garri,* hence three different rules.

And, if they do, the relevant nominal is usually related to the intransitive use of the verb:

(73) a. *Miren harritu egin zen nabarmen* (= b)
surprise do aux clear
"Mary got clearly surprised"

b. *Mirenen harridura nabarmena izan zen* (= a)
Mary.gen surprise clear.art be aux
"Mary's surprise was evident"

c. *Albisteak Miren harritu zuen nabarmen* (↑ d)
news.E surprise aux clear
"The news clearly surprised Mary"

d. (**Albistearen*) *Mirenen harridura nabarmena izan zen* (↑ c)
new.gen Mary.gen surprise clear.art be aux
"(*The new's) surprise of Mary was evident"

In other words, examples (73a) and (73b) are equivalent, but examples (73c) and (73d) are not: Mary's surprise in (73d) relates to her state of being, or becoming, surprised, not to something surprising her.[31]

These facts seem to square well and, indeed follow from, Pesetsky's treatment of the *frighten* class: it appears that nominalization suffixes cannot attach to the corresponding zero causative affix, thus proving the existence of the latter. However, verb formation via zero-suffixation is very common in Basque (cf. Odriozola 2003) and, contrary to the expectations of a Pesetsky-style analysis, many nominalizing suffixes can attach to zero-derived transitive verbs (74), even in cases where there is an inchoative-transitive alternation (75):

(74) transitive verbs derived via zero-suffixation[32]

	N/Adj	transitive verb	derived nominal
a.	*azal* "skin"	*azal(du)* "explain"	*azalpen, azalkuntza* "explanation"
b.	*bero* "hot"	*bero(tu)* "heat"	*beroketa* "heating", *berogailu* "heater"
c.	*garbi* "clean"	*garbi(tu)* "clean"	*garbiketa* "cleaning", *garbitzaile* "cleaner"
d.	*zapal* "flat"	*zapal(du)* "opress, step"	*zapalketa, zapalkuntza* "opression"
e.	*zuri* "white"	*zuri(tu)* "peel"	*zuriketa* "peeling", *zuritzaile* "peeler"

[31] In other words, even if it might disputably be related to the transitive use of *harritu*, it is a result nominal, not an event nominal (cf. Grimshaw 1990: 119ff).

[32] The verbs *azaldu, berotu* and *zuritu* also have an inchoative alternation with their corresponding meaning ("appear", "turn hot", and "turn white").

(75) verbs with an inchoative-transitive alternation

	inchoative verb	transitive verb	derived nominal
a.	*erre* "get burned"	*erre* "burn"	*erreketa* "burning", *erretzaile* "smoker, burner"
b.	*hazi* "grow"	*hazi* "grow, breed"	*haziera, hazkuntza* "breeding", "(2) education"
c.	*hedatu* "get extended"	*hedatu* "extend"	*hedapen, hedakuntza* "extension"
d.	*hil* "die"	*hil* "kill"	*hilketa* "killing", *hiltzaile* "killer"
e.	*ireki* "open"	*ireki* "open"	*irekiera* "opening", *irekitzaile* "opener"
f.	*itxi* "close"	*itxi* "close"	*itxiera* "closure", *itxigailu* "closer (machine)"
g.	*itzuli* "return"	*itzuli* "return, translate"	*itzulketa* "return", *itzulpen* "translation"

Furthermore, as *-garri* suffixation demonstrates, even *kezkatu* verbs derived from a zero-suffix can take certain suffixes, even though an event nominalization is impossible:

(76)	noun/adjective	psych verb	derived word
a.	*kezka* "worry	*kezkatu* "worry"	*kezkagarri* "worrying, worrisome"
b.	*lotsa* "embarrassment"	*lotsatu* "embarrass"	*lotsagarri* "embarrasssing"
c.	*haserre* "anger"	*haserretu* "anger"	*haserregarri* "infuriating", *haserretzaile* "infuriater"
d.	*maite* "dear"	*maitatu* "love"	*maitagarri* "lovable", *maitale* "lover"

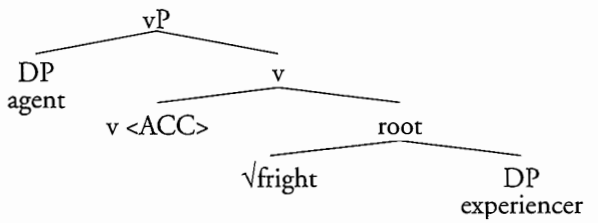
The existence of productive [zero suffix + lexical suffix] morphological processes in Basque cast serious doubts on Pesetsky's account of the ungrammaticality of data like (72) above, ungrammaticality which can hardly be attributed to the subcategorization restrictions of the zero-affix.[33] Instead, it seems that the impossibility of nominalizing *frighten/preoccupare/kezkatu* verbs is independent of the presence of a zero suffix and must be handled in a different way (cf. Grimshaw's explanation in terms of lack of external argument).

[33] Pesetsky can always argue that Basque zero-affixation, unlike its English counterpart, is not syntactic, but lexicon-internal; or else, that the Basque zero suffix is free to attach to any other suffix. This argumentation leaves us with a puzzle: we would need two sources to explain the lack of (event) nominalization with the *kezkatu/frighten/preoccupare* class: a lexical explanation for the Basque case *and* a morphosyntactic one in terms of subcategorization restrictions on the zero-suffix for the English case. Two sources for a restriction that looks pretty much the same in both languages.

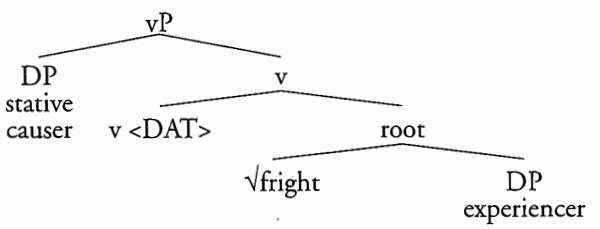
5.4. Arad's "stative little v" and *-garri*

In a couple of articles (cf. 1999a, 1999b) written in the minimalist framework, Arad proposes to treat object-experiencer verbs of the *frigthen-preoccupare* type as roots which can be complement to two different types of "v": "standard little v", which gives rise to the active interpretation of this type of psychological verb; and "stative little v", which gives rise to the non-active reading. Both types of little v share two transitivity features: they project an external, causer, argument and they establish an agreement relation with the experiencer argument (in fact, an argument of the root verb proper); the only difference is that the case checked by stative little v tends to be dative, rather than accusative:[34]

(77) a. standard little v



b. stative little v



The two types of little v differ in their semantics: standard little v implies the existence of an agent that brings about a change of state in the experiencer; stative little v, on the other hand, implies the existence of a stimulus that co-occurs with the mental state it causes; in the latter case, Arad claims, there is no change of state proper. I illustrate the semantic difference in (78):

(78) a. *Ana frigthened Laura deliberately* (action → mental state)
b. *This problem concerned Laura* (perception of stimulus/mental state run parallel)[35]

---

[34] When the object case is accusative, Arad (1999a: 16) assumes this is lexical or inherent case and regards stative v as deffective.

[35] Arad mentions a third, eventive and non-agentive, reading in a footnote. This seems crucial in Basque, given that most *kezkatu* verbs participate in the inchoative-transitive alternation (Levin 1983, Artiagoitia 1995, Etxepare 2003); that is to say, most of them are not stative but change of state verbs, whether they have an agent subject or not.

The alleged psych properties of *preoccupare* verbs would only obtain when the relevant root combines with stative little v, but they would not be properties of a given verb, but a consequence of the combination of the relevant root with stative little v. In other words, one should talk about the syntactic properties of stative constructions rather than syntactic properties of psych verbs in general.

Space limitations prevent us from giving a full-fledged account and critique of Arad's approach to psych verbs,[36] but the presentation above suffices to show that adapting Arad's view on *frigthen-preoccupare* type verbs to Basque would be imcompatible with a unified treatment of *-garri*. Why is this so? Because Arad proposes that the *frighten-preoccupare* verbs have external arguments, be it agentive or not, all the way through. In other words, under her account, there is nothing in common between the non-agentive subjects of the *frighten-preoccupare* type and the direct object of a regular transitive verb.[37]

6. Final remarks

In this article, I have argued that *-garri* is essentially a deverbal suffix that attaches to verbs that take two arguments; the suffix gives rise to both adjectives and nouns but only the former value is productive in present Basque. The two traditional values of the suffix, the active and the passive, can be reduced to one provided Belleti and Rizzi's original insight that *preoccupare/frighten* verbs lack an external argument is adopted; as shown here, verbs that take *-garri* in the active value do indeed have the properties of *preoccupare* and the like. Arguments that bear the instrumental theta-role are also affected by *-garri*; this fact suggests that they, too, be treated as internal arguments in Basque.

A unified analysis of *-garri* is possible within Grimshaw's framework: *-garri* takes the innermost argument of a two-argument verb and makes it the external argument of the newly created adjective; in this regard, *-garri* simply unifies what seems to be two different suffixes in English *(-able,* and the adjectival suffix *-ing* described in Brekke 1988). Theories that treat *preoccupare/frighten* verbs as causatives or regular transitives fail to predict that *-garri* suffixation might be one single rule and force a dual explanation for the suffix.

References

Albizu, P., in press, "Sobre la distribución sintáctica de las formas finitas del verbo vasco: condicionamiento léxico y sintáctico", *ASJU-International Journal of Basque Linguistics and Philology* 34.1.

Arad, M., 1999a, "What counts as a class? The case of psych verbs", in L. Pylkänen, A. van Hout, & H. Harley (eds.), *Papers from the UPenn/MIT Roundtable on the Lexicon*, MIT Working Papers in Linguistics 35, 1-23.

[36] See the previous note. Other problematic aspects of her proposal include the following: she takes NOM-DAT verbs as the unmarked pattern for non-agentive object-experiencer verbs, but the general tendency seems to be for these verbs to behave as regular transitives: cf. NOM-ACC in English and French (Ruwet 1993) or ERG-ABS in Basque. Arad (1999a) also obviates the fact that in *piacere* verbs the dative argument appears to be the subject, rather than the object (cf. Belleti and Rizzi 1988).

[37] This problem also extends to Oyharçabal's (2002, 2003b) treatment of psychological causative verbs, which he analyzes just like regular causative verbs following Pylkkänen (2002).

—, 1999b, "On little *v*", in K. Arregi et al. (eds), *Papers on Syntax and Morphology Cycle 1*, MIT Working Papers in Linguistics 33, 1-25.

Artiagoitia, X., 1995, "*-garri* atzizkiaren izaera bikoitzaz: zergatik den maitagarria bezain mingarria" [On the double nature of the suffix *–garri*: why it is both harming and lovable"], *ASJU-International Journal of Basque Linguistics and Philology* 29.2, 355-405.

—, 2000a, *Hatsarreak eta Parametroak lantzen* [Working with Principles and Parameters], Vitoria-Gasteiz: Arabako Foru Aldundia-EHU/UPV.

—, 2000b, "Oharrak aditz ezergatiboen inguruan" [Remarks on unergative verbs]. Paper delivered at the Basque Summer University, Pamplone.

—, 2002, "The functional structure of the Basque noun phrase", in X. Artiagoitia, J. A. Lakarra and P. Goenaga (eds.), *Erramu Boneta: Festschrift for Rudolf P. G. de Rijk*, Supplements of *ASJU*, Bilbao: Euskal Herriko Unibertsitatea, 73-90.

Azkarate, M., 1990, "Euskal *-garri* vs. latinezko *-bilis* bi balio desberdin" [Basque *-garri* versus Latin *-bilis*: two distinct values], in *Patxi Altunari Omenaldia*, Donostia: Mundaiz, 19-39.

—, 1995, "Sobre los compuestos de tipo 'sustantivo + sust/adj deverbal' en euskera", in P. Goenaga (ed.), *De Grammatica Generativa*, Donostia: Supplements of *ASJU*, 39-50.

Azkue, R.M. de, 1905, *Diccionario vasco-español-francés*, Bilbao. [Reed.: 1969, La Gran Enciclopedia Vasca].

—, 1925, *Morfología vasca*, Bilbao. [Reed.: 1969, La Gran Enciclopedia Vasca].

Belleti, A. and L. Rizzi, 1988, "Psych verbs and Theta Theory", *Natural Language and Linguistic Theory* 6, 291-352.

—, and —, 1991, "Notes on Psych-Verbs, θ-theory, and Binding", in Freidin (ed.), *Principles and Parameters in Comparative Syntax*, Cambridge MA: MIT Press, 132-162.

Brekke, M., 1988, "The Experiencer Constraint", *Linguistic Inquiry* 19.2, 169-180.

Chomsky, N., 1995, *The Minimalist Program*, Cambridge MA: MIT Press.

Emonds, J., 1991, "Subcategorization and Syntax-Based Theta-Role Assignment", *Natural Language and Linguistic Theory* 9.3, 369-429.

—, 2000, *Lexicon and grammar: the English Syntacticon*, Berlin: Mouton de Gruyter.

Etxepare, R., 2003, "Psychological predicates", to appear in J. I. Hualde & J. Ortiz de Urbina (eds.), *A grammar of Basque*, Berlin: Mouton de Gruyter.

Fabb, N., 1988, "English affixation is constrained only by selectional restrictions", *Natural Language and Linguistic Theory* 6, 527-540.

Gracia, Ll. et alii, 2000, *Configuración morfológica y estructura argumental: léxico y diccionario*, Bilbao: EHU.

Grimshaw, J., 1990, *Argument Structure*, Cambridge MA: MIT Press.

Herschensohn, J., 1999, "What does Zero Syntax add to an analysis of French psych verbs?". In E. Trevino and J. Lema (eds.) *Semantic Issues in Romance Syntax*, Amsterdam: John Benjamins, 105-119.

Laka, I., 1993, "Unergatives that Assign Ergative, Unaccusatives that Assign Accusative", In J. Bobaljik and C. Phillips, eds., *Papers on Case & Agreement I*, MIT Working Papers in Linguistics Volume 18, 149-172.

Levin, B., 1983, *On the Nature of Ergativity*, MIT PhD.

— and M. Rappaport, 1986, "The formation of adjectival passives", *Linguistic Inquiry* 17.4, 623-61.

—, 1988, "Non-event *-er* nominals: A probe into argument structure", *Linguistics* 26, 1067-83.

Lieber, R., 1992, *Deconstructing Morphology*, Chicago: University of Chicago Press.

Longobardi, G., 1994, "Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form", *Linguistic Inquiry* 25.4, 609-665.

—, 2000, "The structure of DPs: some Principles, Parameters and Problems". In C. Collins and M. Baltin (eds.), *The Handbook of Contemporary Syntactic Theory*, Oxford: Blackwell, 562-601.

Odriozola, J.C., 2003, "Verb-deriving processes in Basque", this volume, 185-222.
Ortiz de Urbina, J., 1989, *Parameters in the Grammar of Basque*, Dordrecht: Foris.
Oyharçabal, B., 2002, "Euskal kausatibo lexikalak" [Basque lexical causatives], *Lapurdum* 7, 271-294.
—, 2003a, "*Kor* atzizkiaz" [On the suffix *–kor*], in J.M. Makazaga & B. Oyharçabal (eds.), *Euskal gramatikari eta literaturari buruzko ikerketak XXI. mendearen atarian. Gramatika gaiak*, Iker-14 (I), Euskaltzaindia, Bilbo, 357-383.
—, 2003b, "Lexical causatives and causative alternation in Basque", this volume, 223-253.
Pesetsky, D., 1987, "Binding problems with experiencer verbs", *Linguistic Inquiry* 18.1, 126-140.
—, 1995, *Zero Syntax*, Cambridge MA: MIT Press.
Pylkkänen L., 2002, *Introducing Argument*, MIT PhD dissertation.
Rebuschi, G., 1993, "Basque Pronouns and Relativized Locality", in J.I. Hualde and J. Ortiz de Urbina (eds.), *Generative Studies in Basque Linguistics*, Amsterdam: John Benjamins, 115-144.
Ruwet, N., 1993, "Les verbes dits psychologiques: trois théories et quelques questions", *Recherches linguistiques de Vincennes* 22, 95.124.
Sarasola, I., 1996, *Euskal Hiztegia*, Donostia: Kutxa Fundazioa.
Villasante, L., 1974, *Palabras vascas compuestas y derivadas*, Oñate: EFA.
Williams, E., 1981, "Argument Structure and Morphology", *Linguistic Review* 1, 81-114.

VERB-DERIVING PROCESSES IN BASQUE

Juan Carlos Odriozola Pereira

Abstract

This paper assumes that the Basque suffix -tu *is an inflection suffix with a perfective value in analytical verb forms. However, it shows that addition of this suffix is the only overt change required to turn items of several categories into a participle (the verb citation form in Basque). The paper focuses on this derivational process, and shows that the auxiliary valences, theta-roles of arguments and aspect features of derived verbs all are closely related to subcategorization features of derivational bases, which are nouns, adjectives, adverbs and even postpositional phrases.*

Introduction

A Basque analytical[1] verb consists of an uninflected form of the lexical verb and a conjugated auxiliary.

(1) a. Jonek Peruri lanak itzuli dizkio
Jon-ERG Peru-DAT papers-ABS give-back-PERF AUX-A-D-E
'Jon gave back papers to Peru'

b. Jon ez da konpondu ondo Perurekin
Jon not AUX-A get on-PERF well Peru-with
'Jon did not get on well with Peru'

The uninflected form in (1a-b) bears a perfect aspect ending (Laka 1993a: §2) that, depending on the lexical item, may be either *-i* (2a) or *-tu* (*-du* after a nasal o lateral) (2b).[2]

* I would like to thank Axun Aierbe, Asier Alcazar, Xabier Artiagoitia, Elixabete Pérez-Gaztelu, Marijo Ezeizabarrena and Itziar Laka for their valuable comments and suggestions. All errors are, as always, my own.

[1] See Hualde (2003: §3.5.2.4.) for a description of synthetic forms and Alcazar (2002) for their set of aspectual readings.

[2] See a more accurate description of inflection suffixes in Hualde (2003: §3.5.1.).

(2) a. *itzul-i,*
give-back-PERF
'to give back'

b. *konpon-du*
get on-PERF
'to get on'

The uninflected form of lexical verbs can take other endings that are clear inflection suffixes: the future suffix *-ko* attached to the perfect form (3), and the imperfective *-tzen* (4).

(3) a. Itzul-i-ko dizkio
give-back-PERF-FUT AUX-A-D-E
'He will give them back to him'

b. Ez da ondo konpon-du-ko
not AUX-A get on-PERF-FUT well Peru-with
'He will not get on well with him'

(4) a. itzul-tzen dizkio
give-back-IMPF AUX-A-D-E
'He usually gives them back to him'

b. Ez da ondo konpon-tzen
not AUX-A get on-IMPF
'He usually does not get on well with him'

The uninflected form may also be constituted by the verb radical itself.

(5) a. itzul diezazkioke
give-back AUX-A-D-E
'He can give them back to him'

b. Ondo konpon daiteke
well get on AUX-A
'He can get on with him'

Mitxelena (1961: §4.5., 8.5., 12.8)[3] claims that *-tu* came into Basque in Latin verbal forms that were borrowed before the literary period of Basque

(6) *aditu* 'to listen' cf. Lat. *auditum*
arbuiatu 'to disdain' cf. Lat. *repudiatum*

The suffix *-tu* replaced patrimonial *-i* in some verbal items. In fact, today's Basque often shows two morphological counterparts for this participial form

(7) *ekoitz-i, ekoitz-tu* 'to produce', *sinets-i, sinis-tu* 'to believe'

On the other hand, some patrimonial verbs have a single form with *-tu* in today's Basque:

(8) *antola-tu* 'organize', *gal-du* 'to lose', *sar-tu* 'to enter'[4]

---

[3] See also Mujika (1982: 285) and Laka (1995b: §3.1).

[4] There is evidence in literary Basque syntax indicating that it must have been a form without suffix that had a perfect value. See Lafon (1943).

The suffix *-tu* is the only one available for the perfective form of verbs borrowed from other languages.[5]

(9) *ametitu* 'to admit' cf. Spanish 'admitir', *bainatu* 'to bathe' cf. Spanish *bañar, tratatu* 'to treat' cf. Spanish *tratar*

Basque has a very large number of non-verbal items whose verb counterpart bears the suffix *-tu*. In fact, the nonfinite form bearing the ending *-tu* is the usual verb citation form for speakers of Basque. Even if *-tu* was not borrowed from Latin, modern-day Basque can derive verbs only by means of *-tu*, which actually adjoins to items of several grammatical categories.

Therefore, on the one hand some authors (Laka 1993a, Haddican 2001) assume that *-tu* and *-i* and *-tzen* all are inflection suffixes bearing different aspectual values,[6] while on the other hand, Gràcia et al. (2000: §3.1.3.2) have shown that *-tu* is similar to Spanish and Catalan suffixes *-ar*, *-er* and *-ir* in some ways: they all may be taken not as derivative but as inflectional suffixes. In any case, the presence of the suffix itself is enough to change the category of the base to verb:

(10) Bas. *lodi*, Cat. *espès*, Sp. *espeso* 'thick'
Bas. *loditu*, Cat. espessir, Sp. *espesar* 'to thicken'

Therefore this paper assumes that the Basque suffix *-tu* is an inflection suffix with a perfective value in analytical verb forms. However, one could informally say that addition of the sufix *-tu* is sufficient to turn certain non-verbal items into verbs, i.e. into the participial, or citation form of Basque verbs. This paper focuses on the derivational process that yields verbs from several Basque items. Although we are not assuming that *-tu* is a derivational suffix, the citation throughout this work of a wide set of items yielding derived verbs has led us to use *-tu* to distinguish the derivational basis from the derived verb.

Sometimes the English translation of a derivational base will be both a nonverbal item and a verb. In such cases we will give only the base itself: in (11) it must be understood that noun *zati* 'piece' gives rise to a verb that should be translated to the same English word '(to) piece'. In cases in which English does not have the same entry for both base and verb, the English translation of the derived verb will be stated (12a). When, besides the attachment of *-tu* some (other) morphological changes occur, the derived verb will also be stated (12b).

(11) *zati* 'piece'

(12) a. *giltza* 'key' (> 'to lock')
b. *lore* 'flower' (> *loratu*)

This does not mean, however, that we will disregard the inflectional, or aspectual point of view of *-tu*. In our glosses it will be appear as PERF (13a), together with the other inflection suffixes of Basque (13b-c).

[5] See Alberdi (2003) in this book.

[6] See other points of view in Ortiz de Urbina & Uribe-Etxebarria (1991) and Artiagoitia (1995: §3, 4, 5).

(13) a. Jon ez da haserretu Mirenekin
Jon not AUX-A get angry-PERF Miren-with
'Jon did not get angry with Miren'

b. Jon ez da haserretzen Mirenekin
Jon not AUX-A get angry-IMPF Miren-with
'Jon does not usually get angry with Miren'

c. Jon ez da haserretuko Mirenekin
Jon not AUX-A get angry-PERF-FUT Miren-with
'Jon will not get angry with Miren'

In this work we will approach the verb-deriving processes in Basque from a merely descriptive point of view. Our point of departure is the work of Gràcia et al. (2000), which describes Basque, Catalan and Spanish derived verbs by means of category and aspectual features and argument structure of both derivational base and derived items. The aim of the present paper is to provide a more accurate description of the behaviour of *-tu*, which is involved in what has been taken as different processes of derivation. Basically we will provide new data about subcategorization features of bases and aspect features of derived verbs. In addition, we hope to provide an approach to both auxiliary valences and theta-roles of arguments of derived verbs.

Section 1 gives a brief description of some general points of Basque grammar involved in Basque verb-deriving processes.

We assume that the main feature that distinguishes several processes of derivation by *-tu* is the grammatical category of the item to which the suffix adjoins. Therefore, sections 2 to 5 correspond to nominal, adjectival, adverbial and postpositional bases.

1. Some general issues in the grammar of basque

We believe that readers not familiar with Basque really need some brief pointers focusing on certain aspects of the language. Moreover, the attempt to shed light with new data on phenomena that previous works (Gràcia et al. 2000, Hualde 2003: §3.7.4.) have pointed out must be preceded by some general considerations about this language.

Section 1.1 provides a brief description of the Basque auxiliary and case-marking system.

Gràcia et al. (2000) provide the argument structure of derived verbs and the consequent inheritance from the base. They also provide the lexico-conceptual structures consisting of semantic relationships of logical *participants* of the event expressed by derived verbs. They express argument structures following Grimshaw's framework (1990: §2) in which an argument structure is a set of arguments represented in terms of prominence relationships. In this work we will attempt to provide semantic relationships of participants mapped to syntax. That is, we will deal with habitual theta-roles of arguments. Section 1.2. is concerned with some minimal terminological and theoretical assumptions about theta-roles provided in order to make a description of semantic relationships in the several classes of derived verbs.

Section 1.3 offers an approach to a classification in aspect classes of derived verbs that will be defined in terms of sets of aspect readings related to sets of verbal forms available for each verb class.

1.1. Auxiliary valence and case-marks

Several types of Basque verbs, both derived and non-derived, are broadly described in Etxepare (2003), where two domains are made clear: lexical structure of verbs and morphological structure of the auxiliary. Etxepare points out that terms such as monovalent, divalent or trivalent are related to the lexical structure of verbs, while terms such as intransitive, transitive and ditransitive are concerned with the morphological structure of the Basque auxiliary.

Basque has an ergative/absolutive case-mark system. The morphologically non-marked absolutive case is assigned to both subjects of intransitive (14a) and objects of transitive (14b) verbs. The marked case ergative is habitually assigned to subjects of transitive verbs (14b). Noun phrases case-marked ergative appear before noun phrases case-marked absolutive. The marks of the cases in arguments are *-k* for ergative (ERG) and zero for absolutive (ABS).[7] The third case-mark is dative *-(r)i* (DAT).

(14) a. Katua etorri da
Cat-ABS come-PERF AUX-A
'The cat came'

b. Ardi-txakurrak katua ikusi du
Sheepdog-ERG cat-abs see-PERF AUX-A-E
'The sheepdog saw the cat'

Laka (1993a) describes Basque auxiliary morphology, which shows (obligatory) agreement clitics with these three arguments. We will express the corresponding clitics in auxiliaries as E, A and E in English glosses.

Ditransitive auxiliaries appear with noun phrases case-marked ergative, dative and absolutive (15a). Despite the obligatory agreement with three noun phrases, some of these verbs appear with two arguments case-marked ergative and dative (15b):

(15) a. Jonek Josebari katu bat oparitu dio
Jon-ERG Joseba-DAT cat a-ABS give-PERF AUX-A-D-E
'Jon gave a cat to Joseba'

b. Nik ardi-txakurren jabeei deitu diet
I-ERG sheepdog-GEN owners-DAT call-PERF AUX-A-D-E
'I called the sheepdog's owners'

[7] In a noun such as *katu* 'cat', *-a* is the determiner, which will not be expressed in the glosses of this work. Absolutive is not an overt case. Therefore, *katua* is a (determiner) phrase case-marked absolutive, and *katuak* a (determiner) phrase case-marked ergative.

Most verbs with a transitive auxiliary appear with noun phrases case-marked ergative and absolutive.[8]

(16) Jonek beste katu bat ekarri du
Jon-ERG another cat one-ABS bring-PERF AUX-A-E
'Jon brought another cat'

(17) a. Alexek otoitz egiten du
Alex-EGR prayer egin-IMPF AUX-A-E
'Alex prays'

b. Mikelek barau egin du
Mikel-ERG fast do-PERF AUX-A-E
'Mikel went on a fast'

Some Basque verbs take a transitive auxiliary but they have a single argument case-marked ergative. Most of them (18a) are the single-verb counterparts of complex predicates (17). Furthermore, some of the single-verb counterparts (18b) take an intransitive auxiliary bearing agreement with a single noun phrase case-marked absolutive (17b).[9]

(18) a. Alexek otoizten du
Alex-ERG pray-IMPF AUX-A-E
'Alex prays'

b. Mikel barautu da
Mikel-ABS fast PERF AUX-A
'Mikel fasted'

The intransitive auxiliary appears with a single noun phrase case-marked absolutive.

(19) Jon joan da
Jon-ABS go-PERF AUX-A
'Jon has gone'

On the other hand, (20) illustrates an intransitive/transitive alternation of the auxiliary that is very common in Basque.[10]

(20) a. Jon alaitu da
Jon-ABS gladden-PERF AUX-A
'Jon cheered up'

b. Ustekabeak Jon alaitu du
Surprise-ERG Jon-A gladden-PERF AUX-A-E
'The surprise gladdened Jon'

---

[8] Laka (1993b) claims that *otoitz* 'prayer' in the example (17a) of the text is a determinerless phrase case-marked absolutive. See note 7 and section 2.9 for Basque complex predicates such as *otoitz egin* 'to pray' in which the noun does not bear any overt mark. Phrases without a determiner and a quantifier are not allowed in Basque syntax and, since in this paper we do not deal with any theoretical assumptions about complex predicates, glosses will express nothing about the case of these nouns inside complex predicates.

[9] See Laka (1993b).

[10] See Zabala (2003b) in this book.

The verb in (20a) takes an intransitive auxiliary with a noun phrase case-marked absolutive just like the pattern in (19). The same verb in (20b) takes a transitive auxiliary and two noun phrases case-marked ergative and absolutive.

Therefore, Basque subjects are assigned ergative by divalent verbs (that take a transitive auxiliary) and are assigned absolutive by monovalent verbs (taking an intransitive auxiliary). Ergative is also assigned by monovalent verbs (bearing a transitive auxiliary).

Finally there is another kind of ditransitive auxiliary that appears with arguments case-marked dative and absolutive:

(21) Josebari bururatu zaio konponketa
Joseba-DAT come-to-mind-PERF AUX-A-D
'The solution occurred to Joseba'

1.2. Argument structure and theta-roles

Section 1.1. explains that the number of agreement marks in the Basque auxiliary does not necessarily match the number of arguments of the verb. In the words of Etxepare (2003), intransitive, transitive and ditransitive auxiliaries are not necessarily related to monovalent, divalent or trivalent argument structures.

In this regard, it should be noted that what we have described noun phrases bearing one of the three case-marks have a neutral order that mirrors the agreement morphology of the verb. The combinations available for argument noun phrases are expressed in (21).[11]

(22) ergative/dative/absolutive
ergative/dative
ergative/dative/absolutive
dative/absolutive
absolutive

Regarding argument classes, for the descriptive purposes of this work we assume Pustejovsky's (1995) classification (English examples and underlining are Pustejovsky's):

True arguments are the syntactically realized parameters of the lexical items.

(23) John arrived late

Shadow arguments are parameters that are semantically incorporated into the lexical item. They can be expressed only in some cases.

(24) Mary buttered her toast with an expensive butter

[11] See Laka (1993a) for the intricate morphology of agreement marks, and Zabala & Odriozola (1996) for specifications about the relationship between noun phrases marked by a certain case and their prominence in the clause.

True adjuncts are parameters that modify the logical expression, but are part of the situational interpretation, and are not tied to any particular lexical item's semantic representation. These include adjunct expressions of temporal or spatial modification:

(25) Mary drove down to New York on Tuesday.

For the descriptive purposes of this work we borrow from Theta-Role Theory some (semantic) labels generally accepted in Linguistics.

We assume that agent is the entity that executes the action expressed by the predicate. Inanimate agents (26b) are taken as causes.

(26) a. Josebak Xixka garbitu du
Joseba-ERG Xixka-ABS clean-PERF AUX-A-E
'Joseba washed Xixka'

b. 'Porrotak ikaratu du Jon urteetan zehar
Failure-ERG scare-PERF AUX-A-E Jon-ABS years-in through
'Failure scared Jon for years'

In this work we follow Artiagoitia (2000), who assumes that experiencer is an animate being that experiences a psychological state. We therefore assume that in (27-28) *Joseba* is but *Iñigo* is not an experiencer.

(27) Joseba haserretu da
Joseba-ABS get angry-PERF AUX-A
'Joseba got angry'

(28) Iñigo gizondu da
become a man-PERF AUX-A
'Iñigo became a man'

Artiagoitia defines patient as the entity that suffers the change of place expressed by the verb. Theme is the entity that suffers the change of state expressed by the verb. Gràcia agrees with Artiagoitia in that theme and patient are not easy to distinguish. We embrace both concepts by means of the covert term theme.

Some Basque verbs (→§2.1.2) express both psychological states (29) and changes in (psychological) states (30a-b).

(29) Heriotzak Marisa ikaratzen du
Death-ERG Marisa-ABS frighten-IMPF AUX-A-E
'Death frightens Marisa'

(30) a. Istripuek Marisa ikaratzen dute
accident-PL-ERG Marisa-ABS frighten-IMPF AUX-A-E
'Accidents frightens Marisa

b. Marisa ikaratu da gaur
Marisa-ABS get frightened-PERF AUX-A today
'Marisa got frightened today'

However, we assume that in every event in which a psychological feature is involved, there is an experiencer, even if it is a change of (psychological) state that an argument suffers. On the other hand, we assume that any change of place bears a theme such as *Alex* in (31).[12]

(31) Alex aurreratu da
Alex-ABS go ahead-PERF AUX-A
'Alex went ahead'

Artiagoitia defines beneficiary as the entity that is affected advantageously or adversely by the action expressed by the verb. Minkoff (1997) has defined beneficiary as something that the theme is understood to be for. We assume that beneficiary is animate and therefore, it seems clear that *belarritakoak* in (32) is not a beneficiary but a theme.

(32) Jonek eta Josebak urre zaharraz urreztatu nahi dituzte
Jon-ERG and Joseba-ERG gold old-*z* gold-plate-PERF want AUX-A-E
beren belarritakoak
'Jon and Joseba want to gold-plate their earrings with old gold'

Furthermore, we assume that theme does but that beneficiary does not experience the action of the verb. Therefore *Lehendakariari* is a beneficiary, and *Irlanda* is a theme.

(33) Legebiltzarkideek zerbait galdetu zioten
members of parlament-ERG something-ABS ask-PERF AUX-A-D-E
Lehendakariari
president-DAT
'The members of parliament asked the President something'

(34) Irlanda baketu da
Ireland-ABS get peace-PERF AUX-A
'Ireland got peace'

Gràcia et al. (2000) defines instrument as the inanimate entity used to execute the action of the verb. We assume that the shadow argument of *giltzatu* 'to lock with a key' is an instrument, *giltza* 'key'.

(35) Josebak ez du giltzatzen baserriko atea
Joseba-ERG not AUX-A-E lock-IMPFF door-ABS
'Joseba does not lock the door of the farmhouse'

Finally Artiagoitia notes the possibility that verb complement (embedded) clauses bear a theta-role:

---

[12] See Minkoff (1997), which explores the lexico-interpretational animacy entailment that (optionally) converts causer, theme, patient and goal to agent, volunteer, sensor and beneficiary. See also his restrictive interpretation of experiencer.

(36) Internetetik galdetu dit polita naizen
Internet-by ask-PERF AUX-A-D-E whether good-looking am-wether
'He asked me whether I am good-looking by Internet'

Here, we will use quotations that pick up both case-marks of arguments and theta roles. So, the pattern of *garbitu* 'to wash' in (26a) is agent (ERG)/theme (ABS), which expresses both ergative/absolutive cases and agent/theme thematic relationships of arguments *Josebak* and *Xixka.* This quotation is also able to express the auxiliary alternation (→§1.1) and so, the pattern of *ikaratu* 'to frighten' in (30a-b) is expressed as (agent(ERG))/experiencer (ABS). Finally this quotation system is concerned with overt order of arguments and the pattern of *ikaratu* is distinguished from the one of *erdeinatu* 'to disdain' which will be expressed as experiencer (ERG)/agent (ABS).

Finally, general semantics will be described by means of general paraphrases such as 'to become N', 'to make A', 'to use N', in which the specification of grammatical category corresponds not only to a semantic concept but also to the grammatical category of the derivational base itself.

1.3. Set of aspect readings

Vendler's aspect classes of events are widely certified in the literature about verbs of languages of the world. Although aspect classes were claimed to be lexical, today it is well known that differences between states, activities, accomplishments and achievements are not strictly lexical (Verkuyl 1989). However these labels are useful as descriptors of (four) sets of compositional aspect readings (De Miguel 1999). Therefore we will be speaking about syntactic expressions of four sets of compositional readings.

1.3.1. Pustejovsky (1995) describes both achievements and accomplishments as a set of two subevents. The former denotes a process and a result state that heads the whole event. The latter denotes a process that heads the whole event and a result state. In both cases processes are limited, i.e., they denote a point at which process occurs or finishes.

1.3.1.1. Some Basque postpositional phrases express a measurable time in which the event occurs, i.e., they bear a frame reading. Certain Basque verbs (37a) do not accept this kind of frame readings. These verbs do not allow the quantifier *asko* 'very much/often' in its intensive reading (37c). We assume that they denote events that occur at a point in time and we will call them punctual achievement verbs.

(37) a. *ahaidetu* 'to become related' (→§2.2.3), *aholkatu* 'to advise' (→§2.9.4), *mailegatu* 'to borrow' (→§2.10.2), *koroatu* (> 'to coronate/crown') (→§2.9.9), *azaleztatu* 'to bind' (→§4.2.2)

b. *Ordu erdian ahaidetu dira
hour half become relative-PERF AUX-A
'They became related in half an hour'

c. *Asko ahaidetu dira
a lot become relative-PERF AUX-A very much
'They became related very much'

Some of these verbs allow non-frame readings for postpositional phrases (38b) but the intensive reading of the quantifier is always avoided.

(38) a. *konturatu* 'to realize' (→§1.3.1.1)

b. Bi hilabetetan konturatu nintzen zail izango zela
two months-in realize-PERF AUX-A hard be-future AUX-A-that
'I realized two months later that things would be hard'

c. *Asko konturatu naiz
very much realize-PERF AUX-A
'I realized very much'

1.3.1.2. Some Basque verbs behave like punctual achievement verbs in that they do not accept the quantifier *asko* in their intensive reading (39c). However they admit frame readings (39b), since they consist of processes that finish at a point. We will call them terminative achievements.

(39) a. *jabetu* 'to take hold' (→§2.2.2) (*loratu* 'to flower' (→§2.4.1), *amaitu* (>'to finish') (→§2.10.4), *sailkatu* 'to classify' (→§4.1.3), *porturatu* 'to bring/put into port' (→§5.1.2)

b. Arrosondoa bi egunetan loratu da
rose-ABS two days-in flower-PERF AUX-A
'The rose bloomed in two days'

c. *Arrosondoa asko loratu da
very much flower-PERF AUX-A
'The rose flowered very much'

1.3.1.3. Some Basque verbs accept both frame and intensive readings, since they express both the starting of an event and subsequent identical subevents that yield more and more intensive results. In fact universal quantifiers sound odd with this class of verbs (40b) but both intensive reading of *asko* 'very much' (40c) and even progressive modifiers are licensed (40d). We will call them ingressive achievements.[13]

(40) a. *gihartu* 'muscle' (→§2.10.4), *ahuldu* 'to weaken (→§3)

b. ?Guztiz gihartu da
completely become brawny-PERF AUX-A
'He developed his muscles completely'

c. Asko gihartu da
very much become brawny-PERF AUX-A
'He developed his muscles very much'

[13] De Miguel (1999: §46.3.2.4) distinguishes between ingressive, progressive and terminative verbs.

d. Oraindik gehiago gihartu da
more become brawny-PERF AUX-A
'He developed his muscles even more'

On the other hand, there are several verbs that bear two meanings and sets of readings. The former correspond to a punctual reading (41). The latter have a terminative set of aspect readings (42).

(41) *apaiztu* 'to ordain' (→§2.2.3), *gizondu* 'to turn into a man'

(42) *apaiztu* 'to get the appearance of a priest', *gizondu* (→ §2.2.2) 'to become a man'

1.3.2. Intensive modification is not available to the Basque accomplishment set of aspect readings (43c).

(43) a. *ikaztu* 'to become coal' (→§2.2.1), *birrindu* 'to crumble' (→§2.3.3), *larrutu* 'to skin' (→§2.4.3)

b. *Ogia asko birrindu du
Bread-ABS very much crumble-PERF AUX-A-E
'He crumbled the bread very much'

Frame modifiers and universal quantification are available here.

(44) a. Ogia bi minutuan birrindu du
Bread-ABS two minutes crumble -PERF AUX-A-E
'He crumbled the bread in two minutes'

b. Ogia guztiz birrindu du
Bread-ABS completely crumble-PERF AUX- A-E
'He crumbled the bread completely'

1.3.3. States are supposed to be non-limited and non-dynamic events. Basque state verbs admit quantification of the event (45b) but universal quantifiers sometimes sound odd (45c). As could be expected frame adverbs give rise to a reading of (a part of) the time the state occurs in (45d). Unlike in limited events, no change occurs at that time.

(45) a. *erdeinatu* 'to disdain', *damutu* 'to regret', *ikaratu* 'to frighten' (→§2.1.1), *gogoratu* 'to remember' (→§5.1.1)

b. Jonek Estibaliz asko erdeinatzen du
Jon-ERG Estibaliz-ABS very much disdain-IMPF AUX-A-E
'Jon disdains Estibaliz very much'

b' Damutu naiz zu irainduaz
regret-PERF AUX-A you-ABS insult-*az*
'I regretted having insulted you'

c. ?Jonek Estibaliz guztiz erdeinatzen du
Jon-ERG Estibaliz-ABS completely disdain-IMPF AUX- A-E
'Jon disdains Estibaliz completely'

d. Hiru urtean erdeinatu du eta aurrerantzean ere
three years-in hate-PERF AUX-A-E and from now on
erdeinatuko du
hate-PERF-FUT AUX- A-E
'He disdained her for three years and will disdain her from now on'

Erdeinatu verbs, i.e., Basque *disdain* verbs take a non-habitual (45b) or habitual (46) state reading for all the verbal forms,

(46) Jonek Maite erdeinatzen du zenbaitetan
Jon-ERG Maite-ABS disdain-IMPF AUX-A-E
'Jon disdains Maite sometimes'

Ikaratu verbs, i.e., Basque *frighten* verbs get state readings (47) as Basque *disdain* verbs do (45-46).

(47) Gizakiak Jon ikaratzen du
Mankind-ERG Jon-ABS scare-IMPF AUX-A-E
'Mankind scares Jon'

However, unlike Basque *disdain* verbs, Basque *frighten* verbs may take an achievement change of state in some contexts related (48a) or not (48b) to habitual events:

(48) a. Josebak Jon ikaratzen du zenbaitetan
Joseba-ERG Jon-ABS scare-IMPF AUX-A-E sometimes
'Joseba scares Jon sometimes'

b. Josebak Jon ikaratu du gaur
Joseba-ERG Jon-ABS scare-PERF AUX-A-E today
'Joseba scared Jon today'

Notice that the imperfect suffix *-tzen* licenses a unique state reading in some Basque *fear* verbs (45-46), but it may denote change of states in Basque *frighten* verbs (48).

On the other hand, some Basque *disdain* verbs such as *erdeinatu* 'to disdain' just license the reading of a past time frame for the state (45c), whereas some other Basque *fear* verbs such as *damutu* (45b') 'to regret' are related to the start of the state that still lasts in the present.

Some other verbs with psychological meaning are the result of different deriving processes but they pattern with Basque *frighten* verbs in their set of aspect readings.

(49) a. Hiri horretan Joxez gogoratzen da beti
city that-ABS see-IMPF-*ean* Joxe-*z* remember-IMPF AUX-A
'He always remembers Joxe when he is in that city'

b. Dena gogoratzen dut
All-ABS remember-IMPF AUX-A-E
'I remember everything'

1.3.4. Activities are non-limited dynamic processes. Basque verbs bearing activity sets of readings admit the quantifier *asko* which may take either an habitual (50b-b') or an intensive (50c) reading. As is well known, they denote a subject-controlled event.

(50) a. *barautu* 'to fast' (→§2.9.2), *babestu* 'to protect' (→§2.10.6), *harrikatu* 'to stone' (→§4.1.1)

b. Asko dantzatzen zuen
a lot dance-IMPF AUX-A-E
'He used to dance a lot'

b' Gutxiago dantzatzea erabaki du
less dance-*tzea* decide-PERF AUX-A-E
'He decided to dance less'

c. Alexek asko babesten du Ane
Alex-ERG very much protec-IMP AUX-A-E Ane-ABS
'Alex protects Ane very much'

2. Verbs derived from nouns

Gràcia et al. (2000) describe seven types of derivation from nouns, which seems to be a very wide range of items from this grammatical category. Classes of derivation are described in terms of paraphrases, argument-structure inheritance and aspect features of the derived verbs. We will focus on the classes of nouns that act as the base of the derivation, so we will describe more than twenty subclasses that will provide new data on the classes of nouns themselves, type of auxiliary, inheritance of argument structure and aspect features of the derived verb.

It is well known that some nouns may project predicates, i.e. they have argument structure. Section 2.1 deals with psychological nouns that filter their argument (structure) to both complex predicates and derived verbs. In addition, Zabala (1993: §4.4.6.) has described two more classes of nouns that may head nominal predicates: profession nouns and event nouns. Section 2.2 deals with nouns that are predicate-likely in some points. Most of them are nouns related to human relationships. Moreover, verbs derived from them express a change of state that may also be expressed by the noun itself and a copula. These derived verbs pattern in some points with verbs derived from adjectives (→§3), which could be expected from the not so clear limit separating Basque nouns and adjectives. In this case we will deal only with the structure of verbs.

Sections 2.3. to 2.7. describe verbs derived from several classes of nouns, without assuming any argument structure for the latter: nouns expressing a part of the whole (→§2.3.) or inalienable possession (→§2.4.), relational location nouns (→§2.5.), meal nouns (→§2.6.), and instrument nouns (→§2.7.).

Section 2.8. is concerned with weather-derived verbs. In section 2.9. we see verbs derived from event nouns that sometimes clearly filter their arguments to the verb. Finally, section 2.10. is concerned with result nouns.

2.1. Psychological nouns

Basque exhibits the two classes of psychological verbs (→§1.3.3.)[14] that are well described in other languages. Most of them are derived from nouns that designate a psychological state.

We assume that Basque psychological verbs of the first class such as *erdeinatu* 'to disdain' (51) bear an experiencer (case-marked ergative) and an agent (case-marked absolutive).

(51) a. *damu* 'regret' (> 'to regret'/'to back down'), *deitore* 'lamentation' (> 'to lament'), *erdeinau* (*erdeinatu* 'to disdain'), *irrika* 'longing' (> 'to long for'), *pairu* 'suffering' (> 'to suffer')

b. Jonek Estibaliz erdeinatzen du
Jon-ERG Estibaliz-ABS disdain-IMPF AUX- A-E
'Jon disdains Estibaliz'

Psychological verbs of the second class suchs as *ikaratu* 'to frighten' bear a cause-agent (case-marked ergative) and an experiencer (case-marked absolutive).[15] Nevertheless, unlike English *frighten* verbs, Basque *frighten* verbs follow an alternation auxiliary pattern (cause-agent (ERG))/experiencer (ABS) pattern.[15]

(52) a. *ardura* 'worry, care' (> 'to worry'/'to worry about'), *adore* 'courage' (> 'to encourage'/'to become animated'), *ikara* 'fright' (> 'to frighten/'to become frightened'), *lotsa* (> 'to put to shame/'to be ashamed, to blush'), *poz* 'joy' (> 'to gladden'/'to be glad')

b. Heriotzak Jon ikaratzen du
Death-ERG Jon-ABS frighten-IMPF AUX-A-E
'Death frightens Jon'

c. Jon erraz ikaratzen da
Jon-ABS easy become frightened-IMPF AUX-A
'Jon is easily frightened'

---

[14] See Artiagoitia (2003) in this book.

[15] It is well known that agentivity usually is not denoted by the non-prominent argument. Moreover, it seems odd that Basque shows both patterns, experiencer (ERG)/agent-cause (ABS) and cause-agent (ERG)/experiencer (ABS). The problem was pointed out by Laka (1993a) for Basque, but in fact, psychological verb grammar is intricate in several languages as Grimshaw (1990: §2.3.) pointed out. She claims that the non-experiencer argument is a theme in both verb classes. Etxepare (2003: §4.1.4.6.3.) assumes that the role of argument case-marked ergative in Basque *frighten* verbs is a stimulus-cause. In fact the agentivity related to ergative case-mark matches in a very wide set of both non-psychological and psychological Basque verbs. See Artiagoitia (2003), Oyharçabal (2003) and Zabala (2003b) in this book.

Therefore, the paraphrase for Basque *fear* verbs is 'to feel N' but two paraphrases are available for transitive and intransitive auxiliary counterparts of Basque *frighten* verbs, as can be seen in English translations. The former's paraphrase (51b) is 'to make feel N' and the latter may be paraphrased as 'to feel N'.

Sometimes there is a complex predicate counterpart for the derived verbs of both classes. Zabala (2002) describes Basque complex predicates as a problem of limit between syntax and (compositional) morphology. Complex predicates bear in most of the cases a nude noun or adjective and a verb. She claims that there are two kinds of Basque complex predicates. The former (53b) bears either a copula or an auxiliary whereas the latter (54b) bears either a full verb or a light verb. In the latter (54b) it is the verb that determines the argument structure of the complex. In the former (53b), the nude noun or adjective may project predicates and therefore it gives at least one argument at least one argument to the complex predicate.

(53) a *beldur* 'fright' (< 'to frighten/to become frightened')

b. Jon gizakiaren beldur da
Jon-ABS death-GEN fear is
'Jon is afraid of mankind'

c. Gizakiak Jon beldurtzen du
Mankind-ERG Jon-ABS scare-IMPF AUX-A-E
'Mankind scares Jon'

(54) a. *erdeinu* 'disdain' (< *erdeinatu)*

b. Jonek Estibalizi erdeinu dio
Jon-ERG Estibaliz-DAT disdain have-A-D-E
'Jon disdains Estibaliz'

c. Jonek Estibaliz erdeinatzen du
Jon-ERG Estibaliz-ABS disdain-IMPF AUX-A-E
'Jon disdains Estibaliz'

However, *erdeinu* 'disdain' (54) shows an argument with a dative mark (→§2.9), and *beldur* (53) bears an argument with a genitive mark. Notice that derived verb *erdeinatu* 'to disdain' bears the same agent argument case-marked absolutive in (54b), whereas the cause-agent of *beldurtu* in (53b) is case-marked ergative. We assume that derived verbs inherit both experiencer and cause-agent from these nouns bearing an argument structure inside a complex predicate.

Etxepare (2003) points out that some complex predicates have a unique state reading, whereas the verb counterpart bears a set of readings related to several verb forms in the Basque paradigm. Section 1.3. shows that some Basque *disdain* verbs such as *erdeinatu* 'to disdain' have a unique state reading similar to that of the complex predicate, although an habitual reading of the state is available. Basque *frighten* verbs where a wide set of readings clearly contrasts with the unique state reading of the complex predicate counterpart.

Finally Odriozola (1993) showed that some (literary) derived verbs of this class select embedded clauses. They have a single punctual achievement reading.

(55) *kilika* 'excitement' ('to incite'), *narrita* 'instigation' (> 'to instigate')

Basque *frighten* verbs share all their characteristics described above with verbs derived from physiologic nouns.

(56) *egarri* 'thirst' (> 'to make/get thirsty), *gose* 'hunger'

2.2. Verbs derived from human and material nouns

Under this heading are verbs paraphrased as 'to (make) become a N'. All of them denote a change of state with individual-level results in the sense of Kratzer (1988), i.e., they express the acquisition of non-transitory properties and they usually follow the pattern (cause-agent (ERG))/theme (ABS).

(57) a. *apaiztu* 'to ordain'/'to become a priest'

b. Apaiztu dute
ordain-PERF AUX-A-E
'They ordained him'

c. Apaiztu da
become a priest-PERF AUX-A
'He became a priest'

This is clear in the adjective-predicate counterpart. In fact, Zabala (2003a) points out that individual-level predicates usually appear with the copula *izan* 'to be' and suffix *-a*:

(58) Apaiza da
priest-*a* is
'He is a priest'.

Regarding argument structure, Gràcia et al. (2000) do not specify any for these classes of derivational bases. In fact Gràcia (1994: §72) assumes that some Catalan nouns (59a) similar to what we will see in section 2.2.1 do not have argument structure. Gràcia shows that the result of these Catalan and Spanish verb-deriving processes is not predictable as can be seen in the contrast between the meanings of (59a-b) and (59c).

(59) a. *gas* 'gas' (> *gasificar* 'to turn into gas/to gasify)

b. *gel* 'ice' (> 'to freeze')

c. *carácter* 'character' (> 'to characterize, to give character')

However this is not the case of Basque derived verbs that always pattern with both (57) and (59a-b). Some derived verbs (→§2.2.2., §2.2.3.) not described by Zabala

(1993) nor by Gràcia (1994) bear an experiencer that is very closely related to what the nominal base may predicate for.[16]

(60) *senide* 'relative' (> *senidetu* 'to become related'), *gizon* 'man' (> *gizondu* 'to become a man')

In fact, Gràcia (1994: 73, note 7) pointed out that in the case of new Spanish and Catalan technical words, the result is predictable in that it always patterns with (59a-b).

(61) *os* 'bone' (> *ossificar* 'to ossify')

She says that, intuitively, these nouns behave like adjectives (→§2) in that subject acquires all the properties of the noun.

We distinguish three classes of verbs:

2.2.1. Material nouns derive verbs that show the set of aspect readings of achievements.

(62) *ikatz* 'coal' (> *ikaztu* 'to turn into coal'), *izotz* 'ice' (> *izoztu* 'to freeze'), *lurrun* 'vapour' (> 'to evaporate'), *ozpin* 'vinegar' (> 'to turn (in)to vinegar'), *ur* 'water' (> 'to turn (in)to water')

2.2.2. Some nouns mainly related to human traits show the set of aspect readings of terminative accomplishment.

(63) *adiskide* 'friend' (> 'to become friends'), *gizon* 'men', *haurtu* 'child', *jabe* 'owner' (> 'to take over'), *jaun* 'mister' (> *jaundu* 'to become a gentlemen')

2.2.3. Some verbs derived from human nouns take the set of aspect readings of punctual achievements.

(64) *ahaide* 'relative' (> 'to become related'), *apaiz* 'priest', *erromes* 'pilgrim', *giristino* 'Christian', *gotzain* 'bishop'

2.3. Nouns expressing a part of the whole

Gràcia et al. (2000) describe a derivation process in which the base is a part noun. All of the derived verbs express a change of state of a theme that is seen as broken or divided into parts. The paraphrase is 'to make Ns'.

2.3.1. Some nouns (65a) express unlimited but regular parts of a whole. These parts are closely related achievements that may be indefinitely repeated. That can be seen in the fact that intensive reading provided by the quantifier *asko* 'very much' is ruled out by punctual and terminative achievements but is licensed here with repetitive

[16] Gràcia claims that argument is provided by the derivational suffix.

reading. Notice that the reading of ingressive achievements for *asko* is intensive but theme is affected in the same way as in repetitive readings.

(65) a. *zati* 'piece', *adar* 'branch'

b. Ogia asko zatitu du
bread very much piece-PERF AUX-A-E
'He pieced the bread very much (i.e. he cut/broke the bread into a lot of pieces)'

This argument may also be mapped to syntax as a postpositional phrase.

(66) Zati txikietan zatitzen du ogia
piece little-in piece-IMPF AUX-A-E bread-ABS
'He cut/broke the bread into little pieces' (i.e. 'he diced the bread')

2.3.2. Some other nouns are related to parts of a whole that may be irregular. They produce verbs that express an achievement event that is not necessarily iterative but rather ingressive or terminative. The results with intensive quantifier *asko* are dubious.

(67) a. *apur* 'bit' (> 'to break'), *puska* 'portion, bit' (> 'to break')

b. ?Liburua asko apurtu du
book-ABS very much break-PERF AUX-A-E
'He broke the book very much'

2.3.3. Some nouns are not part of a whole and their derived verbs bear a meaning that is lexicalized to a certain extent. They exhibit the set of aspect readings for accomplishments.

(68) a. *birrin* 'crumb' (> 'to crumble'), *irin* 'flour' (> 'to pulverize')

b. *Ogia asko birrindu du
Bread-ABS very much crumble-PERF AUX-A-E
'He crumbled the bread very much'

c. Ogia guztiz birrindu du
Bread-ABS completely crumble-PERF AUX-A-E
'He crumbled the bread completely'

2.4. Inalienable-possession nouns and deprivation verbs

Some nouns express a kind of inalienable possession. Their derived verbs always denote at least a growing change of state.[17]

[17] See Zabala (2003b) in this book.

2.4.1. Some nouns that are not interpreted as susceptible to deprivation produce verbs that take an intransitive auxiliary. They all have a possessor case-marked absolutive. They express a meaning of growing related to a terminative achievement. Their paraphrase is 'to get N'.

(69) *abar* 'branch' (> 'to grow branches'), *hosto* 'leaf' (> *hostatu* 'to sprout leaves'), *lore* 'flower' (> *loratu*), *ninika* 'bud'

2.4.2. Nouns that can be interpreted as susceptible to deprivation produce verbs with a special auxiliary alternation described by Etxepare (2002).

(70) *ale* 'grain' (> 'to pit/to shake out grain'), *kimu* 'branch' (> *kimatu* 'to hud/to prune'), *luma* (> 'to grow feathers/to pluck')

The intransitive auxiliary counterpart has a meaning of growing.

(71) Txita horia leihoko kaiolan lumatu zen
chick yellow-ABS window-in grow feathers-PERF AUX-A
'The yellow chick in the cage in the window grew feathers'

Transitive counterparts have an accomplishment deprivation meaning paraphrased as 'to deprive of N' and an agent (ERG)/theme or experiencer (ABS) pattern.

(72) Estibalizek oilaskoa lumatu du
Estibaliz-ERG chicken-ABS pluck-PERF AUX-A-E
'Estibaliz plucked the chicken'

2.4.3. Some nouns that are not interpreted as something that grows produce verbs that only have the transitive counterpart.

(73) *larru* 'skin'

2.5. Location nouns

Gràcia et al. (2000) describe Basque verbs derived from nouns that may be paraphrased as 'to go/bring/take/carry to N'. In that work, no restrictions were described for the base, but the derived verbs belong to a closed list of lexical entries. We are dealing with location nouns of two classes that are a kind of shadow arguments whose role would be similar to a goal. All of them show auxiliary alternation and follow an (agent-cause (ERG))/theme (ABS) pattern. They take the set of aspect readings of accomplishments.

2.5.1. Some location location nouns yield verbs whose meaning may be either a physical or a figurative movement.

(74) a. *azal* 'surface' (> 'to explain, to appear), *bazter* 'edge' (> 'to exclude'), *buru* 'head' (> 'to carry out'), *saihets* 'side' (> 'to avoid')

b. *bahe* 'sifter' (> 'to weed out'), *eskola* 'school' (> 'to take to school', *hondar* 'bottom' (> 'to sink'), *kaiola* 'cage', *lerro* 'line' (> *lerratu* 'to glide')

2.5.2. In some cases the bases are relational locative nouns[18] and their derived verbs can more or less be foreseen, since we are dealing with nouns that grammaticalize place. However, their meaning is not so predictable:

(75) *alde* 'side' (> *aldatu* 'to change'), *inguru* 'surroundings' (> *inguratu* 'to go round')

It seems that no new verbs are being produced this way in modern-day Basque.

2.6. Meal nouns

Food-activity nouns yield verbs expressing the activity itself. Their paraphrase is 'to have N' and they take an auxiliary that dialectally may be either transitive or intransitive. Agent is case-marked either ergative or absolutive.[19] They express achievements.

(76) *afari* 'dinner' ('to have dinner/to dine'), *bazkari* 'lunch'

2.7. Instrument nouns

Some instrument and tool nouns yield verbs that have lexically incorporated the noun as an instrument shadow argument.

In most cases the pattern is (ERG)/theme (ABS) and they are either activities or achievements.

(77) *akuilu* 'spur' (> *akuilatu* 'to incite'), *botoitu* 'button' (> 'to button up'), *giltza* 'key' (> *giltzatu* 'to lock'), *laia* 'shovel', *orraz* 'comb'

2.8. Weather nouns

Weather nouns yield transitive verbs that express achievements.

(78) a. *ateri* 'clear weather' (> 'to clear up'), *ilunabar* 'twilight' (> 'to grow dark'), *negutu* 'winter' (> 'to become winter'), *sargori* 'sultry weather' (> 'to become sultry weather')

b. Durangaldean atertu du
Durango area-in clear-up-PERF AUX-A-E
'It cleared up in the Durango area'

[18] See a description of this subclass of Basque nouns in Odriozola (2002).

[19] See Fernández (1997: §2) for theoretical issues involved in the syntactical behaviour of these verbs.

2.9. Event nouns

It is well known (Grimshaw (1990: §3.2, Gràcia 1994) that the subclass of process or event nouns have arguments expressed in syntax with an event reading. In a first approach, we use *event noun* as a semantic label that picks up a subclass of Basque nouns that always bear an event reading. Furthermore, they license genitive phrases or other phrases for the participants in the event that also are expressed in the syntax as arguments of the derived verb. So we can see a set of nouns that give rise to a rather regular verb-deriving process. All of the derived verbs may be paraphrased as 'to make/do/give N'.

There is no doubt about the fact that it is the event noun that has given rise to the verb (and not vice-versa), since the suffix *-tu* was borrowed when nouns and even complex predicates bearing nouns were already available in Basque. In fact, besides derived verbs, Basque often has available a complex predicate counterpart of the second class described by Zabala (2002).[20] They consist of a determinerless noun and a verb, which is what determines the argument structure of the predicate.

(79) Mikelek barau egin du
Mikel-ERG fast do-PERF AUX-A-E
'Mikel went on a fast'

The conjugated complex predicate verb in (79) is transitive and the agent is assigned ergative. Notice that the nude noun appears to be an argument of the verb (Laka 1993b) (→§1.1). In fact, verbs such as *egin* 'to do', *eman* 'to give' or *hartu* 'to take' always take a transitive auxiliary and they usually assign ergative and absolutive case-marks to their arguments (see notes 6 and 7). Derived verbs may take either intransitive (80a) or transitive (80b-c) auxiliaries.

(80) a. Mikel barautu da
Mikel-ABS fast PERF AUX-A
'Mikel fasted'

b. Mikelek arnastu du
Mikel-ERG breathe-PERF AUX-A-E
'Mikel breathed'

c. Gurasoek Alex Txurdinagan bataiatu zuten
parents-ERG Alex-ABS Txurdinaga-*in* baptize-PERF AUX-A-E
'The parents baptized Alex in Txurdinaga'

Neither the derived verbs nor their complex predicates are states and they share a wide range of aspectual interpretations by means of the same verb forms.

(81) a. Mikel barautzen zen astero
Mikel-ABS to fast-IMPF AUX-A every week
'Mikel used to fast every week'

---

[20] See also Etxepare (2003: §4.1.4.5.).

b. Mikelek barau egiten zuen astero
Mikel-ERG fast do-IMPF AUX-A-E every week
'Mikel used to fast every week'

(82) a. Mikel barautu/barautuko da
Mikel-ERG fast-PERF/PERF-FUT AUX-A-E
'Mikel fasted/will fast'

b. Mikelek barau egin/egingo du
Mikel-ERG fast-ABS do-PERF/KO AUX-A-E
'Mikel fasted/will fast'

We distinguish here four classes of verbs derived from event nouns.

2.9.1. Some nouns that express an event not controlled by the subject give rise to verbs with a theme or experiencer case-marked absolutive.

(83) *behaztopa* 'stumble', *irrista* 'sliding' (> 'to slide')

Some of these verbs take an auxiliary and follow an experiencer (DAT)/theme (ABS) pattern.

(84) a. *pilpira* 'heartbeat' (> 'to beat')

b. Untxiari bihotza azkar pilpiratzen zitzaion
rabbit-DAT heart ABS fast beat-IMPF AUX-A-D
'The rabbit's heart was beating fast'

2.9.2. Some event nouns yield verbs having an agent case-marked either absolutive (95a) or ergative (85b).[21]

(85) a. *barau* 'fast', *borroka* 'fight', *hizketa* 'talk', *lasterka* 'running' (> 'to run'), *mintzo* 'speech' (> 'to speak')

b. *arnasa* 'breath' (> *arnastu*), *bidaia* 'journey' (< 'to travel'), *bira* 'turn', *distira* 'sparkle' (> 'to shine'), *hausnar* 'reflection' (> 'to think over').

Odriozola (1993) points out that some (literary) verbs of this class take a clausal complement. Nevertheless they have an intransitive auxiliary.[22]

(86) a. *ahalegin* 'effort' (> 'to try'), *lehia* 'endeavour (> 'to compete'), *oldar* 'attack', *saio* 'try' (> *saiatu)*

[21] See Etxepare (2003).

[22] Besides Odriozola's set of literary verbs, a description of embedded clause complements of Basque verbs can be found in Goenaga (1984).

b. Jon saiatuko da garbi azaltzen
Jon-ABS try-PERF-FUT AUX-A clear explain-IMPF
'Jon will try to explain it clearly'

All the verbs of this class are activities.

2.9.3. Some event nouns yield verbs that take a transitive auxiliary. They are non ingressive achievements or activities and their pattern is cause-agent (ERG)/theme (ABS).

(87) *bataio* 'baptism' (> *bataiatu* 'to baptize'), *epai* 'sentence', *ehiza* 'hunting' (> 'to hunt'), *garraio* 'transport', *zelata* 'spying' (> 'to spy')

2.9.4. Zabala (2002) points out that sometimes the bare noun provides the verb with some features of its own. In fact, it is well known that because of its weakness the light verb somehow needs (one of) the arguments of the noun. However, some event nouns (88a) take a certain noun phrase such as the one case-marked dative in three cases: noun phrases headed by the event noun (88b), complex predicates bearing the even noun (87c), and constructions headed by a verb derived from the even noun (88d).

(88) a. *galde* 'question' (> 'to ask'), *debeku* 'prohibition' (> 'to prohibit') *erregu* 'request'

b. erregua ardi-txakurren jabeei
request sheepdog-GEN owner-DAT
'a request to the sheepdogs' owners'

c. Nik ardi-txakurren jabeei erregu egin diet
I-ERG sheepdog-GEN owners-DAT request do-PERF AUX- A-D-E
etortzeko
come-to
'I request the sheepdogs' owners to come'

d. Nik ardi-txakurren jabeei erregutu diet
I-ERG sheepdog-GEN owners-DAT request-PERF AUX-A-D-E
etortzeko
come-to
'I request the sheepdogs' owners to come'

Thus, derived verbs follow an agent (ERG)/beneficiary (DAT) pattern and they often express punctual achievements.

2.10. Result nouns

Grimshaw (1990: §3.2) and Gràcia (1994) assume that result nouns of both English and Catalan do not have an event structure and therefore, participants that can be seen in the lexical structure of both nouns and derived verbs are not mapped into syntax inside the nominal phrases. In a first approach, we use *result nouns* as the semantic label for nouns such as *otoitz* (89a). We assume that derived verbs are the

causer of the result expressed by the noun that also appears in most of the complex predicates (of the second class) that are available (89b).

(89) a. Alexek otoizten du
Alex-ERG pray-IMPF AUX-A-E
'Alex prays'

b. Alexek otoitz egiten du
Alex-ERG prayer do-IMPF AUX-A-E
'Alex prays'

Finally all derived verbs express activities, as the frequency reading of the quantifier *asko.*

(90) Asko abesten zuen
a lot sing-IMPF AUX-A-E
'He used to sing often'

Syntactical realization of the theme may instantiate a terminative achievement:

(91) Jonek bi minutuan abestu du abestia
Jon-ERG two minutes sing-PERF AUX-A-E song-ABS
'Jon sang the song in two minutes'

We distinguish six classes of result nouns.

2.10.1. It is cross-linguistically well attested that some nouns are incorporated into the derived verb and may sometimes be expressed as themes in the syntax (92b). Some result nouns giving rise to verbs of this class follow an agent (ERG)/(theme (ABS)) pattern.

(92) a. *abesti* 'song' (> *abestu* 'to sing'), *amets* 'dream' (> *amestu*), *dantza* 'dance', *joko* 'game' (> 'to play'), *otoitz* 'prayer' (> *otoiztu* 'to pray)

b. Jonek abesti hunkigarriak abesten ditu
Jon-ERG song touching sing-IMPF AUX-A-E
'Jon sings touching songs'

This subclass of Basque nouns may have a specific interpretation that we will call protocol reading. In fact, songs, dances and so on are a class of actions that license in Basque certain adjuncts that usually modify verbs.

(93) abesti azkarra
song fast
'A fast song'

We assume that nominals express an action that is stipulated and that is susceptible of having an intellectual owner.

2.10.2. Some (abstract) nouns derive verbs with a transitive auxiliary and a special argument structure. Besides the agent (ERG), the theme (ABS), which is often syn-

tactically realized, has the same reference as the nominal base itself. They express punctual achievements.

(94) a. *aholku* 'advice' (> *aholkatu* 'to advise'), *asmo* 'intention' (> *asmatu* 'to guess), *aukera* 'choice', *mailegu* 'loan' (> *mailegatu* 'to borrow'), *usain* 'smell'

b. Euskarak *tu* mailegatu zuen latinetik
Basque-ERG tu borrow-PERF AUX-A-E latin-from
'Basque borrowed *tu* from Latin'

c. Euskarak *tu* mailegua hartu zuen latinetik
Basque-ERG tu loan borrow-PERF AUX-A-E latin-from
'Basque took *tu* from Latin'

2.10.3. Group nouns are also to understood as the result of the event denoted by their derived verbs.

(95) *bilduma* 'collection', *meta* 'stack', *multzo* 'bunch' (> 'to bunch together'), *pila* 'pile', *zerrenda* 'list'

The transitive auxiliary follows an agent (ERG)/theme (ABS) pattern, in which theme is a unity of the group expressed by the base.

(96) Anek pinak bildumatzen ditu
Ana-ERG pin-ABS collect-IMPF AUX-A-E
'Ane collects pins'

When the plural in a noun phrase has a generic reading (97a), it expresses an activity, but non-generic readings and some inflection forms of the verb license an achievement (97b).

(97) a. Jantzi zaharrak pilatzen ditu
clothes old-ABS pile-IMPF AUX-A-E
'He piles old clothes'

b. Iazko jantzi zaharrak komunean pilatu ditu
last year-*from* clothes old-ABS bathroom-in pile-PERF AUX-A-E
'He piled last year's old clothes in the bathroom'

2.10.4. Most of the result nouns express a part of the whole that corresponds to the theme of the derived verb. An auxiliary alternation follows an (agent-cause (ERG))/theme (ABS) pattern and the nouns exhibit an (inalienable) possession relationship with the theme. They take the set of aspect readings of ingressive (98a) and other kinds of achievements (98b).

(98) a. *bular* 'chest' (> 'to sag'), *gihar* 'muscle' (> 'to become brawny'), *gorputz* 'body' (> 'to embody'), *ildaska* 'small groove'

b. *amai* 'end' (> 'to finish'), *azpimarra* 'underlining' (> 'to underline') *era* 'form' '(> 'to establish'), *itxura* 'shape' (> 'to get/give shape'), *sustrai* 'root'

2.10.5. Most of the result nouns express something that belongs to the theme of the derived verb. The pattern is causer-agent (ERG)/theme (ABS) and terminative (99a) or ingressive (99b) achievements are expressed.

(99) a. *arau* 'rule' (> 'to set the rules'), *hesi* 'fence', izen 'name' (> *izendatu* 'to designate'), *koroa* 'crown' (> 'to coronate'), *margo* 'colour' (> 'to paint'),

b. *bake* 'peace' (> 'to bring peace', *korapilo* 'knot' (> korapilatu), *lanbro* 'fog' (> 'to fog over), *oreka* 'balance', *zorabio* 'dizziness' (> *zorabiatu*)

2.10.6. There are several Basque result nouns (100a) that express something that affects an argument of the verb advantageously or adversely.

(100) a. *babes* 'protection' (> 'to protect'), *baldintza* 'condition', *berme* 'guarantee' (> *bermatu*), *muga* 'limit', *oztopo* 'obstacle' (> *oztopatu* 'to hinder'), *zigor* 'punishment' (> 'to punish')

b. Zigorra Konstituzioaren aurkako ideiei
Punishment Constitution-GEN against idea-DAT
'Punishment for ideas against the Constitution'

c. Konstituzioaren aurkako ideiak zigortu nahi dituzte
Constitution-GEN against ideas punish-PERF want AUX-A-E
'They want to punish ideas against the Constitution'

3. Verbs derived from adjectives

In Basque, the productivity of process that derives verbs from adjectives is similar to the one deriving verbs from nouns. In fact one could say that almost any adjective can potentially be involved in a spontaneous derivation resulting in a verb.

Let us see first some of the features common to all verbs derived from adjectives.

a) Almost all of them may be paraphrased by 'to (make) become A'.
b) The pattern is not very variable.

b1) In most cases the pattern is (agent-cause (ERG))/theme-experiencer (ABS). We assume that themes (101a) and experiencer (101b) are inherited from the base.

(101) a. *estu* 'narrow', *labur* 'short' (> 'to become shorter/to shorten'), *txiro* 'poor' (> 'to impoverish'/'to become poor'), *zabal* 'broad' (> 'to spread/broaden'), *zahar* 'old' (> 'to grow old')

b. *alai* 'happy' (> 'to cheer, to make happy'), *harro* 'proud' (> 'to become/make proud'), *haserre* 'angry' (> 'to get angry/to anger') 'sad', *triste* 'sad' (> *tristatu* 'to become sad/to sadden'), *zoro* 'crazy' (> 'to go/drive crazy')

It should be noted that the items in (118b) are psychological verbs similar to the verbs derived from nouns with argument structure (→§2.1)

b2) There are a few that follow experiencer-theme (ABS) (102) or cause-agent (ERG)/experiencer-theme (ABS) (103) patterns.

(102) *axolagabe* 'careless' (> 'to be neglectful'), *ausart* 'brave' (> 'to dare'),

(103) *erraz* 'easy' (> 'to facilitate'), *ezagun* 'well-known' (> *ezagutu* 'to know'), *ezgai* 'unable' (> 'to disqualify')

b3) Etxepare (2003: §2.9.) points out that some Basque verbs may reflexivize by detransitivizing the auxiliary. Most of these verbs are derived from adjectives in Basque:

(104) *apain* (> 'to dress well, to tidy up'), *txukun* 'smart' (> 'to make smart'), *garbi* 'wash', *lehor* 'dry'

(105) a. Josebak Xixka garbitu du
Joseba-ERG Xixka-ABS clean-PERF AUX-A-E
'Joseba washed Xixka'

b. Joseba garbitu da
Joseba-ABS clean-PERF AUX-A
'Joseba washed himself'

c) A few of the verbs derived from adjectives have complex predicate counterparts of the first type, in which the bare adjective gives at least one argument to the complex predicate.

(106) a. *bero* 'hot' (> 'to heat, to get hot'), *nabari* 'patent' (> 'to be apparent', *oker* 'mistake' (> 'to make a mistake'), *zilegi* 'allowable' (> 'to be allowable')

b. Berotu naiz
get hot-PERF AUX-A
'I got hot'

c. Bero naiz
hot AUX-A
'I am hot'

Complex predicates have available a single state reading, whereas derived verbs have the set of readings common to the changes of state (→§2.1).

d) However, most of the verbs derived from adjectives are related to non-state achievements:

d1) In most cases, adjectives derive verbs that designate an ingressive achievement. Their paraphrase is 'to (make) become (more) A'.

(107) *ahul* 'weak' (> 'to weaken'/to grow weak'), *bigun* 'soft' (> 'to soften/to grow soft'), *lodi* 'fat' (> 'to (make) get fat'), *txukun* 'smart' (> 'to make smart'), *zurbil* 'pale' (> 'to (make) get pale')

Adjectives of this type may take a quantifier suffix (108b), and this complex can also forms the base for deriving further verbs of this type (109). This morphological process is available for all the adjectives of this type, although only the derived verbs of some adjectives appear in the dictionary (109b).[23]

(108) a. Luzea da
long is
'It is long'

b. Luzeagoa da
Long-more is
'It is longer'

(109) a. *luze* 'long', *luzeago* 'longer', *luzeagotu* 'to lengthen still more'
b. *bizi* 'living' (> 'to intensify'), *txiki* 'little' (> 'to grow smaller'), *larri* 'critical' (> 'to aggravate'), *sendo* 'strong' (> 'to make stronger')

d2) Punctual achievements are also common.

(110) *antzu* 'sterile' (> 'to become sterile/'to sterilize'), *bikoitz* 'double' (> 'to duplicate'), *isil* 'silent' (> 'to silence'), *oker* 'mistaken' (> 'to make a mistake'), *ezkutu* 'hidden' (> *ezkutatu* 'to hide').

d3) There are some verbs that can take the set of aspect readings of terminative achievements:

(111) *agor* 'dry' (> 'to dry up), *berri* 'new' (> 'to renovate'), *gai* 'able' (> 'to enable'), *xahu* 'wasted (> 'to spend')

Kratzer's (1988) two classes of predicates hold for both adjectival bases and derived verbs. On the one hand individual-level adjectives express non-transitory properties and stage-level adjectives are related to transitory properties. On the other hand, changes of states expressed by derived verbs may be interpreted as the acquisition of non-transitory (→§3.1), and transitory (→§3.2) properties.

3.1. Derived verbs related to individual-level predicates

Basically, stage-level predicates projected by adjectives appear with the verb *egon* 'to be' (112a) whereas individual-level predicates (→§3.2) take *izan* 'to be' (112b). In most cases adjectives expressing stage-level predicates can be bare, whereas individual-level predicates bear the suffix *-a*.[24]

(112) a. haserre dago
anger is
'He is angry'

b. aberatsa da
rich-*a* is
'He is rich'

However, most adjectives can project both stage and invidual-level predicates.

---

[23] Lexical entries for colours are both nouns and adjectives in Basque and have available this kind of derivation.

[24] See a more accurate description in Zabala (2003a).

(113) a. *ahul* 'weak' (> 'to weaken', *garbi* 'clean', *isil* 'silent' (> 'to (be) silent'), *lodi* 'fat' (> 'to (make) get fat'), *txukun* 'tidy' (> 'to 'tidy'), *zurbil* 'pale' (> 'to (make) get pale')

b. lodi dago
fat is
'He is fat (now)'

c. lodia da
fat-*a* is
'He is fat'

Acquisition of individual-level properties is expressed by verbs derived either from individual-level adjectives (114) or from adjectives that can project both classes of predicates (113).

(114) a. *antzu* 'sterile' (> 'to become sterile/to sterilize'), *bikoitz* 'double' (> 'to duplicate'), *eder* 'beauty' (> 'to beautify'), *egoki* 'suitable' (> 'to adjust/to suit/to customize')

b. Antzutu da
become sterile-PERF AUX-A
'She became sterile'

The acquisition of individual-level properties may also be expressed sometimes by the adjective and the copula *egin* 'to become'.

(115) Antzu egin da
sterile become-PERF AUX-A
'She became sterile'

Finally, notice that individual-level properties are expressed as in (116) with the adjective bearing the suffix *-a* and the copula *izan* 'to be'.

(116) Antzua da
'She is sterile'

It must be pointed out that Basque has a very wide range of items that belong to both adjectival and nominal categories.

(117) a. *alargun* 'widow' (> 'to become a widow'), *arrotz* 'stranger' (> 'to become strangers') *ero* 'madman' (> 'to go/drive mad'), *euskaldun* 'Basque-speaker' (> 'to make/become a Basque-speaker'), *etsai* 'enemy, hostile' (> 'to become an enemy')

b. Begoña euskaldun egin da
Begoña-ABS Basque-speaker become-PERF AUX-A
'Begoña became a Basque speaker'

c. Begoña-ABS euskaldundu da
Begoña-ABS become Basque-speaker-PERF AUX-A
'Begoña became a Basque speaker'

These nominals pattern with both adjectives deriving verbs in (113-114) and verbs deriving nouns (→§2.2).

(118) a. Apaiz egin da
Priest become AUX-A
'He became a priest'

b. Apaiztu da
become a priest-PERF AUX-A
'He became a priest'

c. Apaiza da
priest-*a* is
'He is a priest'

In short, all these predicates express (the acquisition of) non-transitory properties.

3.2. Derived verbs related to stage-level predicates

Acquisition of stage-level properties is expressed by verbs derived from stage-level adjectives:

(119) *ados* 'agreed' (> 'to concur/to come to an agreement'), *haserre* 'angry' (> 'to get angry/to anger'), *haurdun* 'pregnant' (> 'to become pregnant'), *jelos* 'jealous' (> 'to become jealous'), *oker* 'mistaken' (> 'to make a mistake')

The acquisiton of transitory properties is expressed by both the derived verb and the stage-level adjective with the copula *jarri* 'to put, to become'.

(120) a. Haserretu da
anger-PERF AUX-A
'She got angry'

b. Haserre jarri da
angry get-PERF AUX-A
'She got angry'

There are some derived verbs that do not correspond to the acquisition of the properties expressed by the adjective. Their paraphrases correspond to secondary predicates, which are not necessary for the grammaticality of the clause, and they take verbs different from *izan, egon, egin* and *jarri.*

(121) a. *nabari* 'obvious' (> 'to realize'), *nabarmen* 'clear' (> 'to make clear'), *ozen* 'loud' (> 'to make loud'), *zabal* 'broad' (> 'to broaden'), *zehatz* 'exact' (> *zehaztu* 'to stipulate clearly')

b. Nabaritu du Iñaki gose zegoela
notice-PERF AUX-A-E Iñaki-A hungry was-*ela*
'He noticed that Iñaki was hungry'

c. Nabari ikusi du Iñaki gose zegoela
obvious see-PERF AUX-A-E Iñaki-A hungry was-*ela*
'He saw clearly that Iñaki was hungry'

Notice that all of the verbs in (121) select an embedded clause. In fact, the adjective predicate is directed to the embedded clause (121b) that was not an argument of the adjectival base. On the other hand, Odriozola (1993) has described a set of verbs derived from (psychological) adjectives that select embedded clauses of several classes, and that are lexicalized as aspectual or modal verbs. These verbs do inherit the argument of the adjective:

(122) *ausart* 'brave' (> 'to dare'), *etsi* 'desperate' (> 'to resign')

4. Verbs derived from adverbs

Basque exhibits a very productive process that derives verbs from adverbs bearing the suffix *-ka* (→§4.1). In addition, there are also some other adverbs (→§4.2, §4.3) resulting from a process that seems to be no longer productive in modern-day Basque.

4.1. Adverbs with *-ka*

Basque verbs may be derived from adverbs ending in *-k*a, a suffix of several meanings. Zabala (1993: §2.2.2.1.3., 2.2.2.3.4., 2.4.) claims that these adverbs have available the projection of nominal predicates, since in some cases they do not modify the verb but rather predicate about a subject (123a). However, it should be noticed that these items may modify the verb as adverbs do (123b).

(123) a. Xixka horzka dago
Xixka tooth-*ka* is
'Xixka is snapping'

b. Xixka horzka defenditu da albaitariarengandik
Xixka-ABS tooth-*ka* to defend-PERF AUX-A
'Xixka defended herself from the veterinarian by snapping'

The adverbs themselves are derived almost spontaneously from a wide range of nouns. Almost all of the verbs take a transitive auxiliary and an agent (ERG)/theme (ABS) pattern. The nouns and adverbs that give rise to verbs are the following:

4.1.1. Instrument nouns yield adverbs with the meaning 'using N'. The derived verbs paraphrase as 'to use N to beat something' and they denote an iterative activity.

(124) a. *harri* 'stone', *mailu* 'hammer', *ezten* 'sting', *ziri* 'stick'

b. harrika daude
stone-*ka* are
'They are throwing stones'

c. Haiek emakumea harrikatu zuten
They-ERG woman-ABS to stone-PERF AUX-A-E
'They stoned the woman'

Notice that the argument inherited from the adverb is the agent.

4.1.2. Derivation from body-part nouns ends up with similar verbs.

(125) a. *adar* 'horn', *moko* 'beak', *hortz* 'tooth'

b. Ahuntza adarka dabil
goat horn-*ka* is
'The goat is goring/butting (something)'

c. Ahuntzak Xixka adarkatu zuen
goat-ERG Xixka-ABS to gore-PERF AUX-A-E
'The goat gored Xixka'

4.1.3. Adverbs derived from event nouns mean 'to be doing N'. The verb's paraphrase is 'to do N once and again', which also corresponds to an iterative activity. The pattern is similar to that of instrument and body-part nouns.

(126) a. *aldarri* 'clamour' (> 'to proclaim', *kolpe* 'knock' (> *kolpatu* 'to beat'), *musu* 'kiss', *oihu* 'shout', *txalo* 'applause' (> 'to applaud')

b. Musuka dabil
kiss-*ka* is
'She is kissing'

c. Argik Doro musukatu zuen
Argi-E Doro-A to cover with kisses-PERF AUX-A-E
'Argi covered Doro with kisses'

4.1.4. Classification nouns give rise to verbs rather different from the ones above. They follow an agent (ERG)/theme (ABS) pattern but the argument inherited from the adverb is a theme. The aspect reading is related to this argument that must be plural and denotes some limit of the event. The verbs are in fact terminative achievements.

(127) a. *sail* 'series', *lerro* 'line', *maila* 'level' (> 'to divide up into levels'), *multzo* 'bunch' (> 'to accumulate'), *txanda* 'turn' (> 'to turn with'), *zati* 'piece' (> 'to split')

b. Liburuak sailka daude
book-ABS series-*ka* are
'The books are (classified) in series'

c. Jonek liburuak sailkatu ditu
Jon-ERG book-ABS to-classify-PERF AUX-A-E
'Jon classified the books'

On the other hand, verbs in (127) lexicalize the movement itself, whatever a movement is. Other verbs above lexicalize the kind of movement. In any case, all derived verbs from adverbs with *-ka* denote a (kind of) movement in the sense of Demonte (1994).

Although there are some (dialectal) exceptions, generally one-argument verbs cannot be derived from this kind of adverb.

(128) a *algaraka* 'laughing', *apurka* 'little by little', *biraka* 'spinning', *iheska* 'fleeing', *zuka* 'using the more formal speech'

b. **algarakatu, *apurkatu, *birakatu, *iheskatu, *zukatu,*

4.2. Locational adverbs

Basque has several lexicalized adverbs[25] that yield a quite productive derivational process. They exhibit auxiliary alternation and an ((agent-cause (ERG))theme (ABS) pattern. On the one hand we have adverbs that consist of a suffix *-ti* attached to a relational location noun (130). Verbs derived from them designate the movement related to an achievement. The pattern is agent-cause (ERG)/theme (ABS) and the argument inherited is a theme.

(129) *behe* 'ground', *gain* 'top', *goi* 'above'
beheiti 'down', *gaindi* 'through', *goiti* 'on high'
beheititu 'to lower', *gainditu* 'to overcome', *goititu* 'to lift up'

Adverbs bearing the (lexicalized) postposition *-ra* 'to' (→§5) also give rise to verbs of a similar kind:

(130) *atzera* 'back', *aurrera* 'forward', *gora* 'up', *kanpora* 'out'
atzeratu 'to put back', *aurreratu* 'to advance/to overtake', *beheratu* 'to lower', *kanporatu* 'to throw out'

4.3. Non-derived adverbs

In Basque there are also a few verbs derived from adverbs of other kinds:

(131) *berandu* 'late' (> 'to delay'), *maiz* 'often' (> 'to do something often')

5. Verbs derived from postpositional phrases

Postpositional phrases headed by two particular suffix postpositions may derive verbs almost spontaneously. The postpositions are *-ra* (directional) 'to' (→§5.1) and *-z* 'with/by' (→§5.2.)

5.1. Postpositional suffix *-ra* (directional) 'to'

The verbal ending *-tu* can be adjoined to postpositional phrases headed by the postpositional suffix *-ra* 'to'. We will classify derived verbs mainly according to the class of the nominal base.

5.1.1. Nouns expressing either parts of the human body or sentiments produce a set of psychological verbs in which movement is rather metaphoric. They are punctual-achievement verbs.

(132) *kontu* 'matter', *kontura* 'to the matter'
konturatu 'to realize'

(133) *gogo* 'mind' (> 'to remember'), *buru* 'head' (> 'to occur')

---

[25] See Odriozola (1999, 2002).

The verb's meaning is not predictable from the postpositional phrase. However, these verbs are closely related to their bases. In fact, postpositional phrases are part of complex predicates whose argument structure is inherited by the derived verb.

(134) a. Joni gogora etorri zaio Joseba
Jon-DAT mind-to come-PERF AUX-A-D Joseba-ABS
'Joseba came to Jon's mind'

b. Jonek Joseba gogoratu du
Jon-ERG Joseba-ABS remember-PERF AUX-A-E
'Jon remembered Joseba'

We assume that they all adhere to an experiencer (ABS) or (ERG)/proposition pattern. They select either an NP or an embedded clause. *Konturatu* 'to realize' takes the intransitive auxiliary, whereas *gogoratu* 'to remember' may take either an intransitive or transitive auxiliary without changing the argument structure.

(135) Joseba konturatu da Jon haserre dagoela
Joseba-ABS to realize-PERF AUX-A Jon angry is-that
'Joseba has realized that Jon is angry'

(136) a. Nik gogoratu dut Dory oraindik gaixorik
I-ERG to remember-PERF AUX-A-E Dory-ABS still ill
dagoela
is-that
'I remembered that Dory is still ill'

b. Ni gogoratu naiz Dory oraindik gaixorik
I-ABS to remember-PERF AUX-A Dory-ABS still ill
dagoela
is-that
'I remembered that Dory is still ill'

Bururatu may take an experiencer case-marked with dative:

(137) Josebari bururatu zaio Durangora joatea
Joseba-DAT to come in mind-PERF AUX-A-D Durango-to to go-*tzea*
'It occurred to Joseba to go to Durango'

5.1.2. Sometimes non-relational location nouns produce verbs that express a true movement. The pattern is agent-cause (ERG)/theme (ABS) or (agent-cause (ERG))/ theme (ABS). They show the set of aspect readings of achievements or accomplishments.

(138) *aho* 'mouth'
ahora 'to the mouth'
ahoratu 'to raise something to the mouth'

(139) *espetxe* 'prison' (> 'to imprison'), *etxe* 'house' (> 'to take home'), *hondo* 'bottom' (> 'to sink'), *itsaso* 'sea' (> 'to set out to sea'), *lur* 'earth' (> 'to land')

Some of them are quite lexicalized.

(140) *azal* 'surface' (> 'to appear, to show up'), *bide* 'road' (> 'to guide'), *esku* 'hand' (> 'to grab')

5.1.3. Section 3.1. shows that the postpositional suffix *-ra* attaches to a relational location noun. These constructions have undergone a lexicalization process to form an adverb. This process seems to be just one of the subcategory features of these relational locations, but in some cases, postpositional phrases are not lexicalized, i.e. they do not appear in Basque dictionaries and they have a predictable meaning. All of them can function as the base of a derivation process that ends up in a verb.

(141) *albo* 'side', *albora* 'to the side'
erdi 'middle', *erdira* 'to the middle'
azpi 'bottom', *azpira* 'to the bottom'
alboratu 'to put aside', *azpiratu* 'to bring down', *erdiratu* 'to put in the middle'

5.1.4. Some non-locational nouns yield verbs that, by means of lexicalization, do not express movement at all:

(142) *argi* 'light' (> *argitaratu* 'to bring to light/to publish'), *begi* 'eye' (> 'to look'), *egun* 'today' (> 'to update'), *plaza* 'square, public life' (> 'to make public')

5.2. Postpositional suffix *-z*

Some verbs are derived from postpositional phrases headed by the postpositional suffix *-z* '(instrumental) by/with'. They all take a transitive auxiliary and follow an agent-cause (ERG)/theme (ABS) pattern. The derivational bases are nouns that sometimes are expressed in the syntax as PP headed also by *-z*:

(143) Jonek eta Josebak urre zaharraz urreztatu nahi dituzte
Jon-ERG and Joseba-ERG gold old-*z* to gold-plate-PERF AUX-A-E
beren belarritakoak
'Jon and Joseba want to gold-plate their earrings with old gold'

The derivational nouns are shadow themes that yield two classes of verbs.

5.2.1. Material nouns give rise to ingressive achievement verbs: on the one hand the quantity of material itself has no limit, and on the other hand, theme has no constrictions on the quantity of material that may be accepted. The paraphrase is either 'to N-plate' (144a) or 'to provide N' (144b).

(144) a. *altzairu* 'iron' (> *altzairuztatu*[26] 'to iron-plate'), *berun* 'lead', *eztainu* 'tin', *urre* 'gold', *lar* 'silver'

b. *aire* 'air' (> 'to ventilate'), *azukre* 'sugar', *koipe* 'grease', *ongarri ur* 'water'

[26] Almost all the verbs derived form posposition phrases headed by *-z* have the infix *-ta-* between the postposition phrase and *-tu*.

5.2.2. *-Z* adjoins to countable nouns. The postpositional phrase gives rise to non-ingressive achievements. In most cases theme may accept only one of the entities expressed by the shadow argument. Most of them are neologisms. The paraphrase is 'to provide with N'.

(145) *azal* 'cover' (> *azaleztatu* 'to bind'), *ohol* 'board' (> *oholeztatu* 'to board up'), *paper* 'paper' (> *papereztatu*)

References

Alberdi, X., 2003, "The transitivity of borrowed verbs in Basque: an outline", in this volume, 23-45, UPV/EHU.

Alcazar, A., 2002, *Aspectual interpretation in Basque.* MA Research Papers, USC.

Artiagoitia, X., 1995, *Verbal projection in Basque and minimal structure.* Supplements of the *ASJU-International Journal of Basque Linguistics and Philology,* 37.

—, 2000, *Hatsarreak eta parametroak.* Euskal Herriko Unibertsitatea eta Arabako Foru Aldundia, Gasteiz.

—, 2003, "The case of an enlightening, provoking and admirable Basque derivational suffix with implications for the theory of argument structure", in this volume, 147-183.

De Miguel, E., 1999, "El aspecto léxico". *Gramática descriptiva de la Lengua Española.* Espasa. Madrid.

Demonte, V., 1994, "La semántica de los verbos de cambio". *Papeles de trabajo. Lingüística Teórica. Cuadernos de Lingüística II.* Instituto Universitario Ortega y Gasset.

Etxepare, R., 2003, "Valency and the auxiliary system. An outline". *A grammar of Basque.* J. Hualde & J. Ortiz de Urbina (eds.), Mouton de Gruyter, Berlin & N.Y.

Fernández, B., 1997, *Egiturazko kasuaren erkaketa euskaraz.* Euskal Herriko Unibertsitateko Argitalpen Zerbitzua, Leioa.

Goenaga, P., 1984, *Euskal sintaxia; konplementazioa eta nominalizazioa.* Ph.D., dissertation, EHU/UPV, Gasteiz.

Gràcia, L., 1994, *Estructura argumental i herència en morfologia.* Universitat de Girona.

—, *et al.,* 2000, *Configuración morfológica y estructura argumental: léxico y diccionario.* Euskal Herriko Unibertsitateko Argitalpen Zerbitzua, Leioa.

Grimshaw, J., 1990, *Arguments Structure.* The MIT Press. Cambridge, Massachussets and London.

Haddican, B., 2001, *Basque Functional Heads,* manuscript. University of New York.

Hualde, J.I., 2003, "Sinthetic verbs" and "Nonfinite forms", in J.I. Hualde and J. Ortiz de Urbina (eds.), *A Grammar of Basque,* Mouton de Gruyter, Berlin & N.Y.

Kratzer, A., 1988, *State-Level And Individual-Level Predicates.* University of Massachussets at Amherst.

Lafitte, R., 1943, *Grammaire basque (Navarro-Labourdin littéraire).* Elkar. Donostia, 1979.

Lafon, R., 1943, *Le système du verbe basque au XVIe siècle.* Elkar. Donostia, 1980.

Laka, I., 1993a, "The structure of inflection: A case study on X° Syntax", in J.I. Hualde & J. Ortiz de Urbina (eds.): *Generative Studies in Basque Syntax:* 21-70. Amsterdam, John Benjamins.

—, 1993b, "Unergatives that assign Ergative". In Bobaljik, J.; Philips, C. (eds.): *Papers on Case & Agreement 1. MIT Working Papers in Linguistics* 18: 149-172.

—, 1995a, "Thetablind case: Burzio's generalization and its image in the mirror". *Workshop on Burzio's Generalization.* Eric Reuland (ed).

—, 1995b, *A Brief Grammar of Euskara, The Basque Language.* http:/www.ehu.es/grammar.

Miguel, E. de, 1999, "El aspecto léxico". *Gramática descriptiva de la Lengua Española.* Real Academia de la Lengua Española. Espasa, Madrid.

Minkoff, S.A., 1997, "Argument structure and animacy entailment". *Theoretical issues at the morphology-syntax interface.* Supplements of ASJU. A. Mendikoetxea & M. Uribe-Etxebarria (eds.).

Mitxelena, K., 1961, *Fonética histórica vasca.* Gipuzkoako Foru Aldundia. 2th ed., Donostia 1976.

Mujika, L.M., 1982, *Latina eta erromanikoaren eragina euskaran. Euskal lexikoaren azterketa bideetan.* Sendoa S.A. argitaldaria. Donostia.

Odriozola, J.C., 1993, *Axularren kasu markadun adizkiak.* Euskal Herriko Unibertsitateko Argitalpen Zerbitzua.

—, 1999, "Postposizioak EGLUn: egitura motak eta gramatika-kategoriak", *Euskera* 44: 2, 243-265.

—, 2002, "Fronteras composición-prefijación y composición sintaxis: construcciones con nombres adverbiales en vasco" (manuscript), UPV/EHU.

Oyharçabal, B., 2003, "Lexical causatives and causative alternation in Basque", in this volume, 221-253.

Pustejovsky, J., 1995, *The Generative Lexicon.* The MIT Press, Cambridge, Massachussets.

Verkuyl, H.J., 1989, "Aspectual classes and aspectual composition", *Linguistics and Philosophy* 12: 39-94.

Zabala, I., 1993, *Predikazioaren teoriak gramatika sortzailean (euskararen kasua).* Ph.D., UPV/EHU.

—, 2002, "Predicados complejos en vasco" (manuscript).

—, 2003a, "Nominal predication: copulative sentences and secondary predication", in J.I. Hualde and J. Ortiz de Urbina (eds.), *A Grammar Basque,* Mouton de Gruyter.

—, 2003b, "Causation and semantic control. Diagnosis of incorrect uses in minorized languages", in this volume, 255-284.

—, & J.C. Odriozola, 1994, "Adjektiboen eta adberbioen arteko muga zehatzik eza", *ASJU-International Journal of Basque Linguistics and Philosophy* 28 (2): 525-541.

—, & —, 1996, "On the relation between DP and TP. The structure of Basque Infinitivals". *Catalan Working Papers in Linguistics* 5 (2): 231-281.

LEXICAL CAUSATIVES AND CAUSATIVE ALTERNATION IN BASQUE

B. Oyharçabal

(IKER-UMR 5478, CNRS)

Abstract

After offering a brief survey of the features of causative sentences in Basque, mainly on the basis of Dixon's (2000) criteria, the paper deals with Basque lexical causatives, which can be used as either causative or unaccusative verbs. The proposed analysis assumes that lexical decomposition is carried out directly according to syntactic principles (Hale & Keyser 1993, Baker 1997, McGinnis 2000), and that different types of causative sentence (morphological vs lexical causatives) correspond to different types of phrase (VoiceP vs VP) selected by the Cause head (Pylkkänen 2001, 2002; Meggerdoomian 2002). The paper shows that in Basque lexical causatives the Cause head selects one of the predicates BECOME *or* GO *only. Other intransitive verbs are excluded from lexical causativization, even those which are superficially similar verbs of change because they are absolutive monadic verbs (reflexive verbs like* orraztu *'comb', verbs of happening like* gertatu *'happen', or verbs of activity like* jostatu *'play'). Three types of lexical causative are distinguished and analyzed following lexical decomposition: verbs of change of (physical) state, verbs of change of place and psychological causatives. Since Basque, unlike Finnish or Japanese, shows a strict correlation between causation and the existence of an external argument, it is assumed that in Basque as in English, the Cause and Voice heads conflate in lexical causatives (Pylkänen 2002).*

There are two main ways to form causative verbs in Basque, illustrated in (2a) and (2b):[1]

---

[1] Abbrevations. ABS: absolutive, ART: article, AUX: auxiliary, CAU: causative, COM: comitative, DAT: dative, ERG: ergative, FOC: focus marker, FUT: future, IMP: imperfective, INE: inesive, INS: instrumental, INTER: interrogative, PL: plural, PAR: partitive, PTP: participle, RFL: reflexive. Finite verb forms such as the auxiliaries [AUX] *da* and *du* incorporate indices for the person and number of the verb's nuclear arguments, which in the literal glosses are indicated to the right of a colon by means of English personal pronouns in the order subject > (direct or indirect) object, regardless of the sequence of morphemes in the Basque forms. Where glosses for finite forms are followed by *three* personal pronouns, the third index represents the dative complement (indirect object). Basque makes no grammatical distinction for the gender of third-person arguments.

(1) *Katua hil da*
cat.ABS die AUX:3SG
"The cat died."

(2a) *Haurrak katua hil du*
child.ERG cat.ABS die AUX:3SG.3SG
"The child killed the cat."

(2b) *Haurrak katua hilarazi du*
child.ERG cat.ABS die.CAU AUX:3SG.3SG
"The child caused the cat to die" or "The child had the cat killed."

In (1) (2a) and (2b) above the noun phrase *katua* "(the) cat" is in the absolutive case, which is zero-marked in Basque; this case identifies both subjects of intransitive verbs and direct objects of transitive verbs. In the above examples the verb *hil* "die" occurs with either an intransitive (1) or a transitive (2) auxiliary. Notice that the DP *katua,* in the absolutive case, keeps the same theta-role throughout, that of undergoer of the change-of-state expressed by the verb, even though it has the syntactic functions of subject in (1) and object in (2). In (1) *hil* occurs as a monadic verb. The noun phrase *katua,* which appears as object and immediate internal argument in (2a,b) has moved to subject position in (1). In (2a,b) the same process is expressed as in (1), namely the death of the cat, with the difference that the *causer* is specified. The *causation* is not of the same kind in (2a) and (2b) and is expressed in different ways. What both sentences have in common is that the subject has done something to bring about the cat's death.

In this article I will use *causer* and *causation* in the way just illustrated and will refer to verbs of the kinds seen in (2a) and (2b) as *causative verbs.* Following Comrie's (1989) typology, the verbs and sentences in (2a) and (2b) will be called *lexical causatives* and *morphological causatives* respectively. In both cases, the verb *hil* "die" is the *base.* The *lexical causative alternation* between (1) and (2a) is the subject of this paper.[2]

1. Features of causative sentences

According to Dixon's (2000) list of criteria for classifying causative formations, Basque causatives can be characterised with regard to three features: (a) the verb base's aspect; (b) its syntactic type; (c) indirectness of the causer's influence. In this introduction I shall begin with a general overview of causative sentences in Basque in which I examine these characteristics of Basque causatives, before moving on to the main subject of the article.

1.1. Dixon's first criterion refers to whether or not the verb base may be a stative verb. This is relevant in Basque not just as a means of classifying causative structures but because Basque does not allow the lexical or morphological formation of causatives

[2] I won't discuss causative verbs including the causative preroot affix *-ra-*, because it is no more productive. I will also exclude from this study control verbs like *laga* or *utzi* 'let' and *behartu* 'compel, oblige', which do not concern us here.

from stative predicates such as **edun* or *eduki* "have", predicate adjective, noun or postpositional phrase + *izan* or *egon* "be", *-tan* or *-tzen jakin* "know" (how to do something), etc., as the following examples show:[3]

(3a) **Otoitzek saindu/ parabisuan izan(arazi)ko zaituzte*
prayers.ERG saint/ paradise.INE be.(CAU).FUT AUX:3PL.2SG
"Prayers will cause you to be a saint/in paradise."

(3b) **Semeari euskaraz (mintzatzen) jakin(arazi) diot*
son.DAT Basque.INS (speak.IMP) know.(CAU) AUX:1SG.-.3SG
"I caused my son to know how to speak Basque."

(3c) **Dirua ukan(arazi) dizut*
money.ABS have.(CAU) AUX:1SG.3SG.2SG
"I caused you to have money."

In the preceding examples, stative predicates are placed in a causative structure and result in ungrammatical sentences. Basque allows the use of a transitive construction with certain stative predicates, such as copulative predicates; but such sentences, which Rebuschi (1984) calls implicative, are not interpreted as causatives:

(3d) *Lankidea aitzinean dut* (or *daukat*)
colleague.ABS in.front have:1SG.3SG
"I have the colleague in front", i.e. "My colleague is in front of me."

(3e) *Lankidea aspaldiko adiskidea dut*
colleague.ABS old friend.ABS have:1SG.3SG
"I have the colleague (as) an old friend", i.e. "My colleague is an old friend."

(3f) *Lankidea eri dut*
colleague.ABS ill have:1SG.3SG
"I have the colleague ill", i.e. "My colleague is ill."

(3g) *Lankidea hotzak hila dut (daukat)*
colleague.ABS cold.ERG dead.ABS have:1SG.3SG
"I have the colleague dead of cold", i.e. "My colleague is freezing."

These are derived by the addition of an external argument (a surface subject, labelled ergative) from copular sentences with predicates in the forms: postpositional phrase + copula (3d), noun phrase + copula (3e), adjectival phrase + copula (3f-g). The presence of this ergative argument triggers replacement of the copula by the transitive verb glossed "have", but the results are not interpreted as causatives.

1.2. The second criterion from Dixon's typology to be considered can be formulated as the question: Can the base verb be transitive? This is relevant to Basque be-

[3] The aspectual restriction linked to causation has been established by Dowty (1979). However, this view has been questioned; see Pylkkänen (1999) for an analysis of causative derivation with stage-level stative verbs in Finnish.

cause it turns out that lexical causatives cannot be derived from a transitive base, but morphological causatives can, as shown by the following examples:

(4a) *Autoa garajean sartu dut*
car.ABS garage.INE put.in AUX:1SG.3SG
"I put the car in the garage."

(4b) *Autoa garajean *sartu/ sarrarazi didazu*
car.ABS garage.INE *put.in/ put.in.CAU AUX:2SG.3SG.1SG
"You made me put the car in the garage."

(4c) *Sagarra jan dut*
apple.ABS eat AUX:1SG.3SG
"I ate the apple."

(4d) *Sagarra *jan/ janarazi didazu*
apple.ABS *eat/ eat.CAU AUX:2SG.3SG.1SG
"You made me eat the apple."

When we want to put the transitive sentences (4a) and (4c) into a causative construction, only the morphologically derived causatives *sarrarazi* and *janarazi* are available; the transitive base forms *sartu* and *jan* cannot acquire causative meanings.[4]

1.3. The third of Dixon's criteria that we shall consider asks whether the causer's influence is indirect or direct. This point is easily confused with extralinguistic issues, for causality in the real world resembles a chain at the end of which it is always possible to attach a further link (Danlos 2001). But as Dixon observes, this question is highly relevant in linguistic causatives, and Basque is no exception, as the following examples show:

(5a) **Oswaldek tiroz hilarazi zuen Kennedy*
Oswald.ERG gunshot.INS die.CAU AUX.PST:3SG.3SG Kennedy.ABS
"Oswald caused Kennedy to die by gunshot", i.e. "Oswald had Kennedy shot."

(5a′) *Oswaldek tiroz hil zuen Kennedy*
Oswald.ERG gunshot.INS die AUX.PST:3SG.3SG Kennedy.ABS
"Oswald killed Kennedy by gunshot", i.e. "Oswald shot Kennedy."

(5b) **Francok tiroz hil zuen Grimau*
Franco.ERG gunshot.INS die AUX.PST:3SG.3SG Grimau.ABS
"Franco killed Grimau by gunshot", i.e. "Franco shot Grimau."

(5b′) *Francok tiroz hilarazi zuen Grimau*
Franco.ERG gunshot.INS die.CAU AUX.PST:3SG.3SG Grimau.ABS
"Franco caused Grimau to die by gunshot", i.e. "Franco had Grimau shot."

(5c) **Erregeak gosez hil zuen presoa*
king.ERG hunger.INS die AUX.PST:3SG.3SG prisoner.ABS
"The king killed the prisoner by hunger."

[4] With a different interpretation, the dative argument being benefactive (4b, c), the starred examples are well formed.

(5c′) *Erregeak gosez hilarazi zuen presoa*
king.ERG hunger.INS die.CAU AUX.PST:3SG.3SG prisoner.ABS
"The king caused the prisoner to die of hunger", i.e. "The king let the prisoner starve to death."

— In (5a) the use of the morphological causative is inappropriate because Oswald shot Kennedy himself. Use of the morphological causative suggests that Oswald was the indirect causer, rather than the agent of "shoot".
— In (5b) it is the lexical causative that is inappropriate, because its use suggests that Franco himself shot Grimau, rather than condemning him to death by firing squad.
— In (5c) the lexical causative is wrong again, because when someone starves to death, the immediate cause of death is hunger or starvation, a process which, at least from the language's point of view, an agent cannot control directly or use as a weapon. Since the causer's influence is indirect, use of the lexical causative is inappropriate. Interestingly, if *gosez* "by hunger" is replaced by *ezpataz* "by the sword" or *tiroz* "by gunshot", which are instruments that the causer can control directly, the sentence is well formed:

(5d) *Erregeak ezpataz hil zuen presoa.*
king.ERG sword.INS die AUX.PST:3SG.3SG prisoner.ABS
"The king killed the prisoner with a sword."

1.4. The aim of this paper is to examine the causative alternation behind Basque lexical causatives of the kind illustrated in (1-2a). First I shall review previous theoretical approaches to the subject and explain my preference for the lexical decomposition approach (§2). Following that I will take a look at the implications of this decision regarding the syntactic features of lexical causatives (§3).

I will then show that the restriction against forming lexical causatives from transitive verbs mentioned above, while true, is only part of the story, for there are further restrictions on the formation of lexical causatives from intransitive verbs. Then I will look at possible connections between such restrictions and a verb's associated case morphology, showing intransitive verbs of the [ERG] type[5] cannot supply lexical

[5] The only apparent exception is *jo* "hit, ring". See the following examples:

(i) *Ezkilek jo dute* "The bells rang"
bell.PL.ERG ring.PTP AUX:3PL
(ii) *Ezkilak jo ditugu* "We rang the bells"
bell.PL.ABS ring.PTP AUX:1PL.3PL

According to Levin & Rappaport Hovav (1995: 140) verbs of emission are inergative verbs. In (i), *jo* is used in such a way and the subject takes ergative case, just like *dirdiratu* "shine, glitter". (ii) shows that the same verb (*jo*) can be used as a causative verb. However, it is not clear that examples in (i-ii) are a case of causative alternation. *Jo* is a polysemic verb ("hit, beat, play (music)"...) often used as transitive verb. Even used as a verb of emission, *jo* can be interpreted as a transitive verb with an unspecified object. Compare (i) with (iii) below:

(iii) *Ezkilek meza jo dute* "The bells rang for mass"
bell.PL.ERG mass.ABS ring.PTP AUX:3PL

causatives. Furthermore, there are some kinds of [-ERG] intransitive verbs which cannot provide lexical causatives either, including all [ABS, DAT] type verbs and also several [ABS] type verbs. I shall conclude that Basque lexical causatives can only be formed from monadic verbs of change, including psych-causatives with an experiencer as object. To explain this, I shall argue that the causative head of lexical causatives selects one of the predicates BECOME or GO, in contrast to morphological causatives with which another syntactic argument *(Voice)* is selected. Finally (§6), following Pylkkänen's (2002) typology which differentiates between a *Cause* head and a *Voice* having an external argument, I will conclude that both Basque and English are languages which conflate both heads.

2. Lexical and syntactic explanation of causative sentences

Like other syntactic alternations associated with the number of arguments of a verb or changes in the way arguments are expressed, such as noun incorporation, passivization, applicatives etc., lexical causatives involve issues concerning the relation between syntax and the lexicon. Approaches to these issues fall into two groups, associated with the lexicalist hypothesis and the syntactic hypothesis respectively.

2.1. In the lexicalist view, the causative alternation is based in the lexicon, in accordance with the Projection Principle (Chomsky 1981). Each lexical entry has an argument structure associated with a verb's meaning and reflected in its syntax. In the case of lexical causatives, both uses of a given verb appear at the level of the lexicon since there are two different argument structures that somehow correspond to them, even though the difference is systematic and limited basically to the presence or absence of an external argument. Within this view, causative alternations have been represented in two ways: either as the addition of an argument, or as the substraction of an argument.

In studies which favour the addition of arguments (Williams 1981), in the argument structure of verbs with causative alternation, a causer argument is added to a monadic verb turning it into a diadic verb, as in the case of labile causatives with a double dictionary entry like *kill* versus *die*. This is the approach taken in EGLU-II (52):

(6) *hil*$_1$ "die" _ *hil*$_2$ "kill"
[NOR] [NOR-NORK]
+inchoative +causative

Works favouring the subtraction of arguments (Levin & Rappaport Hovav 1995, Jackendoff 1997) prefer the opposite analysis, and rather than studying the causative alternation as causativization, they approach it as decausitivization. The verb has a theta-role corresponding to an external argument in its semantic representation, but this does not appear in the argument structure and is therefore not reflected syntactically. In this approach, then, a causative structure is found in the basis of the lexical representation of unaccusative verbs, and this is reflected in many languages where, if one of the forms is marked in verbs with a causative alternation, it is the intransitive form. According to Levin & Rappaport Hovav (1995: 80-81), Chierchia (1989) demonstrates this for

Romance languages with regard to the inchoative-causative alternation: unaccusative verbs take a reflexive form, and causatives the corresponding non-reflexive form. Levin & Rappaport Hovav themselves accept this view and incorporate it into their theory of unaccusativity, in which verbs with lexical alternation have a single representation in Lexical Conceptual Structure but two argument structures, one of which is diadic (the causative) and the other monadic. They appeal to lexical binding, which deletes an external argument, to explain why the external argument of the Lexical Conceptual Structure corresponding to the agent of the causative event fails to be reflected in the argument structure of unaccusative verbs. This is how the two alternating lexical representations of the verb *hil* "die, kill" appear in this theory (cf. Levin & Rappaport Hovav 1995: 108):

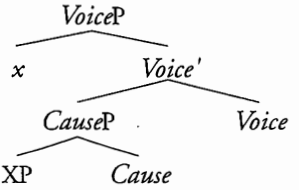
Unaccusative *hil* "die": *hil*: <y>		
	[*x* DO-SOMETHING] CAUSE [*y* BECOME *hil*]	
	Ø	
Lexical binding:	Ø	
Linking rules:		Ø
Argument structure:		<*y*>
Causative *hil* "kill": *hil*: x <y>		
	[*x* DO-SOMETHING] CAUSE [*y* BECOME *hil*]	
Linking rules:	Ø	Ø
Argument structure:	*x*	<*y*>

Here the causative alternation appears in the lexicon, in the organisation of argument structure, and is then reflected in the syntax, according to the lexicalist hypothesis.

2.2. In an alternative approach, following earlier treatments within generative semantics (Lakoff 1968), analysis of the alternation is located in the syntax. Given that, as shown particularly by Baker (1988) and, with reference to causative morphology in Basque, Deustuko Mintegia (1989), a syntactic analysis of syntactic regularities is possible, it was taken for granted that such an analysis would be plausible for causative alternations also. In Minimalism, the occurrence of an external argument is linked to a special syntactic head (cf. Chomsky's (1995: 352) light verb *v*, and Kratzer's (1996) *Voice),* and this line of analysis has recently been pursued in various forms (Megerdoomian 2002, Pylkkänen 2001 (2002): Folli & Harley to appear). I will follow the same approach here, assuming that syntactic regularities, including those which appear in lexical causative alternations, are to be explained syntactically. In this approach, lexical decomposition is carried out directly according to syntactic principles in line with Hale & Keyser's (1993) proposal, but without a separation of syntax and the lexicon. Predicates that arise through decomposition are made to appear in the syntax, each with its unique argument (Baker 1997, McGinnis 2000).

Since causative verbs have a single head in the present proposal, such verbs will take the following syntactic form (where the head is simply referred to as *Cause*, distinct from *Voice* and under it):

(7)



In (7) the complement of the *Cause* head is not specified and is hence valid for different causative types, i.e. both lexical and morphological causatives.

As we have seen, lexical and morphological causatives in Basque have different distributions, and many verbs that can occur with the causative morpheme *arazi* do not have lexical causative alternation, so XP must differ in such cases, but how? That is the issue we are going to study now, with special attention to lexical alternation since that is where we find the greatest number of restrictions. We shall discover, in line with Pylkkänen (2002), that in the case of lexical causatives there is a special relationship between *Cause* and *Voice*.

3. The complement of the Cause head in lexical causatives.

In §1 we saw that with the exception of stative predicates (§1.1), Basque verbs may undergo morphological causativization. In order to delimit the more restricted set of verbs capable of lexical causative alternation in Basque, we may begin with a descriptive characterisation referring to the morphosyntactic properties of such verbs.

Verbs whose subject takes the ergative case do not admit lexical causativization. This applies of course to transitive verbs (§1.2), including the numerous constructions on the pattern "do/make X" (i.e. noun + *egin* "do, make"), such as *eztul egin* "cough", literally "make (a) cough". The restriction also applies to deponent verbs such as *bazkaldu* "have lunch", *dirdiratu* "sparkle, glitter", *etsi* "surrender", *iraun* "last", which have a single nuclear argument that takes the ergative case.

The first of the following examples (8a) illustrates the restriction on an ordinary transitive verb, *jan* "eat".[6]

(8a) **Pellok* *Maddiri* *ogia* *jan dio*
Peter.ERG Mary.DAT bread.ABS eat AUX:3SG.3SG.3SG

(8a') *Pellok* *Maddiri* *ogia* *janarazi dio*
Peter.ERG Mary.DAT bread.ABS eat.CAU AUX:3SG.3SG.3SG
"Peter made Mary eat the bread."

The restriction applies regardless of whether the object is specified (*ogia* "bread") as in (8a,a') or unspecified as in one interpretation of (8b,b').

(8b) **Pellok* *Maddiri* *jan dio*
Peter.ERG Mary.DAT eat AUX:3SG.3SG.3SG

[6] In all the examples in (8), a dative DP corresponds to the *causee*.

(8b') *Pellok Maddiri janarazi dio*
Peter.ERG Mary.DAT eat.CAU AUX:3SG.3SG.3SG
"Peter made Mary eat (it)."

(8c) illustrates the restriction on the light verb *egin* "do, make" in the construction *eztul egin* "cough":

(8c) **Pellok Maddiri eztul egin dio*
Peter.ERG Mary.DAT cough make AUX:3SG.3SG.3SG

(8c') *Pellok Maddiri eztul eginarazi dio*
Peter.ERG Mary.DAT cough make.CAU AUX:3SG.3SG.3SG
"Peter made Mary cough."

(8d) illustrates a similar restriction on intransitive verbs with an ergative subject, which in this case is animate *(semea* "son"), cf. *Semeak bazkaldu du* "The son [ERG] had lunch":

(8d) **Pellok semeari bazkaldu dio*
Peter.ERG son.DAT have.lunch AUX:3SG.-.3SG

(8d') *Pellok semeari bazkalarazi dio*
Peter.ERG son.DAT have.lunch.CAU AUX:3SG.-.3SG
"Peter made (his) son have lunch."

(8e,f) show that the same applies when the base verb's ergative subject is an inanimate (*gerla* "war"), cf. *Gerlak iraun zuen* "The war [ERG] lasted (a long time)". Notice that in this case, the lexical causative construction is barred, regardless of whether the case of the causee is absolutive (8e) or dative (8f).

(8e) **Erregeak gerla iraun zuen*
king.ERG war.ABS last AUX.PST:3SG.3SG

(8e') *Erregeak gerla iraunarazi zuen*
king.ERG war.ABS last.CAU AUX.PST:3SG.3SG
"The king made the war last (a long time)."

(8f) **Erregeak gerlari iraun zion*
king.ERG war.DAT last AUX.PST:3SG.-.3SG

(8f') *Erregeak gerlari iraunarazi zion*
king.ERG war.DAT last.CAU AUX.PST:3SG.-.3SG
"The king made the war last (a long time)."

In these examples the morphological causative (8a′,b′,c′,d′,e′,f′) is possible but the lexical causative (8a,b,c,d,e) is not. Assuming that the occurrence of the ergative case is a realization of the light verb *Voice* on an external argument, it can be inferred by generalisation from these examples that the XP complement of the causative head of the lexical causative cannot be *Voice*P.

Next we would like to find out whether this initial generalisation about verbs with ergative arguments can be extended further. Considering that deponent verbs are associated with a transitive structure in the lexicon (Hale & Keyser 1993, Laka 1993), and all verbs with ergative arguments are at least diadic, let us see if the generalisation can be extended to all polyadic verbs. In that case, the generalisation might have the important syntactic implication that XP in the representation of (7) may only be VP.

To test the validity of this generalisation in Basque descriptively, we must now look at [ABS, DAT] verbs, which have an absolutive and a dative argument, paying special attention to psych-verbs, which although few in number are significant for our study.[7] We need to distinguish between two types of [ABS, DAT] psych-verbs: those in which the experiencer is in the dative and those in which the experiencer is in the absolutive. The former type includes *ahaztu* "forget" and *gustatu* "like", and the latter type includes *urrikaldu* and *errukitu*, which both mean "to pity".

Adhering to the typology usually applied to these verbs (Belletti & Rizzi 1988), *ahaztu* and *gustatu* belong to the *piacere* type of psych-verb (Artiagoitia 1995, 2000). Such verbs do not admit a lexical causative alternation, as the following examples show:

(9a) *Adinarekin kantuak ahaztu zaizkit*
age.COM song.PL.ABS forget AUX:3PL.1SG
"On account of age I have forgotten the songs."

(9b) **Adinak kantuak ahaztu dizkit*
age.ERG song.PL.ABS forget AUX:3SG.3PL.1SG
*"Age has forgotten me the songs."

(9b′) *Adinak kantuak ahatzarazi dizkit*
age.ERG song.PL.ABS forget.CAU AUX:3SG.3PL.1SG
"Age has made me forget the songs."

[ABS, DAT] verbs like *urrikaldu* (in present-day usage)[8] and *errukitu*, which have the experiencer in the abolutive case, do not admit lexical causative alternation either:

---

[7] Communication verbs like *mintzatu* "speak" or *elekatu* 'converse' and some other verbs like *jarraiki* "follow", *ekin* "start doing something" can be used as [ABS, DAT] verbs. They have no causative alternation. However, this is not very revealing because causative alternation is blocked even when they are mere [ABS] verbs; see below §4.2.

[8] Following the data given by the DGV, untill the middle of the 19th century, *urrikaldu* "pity" was an [ABS, DAT] verb in which the experiencer was dative. This use is no longer available in present day Basque (outside except in markedly literary usage). Consider the following contrast:

(i) *Jainkoari urrikaldu zitzaizkion gizonak*
god.DAT pity.PTP AUX.3PL.3SG men.ABS
"God took pity on the men" (old usage)

(ii) *Jainkoa urrikaldu zitzaien gizonei*
god.ABS pity.PTP AUX.3SG.3PL men.DAT
"God took pity on the men" (contemporary usage)

(9c) *Aberatsak bakan urrikaltzen zaizkie behartsuei*
rich.PL.ABS rare pity.IMP AUX:3PL.3PL needy.PL.DAT
"The rich rarely take pity on the poor."

(9d) **Apaizaren erranek aberatsak urrikaldu dizkiete behartsuei*
priest.GEN saying.PL.ERG rich.PL.ABS pity 3PL.3PL.3SG AUX: needy.PL.DAT
"The priest's words made the rich take pity of the poor."

It can be concluded from the data cited so far that in lexical causatives the base verb must be a [ABS] monadic verb, i.e. a verb with a single argument which cannot be ergative. We must now ask whether any such [ABS] verb other than change-of-state verbs can appear as a complement of *Cause*. In fact, it cannot. There are some kinds of [ABS] verb that allow morphological causativization but not lexical causativization, as shown in the following table:

(10) Possibility of lexical causative alternation in major classes of [ABS] monadic verbs:

—Reflexive verbs	
garbitu "get washed", *jantzi* "get dressed", *orraztatu* "comb one's hair"...	NO
—Verbs of activity	
jokatu "play", *jostatu* "play", *mintzatu* "talk"...	NO
—Verbs of happening	
gertatu "happen", *jazo* "happen"...	NO
—Change-of-state verbs	
hil "die", *hautsi* "break", *zabaldu* "spread"...	YES
—Change-of-place verbs:	
atera "leave", *hurbildu* "come close", *joan* "go"...	YES
—Psych-verbs:	
aspertu "get bored", *harritu* "be surprised", *izutu* "be scared"...	YES

(10) shows which classes of NOR-verbs allow lexical causative alternation and which do not.[9] In the next two section we look at some examples which show in greater detail which lexical causatives of NOR-verbs are grammatical.

---

[9] I won't discuss the case of aspect verbs like *hasi* "begin" or *bukatu* "finish". As can be seen in the examples below, these verbs have causative alternation (Pustejovsky 1995: 201):

(i) *Pilota partida hasi/ bukatu zen* "The pelota game started/finished"
pelota game begin finish AUX:3SG

(ii) *Pilotariek partida hasi/ bukatu zuten* "The pelota players began/finished the game"
pelota player.PL.ERG begin finish AUX:3SG

However these aspectual causatives deserve a special analysis. Semantically, the complement of aspectual verbs must be an event. Therefore, only DPs which permits the event reading (by means of coercion) can appear in the transitive construction. This is why, out of context (iii) below is normally understood as (iv), depending on whether Mary is known as a writer or not.

(iii) *Maddik liburua hasi zuen* "Mary began the book"
Mary.ERG book.ABS begin AUX:3SG.3SG

4. [ABS] verbs that do not admit lexical causatives.

Let us first examine the verb classes shown in (10) that do *not* have lexical causatives.

4.1. Reflexive [ABS] verbs

There are two ways to make transitive verbs reflexive in Basque. One is to employ a reflexive pronoun, without altering the verb's transitive structure. The other is to alter the verb's syntax, turning it into a single-argument verb whose argument goes into the absolutive case (Ortiz de Urbina 1989). For most verbs the standard derivation is the one which maintains the verb's transitive form, but some verbs such as *beztitu* "get dressed" or *orraztatu* "comb one's hair" have lexicalized the intransitive reflexive. Consider these examples:

(11a) *Pello beztitu da*
Peter.ABS dress AUX:3SG
"Peter got dressed"

(11b) *Maddik Pello beztitu du*
Mary.ERG Peter.ABS dress AUX:3SG.3SG
"Mary dressed Peter."

(11c) *Pello orraztatu da*
Peter.ABS comb AUX:3SG.3SG
"Peter combed his hair."

(11d) *Maddik Pello orraztatu du*
Mary.ERG Peter.ABS comb AUX:3SG.3SG
"Mary combed Peter's hair."

The verbs in (11) have both intransitive and transitive usages. Nevertheless, they do not display lexical causative alternation because (11b) and (11d) do not incorporate the meanings of (11a) and (11c). In (11a,c) the verb is reflexive, meaning that Peter dresses himself and combs his own hair; whereas in (11b,d) Pello does not dress himself or comb his own hair. The only way to obtain a causative from these reflexive verbs is by means of a morphological causative: *Maddik Pello beztiarazi du* "Mary made Peter get dressed", *Maddik Pello orraztarazi du* "Mary made Peter comb his hair". These causatives are derived from syntactic intransitives that decompose lexically into two co-referential arguments, one internal and one external (Reinhart & Siloni, to appear). Thus the restriction that applies to these verbs arises from their underlying transitivity.

(iv) *Maddi liburua irakurtzen/idazten hasi zen* "Mary began readind/writing the book"
Mary.ERG book.ABS reading/writing begin AUX:3SG

(v), below, shows that the causative alternation is restricted to event-nouns:

(v) **Liburua hasi zen* (vs *Gerla/pilota partida/filma/klasea... hasi zen*)
"*The book began" (vs "The war/the pelota game/the movie/the class... began")

4.2. Unergative [ABS] verbs

In her classification of Basque verbs, Levin (1983, 1989) claims that monadic [ABS] verbs are all unaccusative with the single exception of *mintzatu* "speak". Although it would seem that the ergative case marking is becoming more and more widespread with non stative intransitive verbs, particularly in the dialects of the South (Sarasola 1977), this is an over-generalization for there are many unergative verbs, particularly in Northern dialects, which while semantically being clearly unergative, can or must be used as [ABS] verbs.[10] Some examples of these are *ari izan* "be doing (something)", *bazkaldu* "have lunch", *borrokatu* "fight", *dantzatu* "dance", *elekatu* "talk", *entseiatu* "try", *jauzi* "jump", *jazarri* "attack", *jokatu* "play (a competitive game)", *jostatu* "play, have fun", *mendekatu* "avenge", *mintzatu* "speak", etc. Regarding the analysis of these agentive verbs as unergative verbs, see Perlmutter & Postal 1984, and Levin & Rappaport Hovav 1995: 136.[11]

As the following examples show, unergative NOR-verbs do not alow lexical causatives (12b, 13b):

(12a) *Pello kanpoan jostatu da*
Peter.ABS outside play AUX:3SG
"Peter played outside."

(12b) **Maddik Pello kanpoan jostatu du*
Mary.ERG Peter.ABS outside play AUX:3SG.3SG

(12b′) *Maddik Pello kanpoan jostararazi du*
Mary.ERG Peter.ABS outside play.CAU AUX:3SG.3SG
"Mary made Peter play outside."

(13a) *Nire aurka borrokatu zara*
me.GEN against fight AUX:2SG
"You fought (against) me."

(13b) **Buruzagi berriek nire aurka borrokatu zaituzte*
boss new.PL.ERG me.GEN against fight AUX:3PL.2SG

(13b′) *Buruzagi berriek nire aurka borrokarazi zaituzte*
boss new.PL.ERG me.GEN against fight.CAU AUX:3PL.2SG
"The new bosses have made you fight (against) me."

Indeed, many speakers admit use of *borrakatu* "fight" as a transitive, e.g.

(13c) *Buruzagi berriek zu ere borrokatu zaituzte*
boss new.PL.ERG you too fight AUX:3PL.2SG
"The new bosses fought you too."

---

[10] Following de Rijk (2002) Levins' generalization describes the situation in old Basque (until the begining of the 16th century). In his proposal present-day unergative [ABS] verbs historically either derived from unaccusative verbs (for example, *trabaillatu* "work", when it was borrowed, wasn't unergative and meant "toil"), or they are the result of antipassive constructions (in the case of verbs like *mintzatu* "speak" or *mendekatu* "avenge"). I won't discuss this proposal here.

[11] However, change-of-place verbs, which may also be agentive (even when they don't express the manner of motion), are not included in this class as we shall see in §5.2.

However (13c) is not a causative formation and its meaning is not related to that of (13b′). The same applies to verbs denoting verbal communication such as *elekatu*, *hizkatu*, *mintzatu*, *solastatu*, etc., which all roughly mean "talk", "converse", "speak" in various dialects. In Northern dialects these verbs, while generally intransitive, admit transitive uses too:

(14a) *Pello* *mintzatu da*
Peter.ABS speak AUX:3SG
"Peter spoke."

(14b) *Maddik* *Pello* *eta Jakes* *mintzatu ditu*
Mary.ERG Peter.ABS and James.ABS speak AUX:3SG.3PL
"Mary spoke to Peter and James."

The verb in (14b) is transitive, but the meaning is not causative. We can prove this by inserting a prepositional phrase in (14a), as in (14c), and then trying to make the verb transitive as in (14d):

(14c) *Pello* *bere buruarekin* *mintzatu da*
Peter.ABS his head.COM speak AUX:3SG
"Peter spoke with his head", i.e. "P. spoke to himself."

(14d) **Maddik* *Pello* *bere buruarekin* *mintzatu du*
Maddi.ERG Peter.ABS his head.COM speak AUX:3SG.3SG
*"Mary spoke Peter with his head", i.e. *"M. spoke P. to himself."

(14d) is ungrammatical because in the only possible interpretation (corresponding to (14c)) it is a causative based on an unergative verb of communication.

It is unclear how the case morphology of such unergative verbs should be represented. In the lexical decomposition approach favoured here it is unlikely that we would want to assign different roles to arguments of the following verb pairs: *borroka egin* "fight" [+ERG]/*borrokatu* "fight" [-ERG], *ele egin* "talk" [+ERG] talk/*elekatu* "talk" [-ERG], *hitz egin* "talk"/*hizketatu* "talk" [-ERG], *zintz egin* "blow one's nose" [+ERG]/*zintzatu* "blow one's nose" [-ERG], etc. (Oyharçabal 1993). This kind of alternation is quite regular with some incorporating verbs, such as verbs of communication, where a noun such as *ele*, *hitz* or *solas* is combined with either the light verb *egin* "do, make", or a morphologically empty verb head, giving a simple verb (Hale & Keyser 1993). One possibility is, following Marantz (1991), to treat the Basque ergative as a dependent case and consider its occurrence in relation to the visibility of the object position (Oyharçabal 1994). Since the object is always visible in constructions with a light verb, use of the ergative is obligatory in this case. In incorporations with a head verb whose form is zero, on the other hand, the object position is released and becomes invisible, blocking occurrence of the dependent case, i.e. the ergative, since this needs to be able to 'see' the object in order to occur. In this analysis, full incorporation of N predicts Basque unergative verbs to be [ABS] and deponent verbs[12] to be the exception because in them the incorporated object remains visible.

[12] Since Lafitte (1944), Basque grammars call monadic simple verbs whose unique argument takes ergative case *deponent verbs*.

In any case it is highly significant for the analysis of these [ABS] monadic verbs that they can never appear in lexical causatives, even though, as we have seen in (11e) and (12b), some of them allow transitive formations, favouring the view that their argument is external.

4.3. Verbs of happening

[ABS] verbs of happening do not have lexical causatives either.

(15a) *Nire otoitzen ondotik, bi mirakuilu gertatu dira*
my prayer.PL.GEN in.consequence, two miracle.ABS happen AUX:3PL
"As a result of my prayers, two miracles happened."

(15b) **Nire otoitzek bi mirakuilu gertatu dituzte*
my prayer.PL.ERG two miracle.ABS happen AUX:3PL.3PL
*"My prayers happened two miracles."

Here the single argument cannot be treated as external: verbs of happening are unaccusative. So that cannot be the reason for the ungrammaticality of (15b). Basque is not alone here; English (Levin 1993: 21) and French (16a,b below) behave similarly:

(16a) *Il est survenu/advenu un miracle*
it AUX happen one miracle
"A miracle happened."

(16b) **Mes prières ont advenu/survenu un miracle*
my.PL prayer.PL AUX happen one miracle
*"My prayers happened a miracle."

Survenir and *advenir*, French verbs of happening, occur in a construction in which only unaccusative verbs are possible, with a non-specific subject following the verb. (16b) shows that such unaccusative verbs cannot occur in a lexical causative. In the lexical decomposition proposed below, we will associate lexical causation with predicates of change. However, in verbs of happening there is an existence predicate rather than a predicate expressing a change. This difference allows us to explain the absence of lexical causatives with such verbs. We have seen that some classes of [ABS] monadic verbs do not alternate with lexical causatives. Next we will look at some which do.

5. NOR-verbs that admit lexical causatives

Two types of verbs of change can be distinguished: change-of-state verbs (§5.1) and change-of-place verbs (§5.2). I will treat psych-causatives separately, although I ultimately propose that these can be thought of as verbs of change (§5.3).

5.1. Change-of-state verbs

Change-of-state verbs include verbs that express a change in the form or physical state of the immediate internal argument. Typically they are derived from an adjective, and sometimes from a noun. Some examples follow of Basque change-of-state verbs. Note that some of their meanings are intransitive and some transitive; *-tu* (or *-du)* is an

aspectual suffix (perfective),[13] while *-i* is an older equivalent that is no longer productive. So, deadjectival and denominal verbs in the list below are zero-derived.

arraildu "crack, get drunk", cf. *arrail* n. "crack" & adj. "cracked, drunk"
belztu "blacken, turn black", cf. *beltz* adj. "black"
berotu "heat, get hot", cf. *bero* adj. "hot" & n. "heat"
edertu "make/become beautiful, adorn", cf. *eder* adj. "beautiful"
eritu "become/fall/make ill", cf. *eri* adj. "ill" & n. "illness, ill person"
hautsi "break", cf. *hauts* n. "powder, ash"
hil "die, kill", cf. *hil* adj. "dead" (but this is also the participle of the verb *hil*)
hoztu "get/grow/make cold", cf. *hotz* adj. & n. "cold"
idortu "dry", cf. *idor* adj. "dry"
puskatu "break, break up", cf. *puska* n. "piece, bit"
urtu "melt", cf. *ur* n. "water"
zabaldu "spread, open", cf. *zabal* adj. "broad, wide"

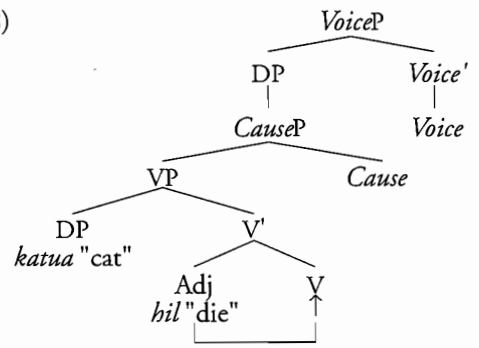
In this group of verbs the derivation of lexical causatives is totally productive, e.g.

(17a) *Udaberriarekin bazterrak laster berdatu ziren*
spring.COM corner.PL.ABS fast turn.green AUX:3PL
"With (the coming of) spring the countryside soon grew green."

(17b) *Udaberriak bazterrak laster berdatu zituen*
spring.ERG corner.PL.ABS fast turn.green AUX:3SG.3PL
"Spring soon turned the countryside green."

In the present analysis such causatives may be represented in two ways depending on whether or not the verb is derived from an adjective (or a noun).[14] If it is, the representation will be as in (18):

(18)



[13] In Basque participle forms are standardly used to cite verbs.

[14] Some change-of-state verbs can probably be analyzed as incorporating a null postpositional head. AOPZ (2000: 438) propose this kind of analysis for verbs like *apurtu, puskatu, zatitu...* "break, smash to pieces, divide...". These derived verbs incorporate a noun (*apur, puska, zati...*) which designates a small piece (of something). In some cases the postposition (*-ka*) may appear: *xehakatu, zatikatu...* Verbs like *lilitu, loratu* "blossom" can also be analyzed following this type of decomposition. See §5.2 and 5.3 below.

As a diagnostic for determining whether intransitive verbs are unaccusative or unergative, Levin & Rappaport Hovav (1995: 91) consider that most change-of-state verbs are unaccusatives because they are externally caused, whereas verbs that express internally-caused changes are unergatives. Nevertheless, there are some change-of-state verbs that have an internal cause (Levin & Rappaport Hovav 1995: 159). Such internally-caused verbs are unlikely to undergo lexical causativization. If a change-of-state is triggered by an internal cause, a subject of a lexical causative expressing an external cause cannot be a direct cause, cf. (5a,b), and such verbs therefore cannot undergo causative alternation.

As Levin & Rappaport Hovav (1995: 99) point out, not all languages deal with this problem in the same way, and some vacillations and contradictions are found even among speakers of a given language. Insofar as the present study is primarily descriptive, let us see what happens in the case of Basque.

As a general rule, Basque appears to tolerate causative alternation in change-of-state verbs that are conceptually analysable as internally-caused. Indeed there are some cases of Basque verbs that admit causative alternation even though in neighbouring languages the formation of a lexical causative from the equivalent verb is blocked on account of internal causation.

Studying Spanish change-of-state verbs, Mendikoetxea (1999: 1599) states that many internally-caused change-of-state verbs may be treated as if they were externally-caused, depending on the type of argument. Thus, if a verb can be used to talk about animals or natural phenomena, it is likely to admit an internal-cause reading that is not available with an inanimate subject, as in the following examples, cf. Mendikoetxea (1999: 1599):

SPANISH:

(19a) *Juan ha ensanchado* (internal cause)
Juan AUX widen.PTP
"Juan has broadened out."

(19b) *La carretera se ensancha en el km 5* (external cause)
ART road RFL widen in ART km 5
"The road widens at kilometre 5."

In Basque it seems that for processes involving inert or inanimate objects (but probably not plants), lexical causatives are possible for all speakers. Processes of change such as melting, rotting and rusting apply to inanimates yet may be thought of as internally caused. In Basque they are treated as externally caused and undergo causative alternation.

Take the French verb *fondre* "melt", for example, which is not amenable to causative alternation, whereas its Basque equivalent *urtu* is:

FRENCH:

(20a) **Le soleil a fondu le verglas*
ART sun AUX melt.PTP ART ice
"The sun melted the ice."

(20a′) *Le soleil a fait fondre le verglas*
ART sun AUX make/CAU.PTP melt ART ice
"The sun caused the ice to melt."

BASQUE:

(20b) *Eguzkiak bideko horma urtu du*
sun.ERG road.GEN ice.ABS melt AUX:3SG.3SG
"The sun melted the ice on the road."

When the undergoer of the change is animate, speakers' judgments differ and are often uncertain. This is illustrated in French and Basque for the verbs Fr. *grossir*, Bq. *loditu* "fatten, grow fat" and Fr. *maigrir*, Bq. *mehatu* "slim, grow thin":[15]

FRENCH:

(21a) *Pierre a grossi/ maigri*
Peter AUX fatten.PTP/ slim.PTP
"Peter grew fat/thin."

(21b) **Les médicaments ont grossi/ maigri Pierre*
ART.PL medicine.PL AUX fatten.PTP/ slim.PTP Peter
"The medicines fattened/slimmed Peter."

BASQUE:

(22a) *Pello loditu/ mehatu da*
Peter.ABS fatten/ slim AUX:3SG
"Peter grew fat/thin."

(22b) %*Erremedioek Pello loditu/ mehatu dute*
medicine.PL.ERG Peter.ABS fatten/ slim AUX:3PL.3SG
"The medicines fattened/slimmed Peter."

In this example with verbs expressing physical changes-of-state in the theme, some Basque speakers, but not all,[16] accept a causative alternation that is hardly acceptable in

[15] Apparently, Spanish data depend upon the speakers. Following Mendikoetxea (1999: 1598) a verb like *adelgazar* "slim" has no causative alternation:

(i) *Pedro adelgazó* "Peter slimmed"

(ii) **Un nuevo medicamento adelgazó a Pedro* "A new medicine slimmed Peter"

However, examples like (iii) appear in dictionaries:

(iii) *Esta medicina te adelgazará* "This medicine will slim you"

[16] In the DGV such examples appear from different dialects:

(i) *Janhari irintsuek loditzen dute* (Harriet) "Floury food slims"

(ii) *Etxeko jatekiak lorittu eiñ nau* (T. Etxebarria) "Homemade food fattened me"

French. The same pattern is observed with the Basque verb *gorritu* "blush" (literally "turn red", from *gorri* "red"). Here we compare this verb with its equivalents in several other languages, as mentioned in the literature:

ENGLISH:

(23a) *Peter blushed*

(23b) **The compliment blushed Peter* (Levin & Rappaport Hovav 1995: 91, 160)

SPANISH:

(24a) *María enrojeció*
María blushed
"María blushed."

(24b) **La enhorabuena enrojeció a María*
ART congratulation blushed ACC María
*"The congratulation blushed María." (Mendikoetxea 1999: 1604)

FRENCH:

(25a) *Marie rougit*
Marie blushed
"Marie blushed."

(25b) **Vos paroles rougirent Marie*
your.PL word.PL blushed Marie
"Your words blushed Marie." (Labelle 1990: 306)

BASQUE:

(26a) *Maddi gorritu zen*
Maddi.ABS blush AUX.PST:3SG
"Maddi blushed."

(26b) %*Zuk esandakoak gorritu egin nau*
you.ERG say.PTP.ERG blush FOC AUX:3SG.1SG
"What you said made me blush."[17]

Thus it would seem that there are very few unaccusative change-of-state verbs in Basque for which speakers unanimously reject causativization. *Hazi* "grow" when applied to plants has no causative use for many speakers, though some westerners admit causation with agentive subjects:[18]

---

[17] *Egin* "make, do" in (26b) is merely a marker of information structure which places the verb *gorritu* in emphatic focus (FOC). The acceptance of the example is easier when the verb is focalized.

[18] With verbs like *loratu* or *lilitu* "bloom" (< *lore*, *lili* "flower"), speakers' judgements are divided with regard to causative alternation; some accept it, while others do not:

(i) %*Maiatzeko eguzkiak gereziondoak loratu ditu*
may.GEN sun.ERG cherry.tree.PL.ABS bloom AUX:3SG.3PL
*"The May sun bloomed the cherry trees."

(27) %*Baratzezainak gure landareak ondo hazi ditu*
gardener.ERG our plant.PL.ABS well grow AUX:3SG.3PL
"The gardener grew our plants well."

(28) **Ongarriak gure landareak ongi hazi ditu*
fertilizer.ERG our plant.PL.ABS well grow AUX:3SG.3PL
*"The fertilizer grew our plants well."

A further pattern exists involving some of the verbs in this class. Certain verbs derived from nouns denoting plant and animal parts enter into a special kind of causative alternation, e.g.

aletu "bear fruit; pick", cf. *ale* "fruit, berry, bean etc."
bihitu "turn to grain; thresh, remove grain", cf. *bihi* "grain, cereal"
kimatu "sprout; prune", cf. *kimu* "shoot, sprout"
lumatu "grow feathers; pluck", cf. *luma* "feather"

AOPZ (2000: 439) and Etxepare (2003) draw attention to these verbs which have an unaccusative use that expresses internal causation, but also a causative use denoting removal of the part expressed by the incorporated noun:

(29a) *Kardua kimatu da*
thistle.ABS sprout AUX:3SG
"The (edible) thistle has sprouted."

(29b) *Jendeek mahastia kimatu zuten*
people.PL.ERG vineyard.ABS prune AUX:3PL.3SG
"The people pruned the vineyard."

So far, lexical causatives in Basque seem to be quite regular, exhibiting few of the idiosyncrasies often associated with lexicalization, in contrast to the next group we shall look at.

5.2. Change-of-place verbs

The class of change-of-place verbs consists exclusively of verbs which express simple or (directed motion, but not manner of motion (Rosen 1984, Levin & Rappaport Hovav 1995).[19] We saw earlier that single-argument agentive verbs are generally unergative, but Levin & Rappaport (1995: 148) established that the verbs in this group form an exception to that generalisation.[20] Even though many of these verbs have an agentive argument, they are unaccusatives and admit causative alternation:

---

[19] Basque generally doesn't lexicalize verbs of manner of motion. Verbs like *run, swim* or *walk* are expressed by using an adverb (indicating the manner) with a change-of-place verb like *etorri* "come", *ibili* "move", *joan* "go", etc. *Igerika/lasterka/oinez etorri/ibili/joan naiz* "I came/moved/went swimming/running/walking".

[20] For Levin & Rappaport (1995) the group is restricted to *verbs of inherently directed motion.* In Basque the group includes simple motion verbs like, for example, *ibili* "move".

(30a) *Kanpora atera zara*
outside go.out AUX:2SG
"You went outside."

(30b) *Kanpora atera zaitut*
outside take.out AUX:I.2SG
"I took you outside."

(31a) *Haurrak oheratu dira*
child.PL.ABS go.to.bed AUX:3PL
"The children went to bed."

(31b) *Haurrak oheratu ditut*
child.PL.ABS put.to.bed AUX:1SG.3PL
"I put the children to bed"

We have seen that there are few change-of-state verbs for which causative alternation is completely impossible for all speakers, but such cases are more numerous among change-of-place verbs:

— With lexical causatives: *agertu* "appear, display", *amildu* "plunge, hurl", *atera* "go out, take out", *elkarretaratu* "come/bring together", *etxeratu* "go/take home", *goititu* "rise, raise", *hurbildu* "approach, bring (to a place)", *igan* "go/take up", *ilki* "come out, bring out", *jaitsi* "go/take down", *jalgi* "bring/take out", *joan* "go, %take", *sortu* "emerge, come into being, be born; bring about, create", *urrundu* "move away", etc.
— Without lexical causatives: *ailegatu, arribatu* "arrive", *erori* "fall", *etorri* "come", *ibili* 'move', *irten* "go out", *jaio* "be born", *jin* "come", *joan* "go", *partitu* "leave", etc.

This distinction is hard to explain. The cases of lack of alternation seem to be the marked ones: they are few in number and constitute a closed list. Verbs derived from adverbs and postpositional forms belong to the group allowing causative alternation; in particular, the alternation is always potentially available for those containing the allative postposition *-ra*, e.g.

> *goratu* "go up, rise; bring up, raise", cf. *gora* "up(wards)", *etxeratu* "go home; take home", cf. *etxe* "house", *etxera* "home(wards)", *lurreratu* "come to the earth, fall to the ground, land; bring down, cause to fall", cf. *lur* "ground, earth", *lurrera* "to the ground, to the earth", etc.

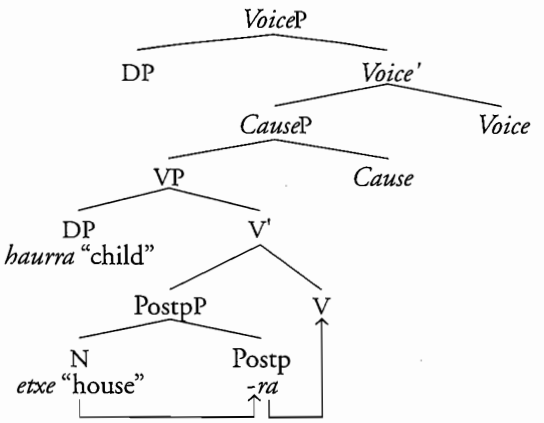
Certain verbs, such as *joan* "go", have a causative use in Northern dialects that is lacking in others (see DGV):

(32) *Ardiak mendira joan ziren*
sheep.PL.ABS to.the.mountain go AUX.PST:3PL
"The sheep went to the mountain."

(33) *%Artzainak ardiak mendira joan zituen*
shepherd.ERG sheep.PL.ABS to.the.mountain take AUX:3SG.3PL
"The shepherd took the sheep to the mountain."

Change-of-place verbs are represented as follows, once again with a derived verb for the sake of clarity:

(34)



The PostpP structure occurs in the case of verbs derived from allative expressions like *etxera* "to the house" or from adverbs like *urrun* "far" and *hurbil* "near". In other cases, such as *jautsi* "descend" or *igan* "rise", direction is an integral part of the verb's sense and the root is the verb's complement (Marantz 1997).

5.3. Causative psych-verbs

The class of causative psych-verbs consists of psych-verbs of the [ERG, ABS] type, i.e. having an ergative subject and an absolutive object, such as *aspertu, enoatu* "bore, be bored", *harritu* "surprise, be surprised", *interesatu* "interest, be interested", *izutu* "frighten, be frightened", *liluratu* "dazzle, fascinate, be dazzled, be fascinated", *poztu* "please, make happy, be pleased, be happy", etc. These are not universally treated as alternating verbs in the literature. While in some of these the subject constitutes the theme, in others the theme turns up as object. Consider the following:

(35a) *Peter fears bears* (35b) *Bears frighten Peter*

It has been suggested in the literature that *Peter* has the same theta-role in both (35a) and (35b), namely experiencer, as does *bears*, namely theme. Syntactically, of course, the theme is the object in (35a) and the subject in (35b), while the experiencer is the subject in (35a) and the object in (35b). This state of affairs is highly enigmatic if one accepts that theta-relations are reflected in syntactic relations (cf. Baker's *Uniformity Theta Assignment Hypothesis*, 1988: 46). To solve this puzzle, Belletti & Rizzi (1988) apply the unaccusative analysis to causative psych-verbs, suggesting that the theme argument occurring as subject is the verb's immediate internal argument in D-structure, and that the experiencer, located above the theme within the VP, takes an inherent accusative case. Since it is not an external argument, the theme argument rises to subject position as with unaccusatives. Thus the theme is a derived subject, as in (36):

(36) Theme$_1$ [VP [V' t_1] Experiencer]

Belletti & Rizzi (1988) provide strong syntactic arguments in favour of this explanation based on the hypothesis that subjects of the *preoccupare-frighten* class are derived (see also Artiagoitia 2003 in this volume), but others reject the thematic analysis on which this explanation is based (Dowty 1991, Pesetsky 1995, Tenny 1995, Baker 1997). Pesetsky (1995) observes that the theta role of *the article* is not quite the same in the following two examples:

(37a) *John is angry at the article* (37b) *The article angered John* AUX:3SG

Pesetsky (1995: 56) points out that in (37a) *the article* is the target of emotion, whereas in (37b) it is the causer of emotion. In (37a) the article is what John's anger is aimed at; in (37b), on the other hand, it is the cause of his anger, but not necessarily what his anger is aimed at. It may be that John thinks the article is well-written and that the article tells of something that makes him angry. According to Pesetsky's account, the thematic analysis of psych-verbs presented by Belletti & Rizzi (1988) is misleading, because in the transitive forms the subject is the causer, as has been suggested repeatedly (see also Dowty 1991, Baker 1997, Arad 1998, Pylkkänen 1999, and in reference to Basque, Zabala 1993: 203). I coincide with this view.

Let us examine the behaviour of causative psych-verbs in Basque, illustrated in (39):

(39a) *Jon enoatu /harritu da /izutu /kezkatu da*
John.ABS be.bored /be.surprised /be.frightened /be.worried AUX:3SG
"John was/got bored/surprised/frightened/worried."

(39b) *Pellok Maddi enoatu /harritu /izutu /kezkatu da*
John.ABS Mary.ABS be.bored /be.surprised /be.frightened /be.worried AUX:3SG
"John bored/surprised/frightened/worried Mary."

The main difficulty in accounting for the causative analysis of such pairs involves binding. It was observed by Artiagoitia (2000: 110) that unusual binding relations may be found with causative psych-verbs. Consider the following (cf. also Artiagoitia's (2000: 110) example with *nazkatu* "sicken"):

(40a) *Nire buruak izutzen nau*
my head.ERG frighten.IMP AUX:3SG.1SG
"I frighten myself", literally: "Myself frightens me."

(40b) *Pello bere buruak izutzen du*
Peter.ABS his head.ERG frighten.IMP AUX:3SG.3SG
"Peter frightens himself", literally "Himself frightens Peter" (or "Peter is frightened by himself.")

In these examples, the reflexive phrase *nire burua(k)* "my head.ERG, i.e. myself" or *bere burua(k)* "his head.ERG, i.e. himself' is the subject of the psych-causative, and is bound by the object. The data in (40) poses several problems. One involves Principle C, which says that referring expressions in a sentence must be unbound; the other involves

Principle A, which says that anaphoric expressions must be bound in their domain. For example, according to Belletti & Rizzi (1988) (41), which is the exact translation of (40b), is a violation of Principle C:[21]

(41) **Himself$_i$ worries John$_i$*

Just as in (41), in (40b) too the object, *Pello*, is a referring expression yet it is bound since it is c-commanded by the subject. If *bere burua(k)* "himself" and *Pello* are co-indexical in (40b), then Principle C is clearly violated. Therefore we should first of all find out if the two phrases in (40) are really co-indexical. When we examine these sentences more closely, some questions arise. For example, the pattern found in (40) is completely ungrammatical if we substitute a reciprocal anaphor as in (42):

(42) **Pello eta Maddi elkarrek izutzen ditu*
Peter.ABS and Mary.ABS each.other.ERG frighten.IMP AUX:3SG.3PL
*"Each other frighten Peter and Mary" (or "Peter and Mary are frightened by each other.")

The reciprocal pronoun *elkar* cannot be used in subject position, whereas *bere burua* "himself" can. How can we explain this difference, which doesn't appear in other contexts, as can be seen in (43a,b)?

(43a) *Pello eta Maddik elkar hilen dute*
Peter and Mary.ERG each.other.ABS hit.FUT AUX:3PL.3SG
"Peter and Mary will kill each other"

(43b) *Pello eta Maddik beren burua hilen dute*
Peter and Mary.ERG their head.ABS kill.FUT AUX:3PL.3SG
"Peter and Mary will kill themselves" (i.e. commite suicide)

In the examples of (43), the two anaphoric expressions obey Principle A, since both are bound in the relevant local domain. I will consider that *X-en burua* is a metonymic anaphor (cf. Safir 1996) and that in such a case the *i*-within-*i* condition is deactived as Rebuschi (1997: 288) proposes:

(44) ... *Maddik*$_1$ ... [DP$_2$ [*ber(e)*$_1$ [*buru*$_2$]]-*a*] ...

On the contrary, there is a sharp contrast between (40a,b) and (42). The expression *X-en burua* can appear in the subject position of the psych-causative (40), while *elkar*

[21] Under Belletti & Rizzi's (1988) analysis, Principle A is not violated in (41), or in (i) either:

(i) *Pictures of himself frighten John*

As seen in (36), in Belletti & Rizzi's (1988) view, the subject is derived from a position where it is c-commanded by *John*. Therefore the anaphor inside of it is bound.

cannot (42). To explain this, I propose that in (40) *bere buruak* is not an anaphor, but an ordinary DP, which is metaphorically used to denotes one 's (uncontrolled) self. On the other hand, the reciprocal pronoun *elkar* is morphologically simple[22] and has to bear the index of the binding DP.

According to our proposal, the expression *X-en burua* can be syntactically autonomous even when it is metonymically used to designate not really the body part, but the whole person, and we would expect that it may also occur outside of psych-causatives. Such is the case in the following examples from various periods and dialects in Basque literature, in which *nire burua*, as subject, denotes the first person (cf. DGV, sub *buru*, p. 2,672):[23]

(45a) ***Nere buruak*** *ere ematen dit* *franko lan*
my head.ERG too give.IMP AUX:3SG.3SG.1SG plenty work.ABS
(Labayen, *Euskal-Eguna*, 92)

"I give myself plenty of work too", lit. "Myself also gives me plenty of work", "I am given plenty of work by myself."

(45b) *Halaz despeditu nahi nuzuia?*
thus take.leave want AUX:2SG.1SG.INTER
Hebetik ioan gabe ***ene buruia***
from.here go without my head.ABS
egin behar duzu *ene nahia*
do must AUX:2SG.3SG my wish.ABS (Dechepare, 207)

"Would you take leave of me thus? Before I (*lit.* myself) depart hence you shall fulfil my wish."

We must adduce some further data, which enforces our proposal. The fact that *X-en burua* may appear as subject of psych-causatives as shown in (40) doesn't rule out that the same expression may also appear in object position. See the examples in (46):

(46a) *Nire buruak izutu nau*
my head.ERG frighten AUX:3SG.1SG
"I frightened myself", literally: "Myself frightened me."

(46b) *Nire burua izutu dut*
my head.ABS frighten AUX:1SG.3SG
"I frightened myself."

In (46a) the expression *nire burua* is the subject and it takes the ergative case. In (46b) the same expression is the direct object and it receives the absolutive case. The

---

[22] I don't take into account historical complexity (*elkar* < *alkhar* < **(h)ark-har* "DEM.ERG-DEM.ABS" (Michelena 1961: 69).

[23] Unlike (45a), example in (45b) (16th century) is rather strange for present-day speakers, because it is difficult not to give the sentence an agentive interpretation (lit. "before I go out from here"). Observe further that the genitive pronoun doesn't have the reflexive form (*neure*) of the genitive pronoun of the 1st person in Dechepare's dialect. Regarding the latter point, see Rebuschi (1995).

two sentences are not exactly synonymous. Speakers perceive a difference in the interpretation of (46a) and (46b), attaching an agentive interpretation to (46b)[24] (Artiagoitia 2000: 110), while (46a) is given a psychological interpretation in which the reason for being frightened resides in one's uncontrolled self. This difference in interpretation corresponds to different uses of the same expression: metonymic anaphor in (46b); metonymic R-expression in (46a).

Now let us look at the representation of psych-causatives. It was noted above that verbs denoting physical states often have an adjective base such as *argal* "thin", *bero* "hot", *handi* "big", *hil* "dead", *hotz* "cold", *lodi* "fat", *luze* "long", *mehe* "thin", *tiki* "small", *zabal* "wide", etc. Verbs formed from words that express psych-states denote changes-of-state too, but are mostly derived from nouns, such as *ahalke* "shame", *arrangura* "worry, preoccupation", *asper* "boredom", *beldur* "fear", *griña* "passion", *izu* "fright", *kezka* "concern", *poz* "pleasure, happiness", etc. Such nouns mostly occur in combination with the intransitive copular verb *izan* "be" as stative predicates, e.g. *ahalke izan* "be ashamed" (literally "be shame"), *arrangura izan* "be worried" (lit. "be worry"), *beldur izan* "be afraid" (lit. "be fear")...; however, they cannot be so used attributively (??*gizon ahalkea* "ashamed man", **gizon arrangura* "worried man", **gizon beldurra* "afraid man", and so on).[25] Such nouns can also occur in postpositional phrases, especially when they occur as a noun phrase rather than a plain noun (Zabala 1993: 544-48):

(48a) *Beldur /ahalke/arrangura /haserre /lotsa naiz*
fear shame/worry anger fear be:1SG
"I am afraid/ashamed/worried/angry/afraid."

(48b) *Kezkaz /beldurrez /pozik nago*
worry.INS fear.INS happiness.PAR be.3SG
"I am worried, afraid, happy."

(48c) *Haserre gorrian naiz*
anger red.INE be:1SG
"I am boiling with rage", lit. "I am in red anger"

I propose that in the decomposition of these kinds of psych-causative there is also a verb of change that selects a PP,[26] where the change consists of entry into a new psych-

---

[24] The agentive interpretation doesn't imply here that the subject is really in control of the process. For myself as for all the speakers I asked (46b) is preferred to describe the following situation: While you were driving on a wet road, you went into a skid and almost had an accident. Which sentence do you prefer to use to describe your feeling: (46a) or (46b)?

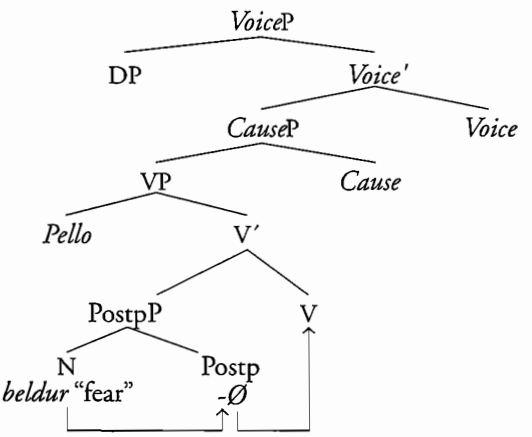
[25] There are ambivalent forms like *haserre* 'anger, angry', which can be used attributively: *gizon haserrea* "the angry man". There are also psych-verbs which can be derived from non ambiguous adjectives, e.g. *alegeratu* "become/make happy", *tristetu* "become/make sad"...

[26] Baker (1997) analyzes causative psych-verbs like *frighten* as a change-of-place whose theme is the emotion and the locative goal the experiencer. The semantic analysis is:

(i) x CAUSE [[FEAR (of x)] GO TO z]

state: x CAUSE [y BE LOC z PSYCH-STATE].[27] On this analysis, underlying a sentence such as *Maddik Pello beldurtu du* "Maddi frightened Pello" there is a PP with an unexpressed head, in which *beldur* "fear" is incorporated in a head-to-head movement which carries forward as far as the *Cause* node.

(49)



6. *Cause* and *Voice* heads

One issue not yet discussed is the type of relation that holds between the causative head and the head of the transitive *Voice* (or *v)* structure. Pylkkänen (2002), inquiring into the relationship between causative morphology and the existence of an external argument, suggests the the two do not always coincide, and notes that in Japanese adversative causatives and Finnish volitive causatives, the causative morphology may leave the verb's valency unchanged, yet the sentence is nevertheless causative. Consider the following Finnish examples:

FINNISH:

(50a) *Maija laula -a*
Maija.NOM sing -3SG
"Maija is singing"

(50b) *Maija -a laula -tta -a* (Pylkkänen 2002: ex. (168))
Maija -PAR sing -CAU -3SG
"Maija feels like singing."

[27] Artiagoitia (pc.) notices that the proposed analysis predicts that verbs like **ahalkez(ta)tu* "become/ make ashamed" or **lotsaz(ta)tu* "become/make afraid"... should be well formed, and this not so. He suggests an analysis where the noun is directly incorporated into V. However, in my view (49) doesn't imply that the lexical realization of the verb has to permit the use of an overt postposition. This is a different matter, which has to do with the way postpositions present in lexical decomposition are phonetically realized within verbs; see footnote 14 above for some other examples.

Adding the causative morpheme *-tta* does not result in the inclusion of another argument, yet the causative morphology conveys an implicit causative event which may be made explicit by a question (Pylkkänen 2002: ex. (174)):

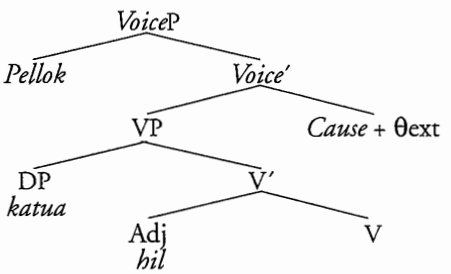
(51) *Minu -a naura -tta -a mutt-en tiedä mikä*
me -PAR laugh -CAU -3SG but-not.1SG know what.NOM
"Something makes me feel like laughing, but I don't know what."

The causative morphology appears in (51) with the subject in the partitive as in (50b). However, in the second part of the sentence the causative question word *mikä* appears. Significantly, the question word cannot express an agent:

(52) **Minu -a naura -tta -a mutt-en tiedä kuka*
me -PART laugh -CAU -3SG but-not.1SG know who.NOM
"Something makes me feel like laughing, but I don't know who."

Hence Pylkkänen (2002) concludes that causation does not always entail an external argument, for which reason the *Cause* head and the external-argument-bearing head (*Voice*) should be differentiated. However, in languages which express lexical causatives through zero morphology, lexical causatives cannot occur without an external argument. This is the case in both English and Basque, where both heads conflate. Pylkkänen calls such a situation Voice-bundling. (53) shows the representation of a lexical causative in this perspective:

(53) *Pellok katua hil du*
Peter.ERG cat.SG.ABS. killed AUX
"Peter killed the cat."



As (53) shows, the external-argument-bearing *Voice* head is associated with the *Cause* head, so causation and the existence of an external argument are linked, unlike Finnish and Japanese. Notice that this structure of causative verbs is similar to that of other transitive verbs. This is why, in some works such as AOPZ (1999: 442), where an intransitive alternation is lacking, some derived verbs not included among the causatives are analysed in causative terms, e.g. in the lexico-semantic structure of verbs such as *babestu* "protect", *zigortu* "punish", etc.

7. In conclusion: lexical causative alternation in Basque occurs with verbs which express a change in the form, location or psych-state of the subject. Aside from certain

idiosyncrasies associated with specific roots, this kind of lexical alternation is highly regular is Basque. We encounter three main types of decomposition, all characterised by a *Cause* head which selects a VP that denotes a change of state or place. In one type, illustrated by (18), the verb BECOME selects and incorporates an adjective or noun, with no further overt morphology, e.g. *edertu* "become beautiful", *handitu* "become big", *haurtu* "become a child" (cf. *eder* "beautiful", *handi* "big", *haur* "child"). In the second type, illustrated by (34), the predicate GO selects an allative PostpP or adverb which inherently expresses direction, e.g. *atera* "go out", *etxeratu* "go home", *hurbildu* "come close" (cf. *ate-ra* "to (the) door", *etxe-ra* "to (the) house", *hurbil* "near"). In the third type, illustrated by (49), the same underlying verb selects a PostpP whose underlying head incorporates the head of its complement. This formation is typical of psych-causatives, e.g. *ahalketu* "be ashamed", *beldurtu* "be afraid", *poztu* "be happy" (cf. *ahalke* "shame", *beldur* "fear", *poz* "happiness"). Like English, Basque conflates the head that expresses causation, which we have called *Cause*, and that which bears an external argument, here called *Voice*. This fact is presumably related to the causative head's zero morphology in causative alternations.

Bibliography

Alberdi, X., 2003, "Euskal aditz mailegatuen erregimena: hurbilketa", in J.M. Makatzaga and B. Oyharçabal (eds.), *Euskal gramatikari eta literaturari buruzko ikerketak XXI. mendearen atarian. Gramatika gaiak,* Iker-14 (I), Euskaltzaindia, Bilbao, 37-60.

Alsina, A., 1992, "On the Argument Structure of Causatives", *Linguistic Inquiry* 23.4, 517-57.

Arad, M., 1998, "Psych-notes", *UCL Working Papers in Linguistics,* 10.

Artiagoitia, X., 1995, "*Garri* atzizkiaren izaera bikoitzaz: zergatik den maitagarria bezain mingarria", *ASJU* 29.2, 355-405.

—, 2000, *Hatsarreak eta Parametroak lantzen,* EHU & Arabako Foru Aldundia.

—, 2003, "The case of an enlightening, provoking and admirable Basque derivational suffix with implications for the theory of argument structure", in this volume, 147-183.

Azkarate, M., Odriozola, J.C., Perez, E. & Zabala I. [AOPZ], 2000, "Análisis de los afijos del Vasco: el sufijo *–tu*". In L. Gràcia Solé, M.T. Cabré Castellví, S. Varela Ortega, M. Azkarate Villar & al. (eds.) *Configuración morfológica y estructura argumental: léxico y diccionario,* Publications of the University of the Basque Country: 436-59.

Baker, M., 1988, *Incorporation theory,* University of Chicago Press, Chicago.

—, 1996, "On the Structural Position of Themes and Goals". In J. Rooryck & L. Zaring (eds.), *Phrase Structure and the Lexicon,* Kluwer, Dordrecht: 7-34.

—, 1997, "Thematic roles and syntactic structure". In L. Haegeman (ed.), *Elements of grammar,* Kluwer, Dordrecht: 137-178.

Belletti, A. & Rizzi, L., 1988, "Psych-verbs and Theta-Theory", *Natural Language and Linguistic Theory* 18, 291-352.

Borer, H., 1994, "The projection of arguments", E. Benedicto and J. Runner (eds.), *University of Massachusetts Occasional Papers in Linguistics* 17.

—, 1993, "The inchoative-causative alternation: a case study in Parallel Morphology", *The Linguistic Review* 8.

Chierchia, G., 1989, "A semantics for unaccusatives and its consequences". Unpublished manuscript, Cornell University, Ithaca, NY.

Chomsky, N., 1981, *Lectures on Government and Binding,* Foris, Dordrecht.

—, 1995, *The Minimalist Program,* MIT Press, Cambridge, MA, & Londres.

Comrie, B., 1989, *Language Universals and Linguistic Typology*, Basil Blackwell, Oxford.
Danlos, L., 2001, "Event coreference in causal discourses". In P. Bouillon & F. Busa (eds.), *The Language of Word Meaning*, Cambridge University Press, Cambridge.
Deustuko Mintegia, 1989, "Inkorporazioa perpaus kausatiboetan". In P. Salaburu (ed.), *Sintaxi teoria eta euskara*, EHU, Donostia, 87-119.
Dixon, R.M.W., 2000, "A Typology of causatives: form, syntax and meaning". In R.M.W. Dixon & A.Y. Aikhenwald (eds.), *Changing valency. Case studies in transitivity*, Cambridge University Press, Cambridge, 30-83.
Dowty, D., 1979, *Word Meaning and Montague Grammar*, Reidel, Dordrecht.
—, 1991, "Thematic Roles and Argument Selection", *Language* 67.3, 547-619.
EGLU-II (*Euskal gramatikaren lehen urratsak*), 1987, Gramatika batzordea. Euskaltzaindia. Bilbo.
Etxepare, R., 2003, "Valency and argument structure in the Basque verb", J.I. Hualde & J. Ortiz de Urbina (eds.), *A grammar of Basque*, Mouton de Guyter, Berlin.
Fodor, J., 1970, "Three Reasons for not Deriving 'kill' from 'cause to die'", *Linguistic Inquiry* 1, 429-438.
Folli, R. & Harley, H., (to appear), "Consuming Results in Italian & English: Flavors of *v*", P. Kempchinsky & R. Slabakova (eds.), *Aspect Book Title*, 1-25, Kluwer, Holland.
Grimshaw, J., 1990, *Argument Structure*, MIT Press, Cambridge, MA.
Hale, K. & Jay Keyser, S.J., 1993, "On argument structure and the lexical expression of syntactic relations". In Hale and Keyser (eds.), 53-109.
— (eds.), 1993, *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, MIT Press, Cambridge, MA.
Halle, M., & Marantz, A., 1993, "Distributed Morphology and the Pieces of Inflection". In Hale and Keyser (eds.), 111-176.
Jackendoff, R., 1997, *The Architecture of the language faculty*, Cambridge, MIT Press, MA.
Kratzer, A., 1996, "Severing the External Argument from its Verb". In J. Rooryck & L. Zaring (eds.), *Phrase Structure and the Lexicon*, 109-138, Kluwer, Dordrecht.
Labelle, M., 1990, "Unaccusatives and Pseudo-Unaccusatives in French". *Proceedings of the North East Linguistic Society* 20.2, 303-317, Pittsburgh.
Lafitte, P., 1944, *Grammaire basque. Dialecte navarro-labourdin*, revized in 1962, fac-simile, Elkarlanean, Bayonne.
Laka, I., 1993, "Unergatives that assign Ergative, Unaccusatives that assign Accusative". In J. Bobaljik & C. Phillips (eds.), *Papers on Cas Agreement I, MIT Working Papers in Linguistics* 18, 149-172, Cambridge, MA.
Lakoff, G., 1968, "Some verbs of change and causation", S. Kuno (ed.), *Mathematical Linguistics and Automatic Translation*, Aiken Computation Laboratory, Harvard University, Cambridge, MA.
Levin, B., 1983, *On the Nature of Ergativity*, Ph-D, MIT, Cambridge, MA.
—, 1989, "The Basque Verbal Inventory and Configurationality", L. Maracz & P. Muysken (eds.), *Configurationality*, 39-62, Foris, Dordrecht.
—, 1993. *English Verb Classes and Alternations. A Prelimenary Investigation*, The University of Chicago Press, Chicago and London.
— & Rappaport Hovav, M., 1995, *Unaccusativity. At the Syntax-Lexical Semantics Interface*, MIT-Press, Cambridge, MA.
Marantz, A., 1997, "No escape from Syntax: Don't try morphological analysis in the privacy of your own lexicon", in A. Dimitriadis & L. Siegel (eds.), *University of Pennsylvania Working Papers in Linguistics* 4.2, 221-225, University of Pennsylvania, Philadelphia.
McGinnis, M., 2000, "Event heads and the Distribution of Psych-roots", A. Williams & E. Kaiser (eds.), *Current work in linguistics: University of Pennsylvania Working Papers in Linguistics* 6.3, 107-144.

Megerdoomian, K., 2002, *Beyond Words and Phrases: A Unified Theory of Predicate Composition*, Ph-D, UCLA.

Mendikoetxea, A., 1999, "Construcciones inacusativas y pasivas". In I. Bosque & V. Demonte (dir.), *Gramática descriptiva de la lengua española*, 25th chapter, II, 1575-1629, Espasa, Madrid.

Michelena, L., 1961, *Fonética histórica vasca*, 2d edition (1976), Publicaciones del Seminario "Julio de Urquijo" de la Diputación Provincial de Guipúzcoa, San Sebastián.

—, DGV (*Dicionario General Vasco*), 1987-2001, 13 vol., Euskaltzaindia, Desclée de Brouwer & Mensajero, Bilbao.

Miyagawa, S., 1989, *Structure and Case Marking in Japanese*, "Syntax and Semantics" 22, Academic Press, London.

Ortiz de Urbina, J., 1989, *Parameters in the Grammar of Basque*, Foris, Dordrecht.

Oyharçabal, B., 1993, "Notion d'objet et cas absolutif en basque", *Le gré des langues* 5, ii-xxi, Paris.

—, 1994, "Contribution de la comparaison typologique à une analyse des rapports ergativité-(in)transitivité en basque". In J-B. Orpustan (dir.) *La langue basque parmi les autres*, Proceedings of the colloquium of Bayonne (27-28/09/1993), 115-147, Izpegi, Baigorri.

Perlmutter, D., 1978, "Impersonal passives and the Unaccusatives Hypothesis", *Proceedings of the Fourth Annual Meeting of the Berkeley Linguistics Society*, 157-189, Berkeley Linguistics Society, Berkeley, University of California.

— & Postal, P., 1984, "The 1-Advancement Exclusiveness Law". In D. Perlmutter & C. Rosen (eds.), *Studies in Relational Grammar 2*, University of Chicago Press.

Pesetsky, D., 1995, *Zero Syntax*, MIT Press, Cambridge, MA.

Pustejovsky, J., 1995, *The Generative Lexicon*, MIT Press, Cambridge, MA.

Pinker, S., 1989, *Learnability and Cognition: The Acquisition of Argument Structure*, MIT Press, Cambridge, MA.

Pylkkänen, L., 1999, "On Causativity and Causation". In C. Tenny & J. Pustejovsky (eds.) *Events as Grammaticals Objects, CSLI Publications*, Stanford, California, 417-442.

—, 2001, "Argument of causatives", ms., MIT, Cambridge, MA.

—, 2002, *Introducing Argument*, Ph-D dissertation, MIT, Cambridge, MA.

Rijk, R. de, 2001, "L'antipassif basque et l'hypothèse de Levin", *Lapurdum* 7, Bayonne.

Rebuschi, G., 1985, "Théorie du liage et langues non-configurationnelles: Quelques donnéees du basque navarro-labourdin", *Euskera* 30.2, 389-433.

—, 1995, "Weak and strong genitive pronouns in Northern Basque: A diachronic perspective". In J. I. Hualde, J.A. Lakarra & R.L. Trask (eds.), *Towards a History of the Basque language*, Johns Benjamins, Amsterdam / Philadelphia, 313-356.

—, 1997, "Absolute and Relativized Locality in the Binding Theory". In id, *Essais de linguistique basque*, Supplements of *ASJU*, University of Basque Country, Bilbao, 279-300.

Reinhart, T. & Siloni, T., (to appear), "Against the Unaccusative Analysis of Reflexives", *Studies in Unaccusativity: The Syntax-lexicon Interface*, CUP.

Safir, K., 1996, "Semantic Atoms of Anaphora", *Natural Language & Linguistic Theory* 14.3, 545-589.

Sarasola, I., 1977, "Sobre la bipartición inicial en el análisis en constituyentes", *ASJU* 11, 51-90.

Tenny, C., 1994, *Aspectual Roles and the Syntax-Semantics Interface*, Kluwer, Dordrecht.

Williams, E., 1981, "On the notions 'Lexically related" and 'Head of a word'", *Linguistic Inquiry* 12.2, 245-274.

Zabala, I., 1993, *Predikazioaren teoriak gramatika sortzailean (euskararen kasua)*, unpublished Ph-D dissertation, University of Basque Country.

CAUSATION AND SEMANTIC CONTROL. DIAGNOSIS OF INCORRECT USES IN MINORIZED LANGUAGES

Igone Zabala

Abstract

Data obtained about the use of language for special purposes can be a valuable tool for checking theoretical predictions concerning semantics and syntax of certain kinds of verbs.

The goal of this paper is to find some kind of test to prevent incorrect uses in Basque specialized texts, without refusing available but previously not achieved uses. We characterize agents, circumstantial causes, instruments and themes according to the features self-initiator, controller, controlled and affected. We also compare unaccusatives which take part in the "causative alternation" with pure unaccusative verbs, with regard to the above-mentioned features. We conclude that both bear subjects which can be defined as internal causes. Neverheless, uncontrollable causes are avoided as affected themes, and either they avoid all kinds of transitive constructions or they allow transitive constructions with circumstantial causes as subject. On the other hand, we claim that clauses with an instrument as subject bear an empty cause, which complement allows restricted kinds of aspectual interpretations. We also show that experiencer predicates can be related to the above mentioned semantic features. When experiencers take part in a transitive construction they are affected objects, but experiencers can also be internal causes. Finally, we explore the selectional behavior of some affixes which derive deverbal nouns or adjectives, and we conclude that the above mentioned semantic features are relevant in the competence of speakers when they generates deverbal nouns or adjectives.

1. Introduction

Basque dictionaries usually provide information about the auxiliary required or allowed by verbal entries. Basque verbs agree with subjects, direct objects and indirect objects, which show ergative, absolutive or dative case, the ergative and dative cases being morphologically realized affixes, while absolutive is a morphologically unmarked case. Therefore when the user of the dictionary knows the auxiliary required by a verb, he also knows the case which each argument must be assigned. However, in order to ensure that the speaker using the dictionary to solve competence doubts will generate acceptable sentences, much further information about arguments is required.[1] For

[1] More data about the information provided by Basque dictionaries for verbal entries in Gràcia et al. (2000: 586-589).

example Basque dictionaries characterize the verb *adierazi* 'to express' with the dyadic or the triadic auxiliary. Example (1a) illustrates a sentence in which this verb is correctly used: an ergative, an absolutive and a dative argument agree with the verb. Nevertheless, dictionaries do not provide the user with any information that will prevent (1b).[2]

(1) a. Jonek *adierazi* digu herrialde hau oso hezea dela.
Jon-ERG notify AUX-3sA-2pD-3sE country this(ABS) very humid is-that
'Jon notified us that this country is very humid.'

b. *Zelai berdeek *adierazi* digute herrialde hau oso hezea dela.
field green-ERGpl. notify AUX-3sA-2pD-3sE country this(ABS) very humid is-that
'The green fields notified us that this country is very humid.'

Since (1a) and (1b) have similar syntactic structures, one may hastily conclude that what accounts for the ungrammaticality in (1a) is animacy: animate subjects must be required with the verb *adierazi* 'to express'. The examples in (2) seem to corroborate this idea, since the verb *iradoki* 'to inform, to suggest' is acceptable with both kinds of subjects, as illustrated in (2).

(2) a. Jonek *iradoki* digu herrialde hau oso hezea dela.
Jon-ERG inform AUX-3sA-1pD-3sE country this(ABS) very humid is-that
'Jon informed us that this country is very humid.'

b. Zelai berdeek *iradoki* digute herrialde hau oso hezea dela.
field green-ERGpl. inform AUX-3sA-1pD-3pE country this(ABS) very humid is-that
'The green fields informed us that this country is very humid.'

However, example (3a) is perfectly acceptable, even with an inanimate subject, which suggests that the verb *adierazi* has more intricate argumental requirements. Comparing the inanimate subjects in (3) *(gezi* 'arrow' and *zelai berde* 'green field'), we conclude that they differ in a crucial aspect: (3a) expresses an intentional eventuality, because there is a voluntary decision of someone for arrows to express the direction of flow. However, what is expressed by *zelai berde* 'green field' can not be intentional.

(3) a. *Geziek* jarioaren noranzkoa *adierazten* dute.
arrow-ERGp. flow-DET-GEN direction-DET(ABS) express AUX-3sA-3pE
'Arrows indicate the direction of the flow.'

b. **Zelai berdeek* *adierazi* digute herrialde hau oso hezea dela.
field green-ERGp. indicate AUX-3sA-1pD-3pE country this (ABS) very humid is-that
'The green fields notified us that this country is very humid.'

[2] Concerning wrong uses of Basque verbs in specialized contexts see Zabala (1995) and Odriozola & Zabala (2000).

Another example of incorrect usage involves the verbs *esleitu* 'assign' and *egotzi* 'attribute', which are also frequently confused, since they bear the same valency, the same auxiliary and the same kind of arguments concerning case.

(4) a. Aurkitu duten substantziari X izena *esleitu* diote.
find AUX-REL substance-DET X name assign AUX-3sA-3sD-3pE
'The substance found has been assigned the name X.'

b. *Substantziari zahartzearen azkartzea *esleitu* diote.
Substance-DAT ageing-GEN acceleration-DET assign AUX-3sA-3sD-3pE
'The substance has been assigned the acceleration of ageing.'

The incorrect sentence in (4b) becomes an acceptable one with the verb *egotzi* 'attribute' (5). In this confusion inanimate objects are involved. What semantic characteristics are responsible for this confusion?

(5) Substantziari zahartzearen azkartzea egotzi diote.
Substance-DAT ageing-GEN acceleration-DET attribute AUX-3sA-3sD-3pE
'They attributed to the substance the acceleration of ageing.'

When speakers use the language for special purposes, verbs, just like any other kind of words, are sometimes used in a new way, and these uses differ apparently from those reflected in general dictionaries. Lexicographic work is based on the real corpus of a language and thus the information contained in dictionaries must be used very carefully. For example, the verb *adierazi* is used in (6) as intransitive with the monadic auxiliary *izan* 'to be', that is, it is used as an unaccusative verb.[3] Nevertheless, dictionaries assign to this verb only a transitive use, by codifying only the dyadic auxiliary. A superficial corrective attitude should perhaps rule out this use, arguing that the wrong auxiliary has been chosen. In fact this kind of positioning is rather frequently found in the context of minorized languages such as Basque, in which speakers, teachers, correctors, language academy members and even linguists are always worried about damage to the language.

(6) Hormona batek agindua ematen dienean, geneak
hormone a-ERG order-DET(ABS) give AUX-3sA-3pD-3sE-when, gene-DETp(ABS)
adierazi egiten dira
express do AUX-3pA
'When a hormone gives the order, gene expression takes place.'

In our opinion, there is a much more constructive, and thus much preferable attitude that we could take. By examining the corpus of a certain language, we collect the different ways in which a verb has been used in this language, but should make allowance for possible uses of this verb that have never been fulfilled because the

[3] Basque intransitive unaccusative verbs take the auxiliary *izan* 'to be', but intransitive unergative verbs take the auxiliary *edun* 'to have'. See for example Levin (1983) and Laka (1995).

required context has never happened. Specialized contexts are good candidates for these possible options to be fulfilled. Moreover, one can expect new uses to be much more frequent in minorized languages which, in their recovering process, are extending more and more to previously non-existent specialized areas. Thus, data obtained about the use of language for special purposes can be a valuable tool for checking theoretical predictions concerning semantics and syntax of certain kinds of verbs.

In this paper we examine the behavior of certain Basque verbs in the light of a number of concepts developed in the fields of generative semantics and syntax, such as agentivity, control, causation and animacy. The goal is to find some kind of test to prevent incorrect uses, without refusing available but previously not achieved uses. Section 2 collects some classic tests for detecting control and intentionality. We characterize agents, circumstantial causes, instruments and themes according to the features self-initiator, controller, controlled and affected. In the end of the section the example in (3a) is explained by attributing to the subject the instrumental theta role. Section 3 explores the behavior of unaccusatives which take part in the 'causative alternation' and compare them with pure unaccusative verbs, with regard to the above-mentioned features. We conclude that both bear subjects which can be defined as internal causes. Internal causes can be characterized at the same time as self-initiator of the event and affected by this event, although whether the cause is controllable or uncontrollable should be distinguished. Controllable causes can also be affected themes and thus can be found in transitive constructions with agent subjects. Uncontrollable causes are avoided as affected themes, and either they avoid all kinds of transitive constructions or they allow transitive constructions with circumstantial causes as subject. At the end of the section the example in (4) is explained by attributing to the subject an internal uncontrollable cause status. Section 4 explores the behavior of the two major classes of psychological predicates from the viewpoint of causation, control and affectedness. Section 5 is concerned with aspectual readings of transitive constructions with instrumental subjects. We attribute to these constructions reportive or ability readings. Different kinds of instruments should yield to different kinds of these aspectual readings. Section 6 explores the way in which syntactic instantiation of the verb affects the possibility of control of the event. In particular, we characterize the behavior of dative oblique arguments when they are interpreted as goal, possessor or starting point. Finally, section 7 is concerned with Basque affixes which derive deverbal nouns or adjectives. We explore their selectional behavior concerning features studied in the other sections, in order to show that they are relevant features in the competence of speakers when they generate new deverbal nouns or adjectives.

2. Causation, intentionality and control

Basque is an ergative language. Transitive constructions are characterized by a subject case-marked ergative and an absolutive object (7a), and they select the auxiliary *edun* 'to have'. As for intransitive constructions, unergative verbs case-mark the subject with ergative case, and also take the auxiliary *edun* 'to have' (7b). On the other hand, unaccusative constructions case-mark the subject with the absolutive case (the case of the direct object in transitive constructions) and take the auxiliary *izan* 'to be' (7c). Therefore Basque subjects are sometimes case-marked ergative and sometimes absolutive. This behavior contrasts with that of accusative languages in which subjects always show nominative case.

(7) a. Jonek atea ireki du.
Jon-ERG door-DET(ABS) open AUX-3sA-3sE
'Jon opened the door.'

b. Jonek bazkaldu du.
Jon-ERG have-lunch AUX-3sA-3sE
'Jon had lunch.'

c. Jon etxera joan da.
Jon(ABS) home-to go AUX-3sA
'Jon went home.'

However, knowing what auxiliary is required and what case is to be assigned to different arguments does not suffice to make sure that a sentence will be correctly generated. It is well known that predicates impose semantic restrictions on their arguments. For example, if we interchange the arguments in (7a) we make the sentence ungrammatical (8a). By contrast, both combinations of arguments are perfectly allowed with the verb *jo* 'hit' in (8b, c).

(8) a. *Ateak Jon ireki du.
door-ERG Jon(ABS) open AUX-3sA-3sE
'The door opened Jon.'

b. Ateak Jon jo du
door-ERG Jon(ABS) hit AUX-3sA-3sE
'The door hit Jon.'

c. Jonek atea jo du.
Jon-ERG ate-DET(ABS) hit AUX-3sA-3sE
'Jon hit the door.'

In the examples in (7) Jon is the argument that initiates all events expressed by the verbs *ireki* 'open, *bazkaldu* 'have lunch' and *joan* 'go'. Furthermore, in (7a) *atea* 'the door' is an object affected by the event initiated by Jon, because the door is opened as a result of this event. An affected argument is an internal argument which undergoes some change (state, location, possessor) (Tenny 1988, 1989, 1994). What is failing in sentence (8a)? Is the subject *ate* 'door' unable to initiate the event? The response must be no, because the verb 'open' can be used as unaccusative with *atea* 'the door' as subject (9). The door itself is able to initiate the action expressed by the verb *ireki* 'open', but it is not able to affect any object different from itself.

(9) Atea ireki da.
door-DET(ABS) open AUX-3sA
'The door opened.'

The sentence in (7c) with the unaccusative verb *joan* 'go' is very similar to (9): the subject itself (Jon) starts the event and is itself affected, because there is a change of location. However, there is an important difference between (7c) and (9), which concerns what is known as 'semantic control'. Authier and Reed (1991) define semantic control as "the possibility of canceling what is denoted by the predicate if the

subject of this predicate decides to stop doing it." Control entails intention. Two classic tests can be used to reveal this intentional control. The complement of obligatory control constructions of subject control verbs such as *try, endeavor, refuse, condescend* and *dare,* and of object control verbs such as *persuade, force, convince, tell* and *order,* must be an intentional action, that is, this action must be within the intentional control of the subject (Berman 1970, Lasnik and Fiengo 1974, Lasnik 1991). Only animate subjects can be subjects of obligatory control constructions:

(10) a. *Atea behartu dute [PRO irekitzen]
door-DET(ABS) obliged AUX-3sA-3pE PRO opening
'They obliged the door to open.'

b. Jon behartu dute [PRO etxera joaten]
Jon(ABS) force AUX-3sA-3pE house-ALAT going
'They forced Jon to go home.'

Control of the action requires animacy and intentionality: only animate beings can bear intention. Jackendoff (1993) argues that "purposes can only go with volitional acts" and all purposes presuppose an intention. The behavior of the two sentences above contrasts also in purpose clauses:

(11) a. Atea ireki da, *jendea sartzeko[4] [unaccusative interpretation]
door-DET(ABS) open AUX-3sA people come in-FOR
'The door opened, to let people in.'

b. Jon etxera joan da, ama agurtzeko.
Jon(ABS) home-to go AUX-3sA mother(ABS) greet
'Jon went home to greet his mother.'

Our first distinction must therefore be that between subjects of self-initiated events there are controllers of the action and there are non controllers of the action denoted by the verb. This will be our starting point for looking at different kinds of subjects. Causatives are transitive constructions in which an argument causes a second argument to be affected. These constructions are assumed to express events that can be broken down into two subevents: a causative subevent and a resultative subevent (Pustejovsky 1991, 1995). Some verbs allow different kinds of cause arguments: agents or interactive causes, instruments, circumstantial causes.[5] Inanimate subjects (instruments and

[4] The purpose clause is allowed with an impersonal interpretation, that is if it means *The door has been opened,* but not if it means *The door opened.*

[5] The term 'cause' is often used for any argument related to the causative subevent, but it is also used to make reference to non-controller causes such as forces of nature (*wind, rain, etc.*). Therefore, we use here the terminology in Kural (1997), i.e. interactive and circumstantial causation. As for interactive causation, there is direct interaction between the cause and the affected object, so the cause is an agent. On the other hand, with circumstantial causation, the cause may only create the circumstances in which the object is affected.

natural forces) are avoided with purposes (12), because they lack intentionality, and thus they are not the controller of the event.

(12) a. *Jonek* atea ireki du, sartzeko [interactive cause = agent]
Jon-ERG door-DET(ABS) open AUX-3sA-3sE, get in-for
'Jon opened the door, in order to get in.'

b. *Giltzak* atea ireki du, *sartzeko [instrument]
Key-DET-ERG door-DET(ABS) open AUX-3sA-3sE, *get in-for
'The key opened the door, *in order to get in.'

c. *Haizeak* atea ireki du, *sartzeko [circumstantial cause]
wind-DET-ERG door-DET(ABS open AUX-3sA-3sE, *get into-for
'The wind opened the door, *in order to get in.'

Moreover, causes and instruments differ in their availability to appear with an agent: instruments are allowed as adjuncts when the agent fills the subject's site (13a) but circumstantial causes are avoided with agents (13b) (Shin 1998). This contrast reveals that instruments can or must be controlled by a controller argument (agent), whereas circumstantial causes avoid being controlled.

(13) a. *Jonek* atea ireki du *giltzaren bidez.*
Jon-ERG door-DET(ABS) open AUX-3sA-3sE key-DET-GEN by means of
'Jon opened the door using the key.'

b. **Jonek* atea ireki du *haizearen bidez.*
Jon-ERG door-DET(ABS) open AUX-3sA-3sE wind-DET-GEN by means of
'Jon opened the door using the wind.'

This makes us think of circumstantial causes and agents as both sides of the same theta position, although instruments represent a different theta position. Following Minkoff (1997), AGENT thematic relations are the result of applying to the argument filling a certain theta position a certain lexical interpretation made available in accordance with abstract syntactic principles, referred to as "animacy entailment". When "animacy entailment" doesn't apply the CAUSER thematic relation is obtained. Animacy is required for a subject to be a controller of the action denoted by the verb. However, animate subjects related to the causative subevent of a causative construction are not necessarily agents. In fact, with a verb such as *ireki* 'to open' in (14) there are two possible readings, which are disambiguated using control adverbs: the adverb *nahita* 'voluntarily' gives us the control reading (14a), and the adverb *nahi gabe* 'involuntarily' gives us the 'out of control' reading (14b). That is, the control reading requires intentionality. Only the control reading is compatible with the AGENT thematic role, and non controller animate causes are therefore circumstantial causes. We call the reading in (14b) 'out of control reading', following the terminology in Demirdache (1997).

(14) a. Jonek atea ireki du nahita [control reading: agent]
Jon-ERG door-DET(ABS) open AUX-3sA-3sE voluntarily
'Jon opened the door voluntarily.'

b. Jonek atea ireki du nahi gabe
Jon-ERG door-DET(ABS) open AUX-3sA-3sE involuntarily
'Jon opened the door involuntarily.'
[out of control reading: circumstantial cause]

However, instruments are avoided with a circumstantial cause (15a), and therefore appear to require voluntary actions. That is, the presence of an instrument in a causative construction characterized by optional animacy entailment makes animacy entailment obligatory. We conclude therefore that even when the instrument fulfills the subject position, an (empty) interactive cause is necessary in the causative subevent (15b). That is, sentences with instrumental subjects require controlled eventualities: instruments presuppose interactive causes. The conclusion is that instruments are not able to initiate an event, but interactive and circumstantial causes are.

(15) a. *Haizeak atea ireki du giltzaren bidez.
wind-DET-ERG door-DET(ABS) open AUX-3sA-3sE key-DET-GEN by means of
'*The wind opened the door using the key.'

b. [INTERACTIVE CAUSE [Giltzak atea ireki du]]
[I. CAUSE [key-DET-ERG door-DET(ABS) open AUX-3sA-3sE]]
[INTERACTIVE CAUSE [The key opened the door.]]

Instruments take part in certain kinds of control constructions. For example, verbs such as *serve*, *help* and *suffice* require an instrument as subject of their clausal complement (Higgins 1973). Since instruments are not controllers, another controller is required in these constructions, in order to control the category PRO.

(16) Giltzak balio du [PRO atea irekitzeko]
key-DET-ERG serve AUX-3sA-3sE door-DET(ABS) open-FOR
'The key serves to open the door.'

Causative verbs often allow unaccusative variants. Such verbs are referred to as verbs with "causative alternance". Basque unaccusative constructions (17a) are very similar to impersonals (17b) (see Levin 1983): both constructions require the auxiliary *izan* 'to be', and both case-mark the subject absolutive. However, the two types of construction behave differently with the adverb *berez* 'spontaneously'. This adverb expresses the absence of an external cause, and it is therefore avoided with impersonals like (17b), since impersonals presuppose a non specified empty subject. *Eskolako atea* 'the school's door' in (17b) is not a subject, but a direct object.

(17) a. Atea (berez) ireki da.
door-DET(ABS) (spontaneously) open AUX-3sA
'The door opened spontaneously.'

b. Eskolako atea zortzietan irekitzen da (*berez).
school-GEN door-DET(ABS) eight-at open-IMPF AUX-3sA (*spontaneously)
'The door of the school is opened at eight o'clock.'

With the verb *ireki* 'open', "animacy entailment" is optional, but as Minkoff (1997) pointed out, there are some verbs which assign a theta role to the subject characterized by an obligatory animacy entailment. This obligatoriness is illustrated by examples such as (18), in which an inanimate subject is clearly avoided with the verb *margotu* 'to paint'.

(18) a. Jonek atea zuriz margotu du.
Jon-ERG door-DET(ABS) white-INSTR paint AUX-3sA-3sE
'Jon painted the door white.'

b. *Haizeak atea zuriz margotu du.
wind-DET-ERG door-DET(ABS) paint AUX-3sA-3sE
'The wind painted the door white.'

Verbs with obligatory agents as subjects absolutely lack unaccusative variants (19). *Atea* 'the door' is an affected object in (19a) but it lacks the intrinsic property necessary for being spontaneously painted. The eventuality expressed by the verb *margotu* 'to paint' requires an external cause. On the other hand, however, the door has the relevant intrinsic property to open by itself. Chierchia (1989) argues that unaccusative variants of causative verbs are similar to reflexives in which the only argument is both the cause and the affected theme. He argues that these constructions have a causative event, which is interpreted statively.[6] Pustejovsky (1995) agrees with Chierchia: causative/ unaccusative alternation should result with causative events non specified for a head. An unaccusative or a transitive causative will be determined by which subevent in the semantic representation is headed: causative transitive verbs without an unaccusative counterpart (*kill, murder*) should be left-headed, and pure unaccusatives should be right-headed. In our opinion this headedness is related to the possibility of the second argument to self-initiate the event and the possibility of this argument to be affected. In the examples in (19) the door is affected by the event initiated by Jon, but it is not able to self-initiate this event.

(19) a. Jonek atea margotu du.
Jon-ERG door-DET(ABS) paint AUX-3sA-3sE
'Jon painted the door.'

b. *Atea (berez) margotu da.
door-DET(ABS) (spontaneously) paint AUX3sA
'The door got painted spontaneously [i.e., by itself].'

The empty subject of impersonal constructions licenses final adjunct clauses. These clauses are avoided with unaccusatives, because they lack an agent, and so they do not express an intentional controlled action:

(20) a. *Atea berez ireki da, ikasleak sartzeko.
door-DET(ABS) spontaneously open AUX3sA, student-DETpl(ABS) get in-for
'The door got opened spontaneously [i.e. got itself open] in order for the students to get in.'

---

[6] Reflexive constructions would differ from unaccusatives because the causative event is dynamic.

b. Eskolako atea zortzietan irekitzen da, ikasleak
school-GEN door-DET(ABS) eight-at open-IMPF AUX-3sA student-DETp(ABS)
sartzeko.
get into-for
'The door of the school is opened at eight o'clock, in order for the students to get in.'

The adverb *berez* 'spontaneously' is a subject-oriented adverb which invalidates every possibility for an external cause to be presupposed. If this adverb applies to an agent (21) it invalidates the possibility of an external causer.

(21) Jonek berez ireki du atea (inork
Jon-ERG spontaneously open AUX-3sA-3sE door-DET(ABS) (anybody-ERG
behartu gabe).
force without)
'Jon opened the door spontaneously (, without being forced by anybody).

In this sense, sentences with an instrumental subject behave similarly to impersonal constructions with regard to the adverb *berez* 'spontaneously'. That is, these constructions require an (empty) cause, and therefore, the negation of this cause via an adverb is avoided (22).

(22) Giltzak atea ireki du (* berez).
Key-DET-ERG door-DET(ABS) open AUX-3sA-3sE spontaneously
'The key opened the door spontaneously.'

There are verbs with transitive/unaccusative alternations which exhibit interesting behaviour. An example is the case of the verb *abiatu* 'set off', taken from Etxepare (2003). In (23a) we have an unaccusative verb which self-initates the event. An inanimate subject is avoided with this verb, so the subject must be a controller (23b).

(23) a. Hiru langile abiatu dira.
three worker(ABS) set.off AUX-3pA
'Three workers set off.'

b. *Projektu berri bat abiatu da. [with the unaccusative reading]
project new one set.off AUX-3sA
'A new project set off.'

By contrast, in the transitive variant (24) the object must be an affected object (24a), and animate objects are avoided (24b). In a lexical causative construction, only one of the arguments can be a controller.

(24) a. Enpresak proiektu berri bat abiatu du.
company-DET-ERG project new one set.off AUX-3sA-3sE
'The company set off a new project.'

b. *Enpresak hiru langile abiatu ditu
company-DET-ERG three worker set.off AUX-3sA-3sE
'The company set off three workers.'

The event 'workers set off' can be caused but not controlled. This causation can be instantiated as a morphological causative (25). The subject of this construction is a circumstantial cause, and circumstantial causes are characterized by the absence of control.

(25) Enpresak hiru langile abiarazi ditu.
Company-DET-ERG three worker(ABS) set.off.CAUSE AUX-3pA-3sE
'The company set off three workers.'

The following table summarizes semantic characteristics reviewed in this section, related to different kinds of arguments.

	Agents	Circumstantial Causes	Instruments	Themes
intentionality = controller	+	–	–	–
self-initiator	+	+	–	–
controlled	–	–	+	+
affected	–	–	–	+

Turning to the example (3a), repeated here as (26a), we can characterize the inanimate subject *gezi* 'arrow', as an instrument, since it is compatible with an agent (26b), that is, it can be controlled and it is unable to express a self-initiated action (26c).

(26) a. *Geziek* jarioaren noranzkoa adierazten dute.
arrow-ERGp. flow-DET-GEN direction-DET(ABS) express AUX-3sA-3pE
'Arrows indicate the direction of the flow.'

b. Liburu honen egileek gezien bidez adierazi dute
book this-GEN autor-ERGp. arrow-GENpl. by means of express AUX-3sA-3pE
jarioaren noranzkoa.
flow-GEN direction
'The authors of this book expressed the direction of the flow using arrows.'

c. [CAUSE [*Geziek* jarioaren noranzkoa adierazten dute]]
arrow-ERGp. flow-DET-GEN direction-DET(ABS) express AUX-3sA-3pE
'Arrows indicate the direction of the flow.'

d. *Geziek berez adierazten dute
arrow-ERGp. spontaneously express AUX-3sA-3pE
jarioaren noranzkoa.
flow-GEN direction-DET(ABS)
'Arrows spontaneously express the direction of the flow.'

We can conclude that the verb *adierazi* 'to express', even with inanimate subjects, requires causative constructions expressing controlled eventualities or situations: the subject must be either human or an instrument.

3. External causes and internal causes

Intransitive predicates express an internal cause eventuality, when the only argument of the verb bears an intrinsic property that is responsible for the event to take place (Levin and Rappaport Hovav 1995: 92). In addition, internal cause verbs grammaticalize sometimes as unergative and sometimes as unaccusative (Mendikoetxea 1999).

(27) a. Haurra jaio da.
baby-DET(ABS) born AUX-3sA
'The baby is born.'

b. Izarrek distiratzen dute.
Star-DET-ERG shine-IMPER AUX-3sA-3pE
'Stars are shining.'

Internal cause verbs are predicated of animate beings or natural phenomena which bear intrinsic properties required for the event denoted to be possible. Mendikoetxea considers two tests to be relevant in order to delimit external and internal cause verbs: internal cause verbs lack transitive variants and are avoided with adverbs of the type *berez* 'spontaneously'. Moreover, she argues that internal cause verbs impose strict restrictions on the subject, in the sense that only a few entities are available for this function: only mammalians are born, only certain kinds of plants bloom, only metals go rusty, only certain kinds of objects shine and so on.

(28) a. Haurra jaio da.
baby-DET(ABS) born AUX-3sA
'The baby is born.'

b. *Sendagileak/ *sendagaiak/ *lurrikarak
doctor-DET-ERG/medicine-DET-ERG/earthquake-DET-ERG
haurra jaio du.
baby-DET(ABS) born AUX-3sA-3sE
'The doctor/the medicine/the earthquake borned the baby.'

c. ? Haurra berez jaio da.
baby-DET(ABS) spontaneously born AUX-3sA
'The baby was born spontaneously.'

Nevertheless, the sentence in (28c) can be improved if the sense of the adverb *berez* 'spontaneously' is justified, or if a circumstantial cause is added as a postpositional adjunct. Naturally, the circumstantial cause adjunct is incompatible with the adverb *berez*.

(29) a. Haurra berez jaio da, oxitozinarik gabe.
baby-DET(ABS) naturally born AUX-3sA oxitozine-PART without
'The baby was born naturally, without oxitozine.'

b. Haurra jaio da, lurrikararen kausaz.
baby-DET(ABS) born AUX-3sA, earthquake-because of
'The baby was born because of the earthquake.'

In any case, it is difficult to see how subjects of unaccusatives without transitive alternance (*haurra jaio da* 'the baby is born') differ from subjects of unaccusatives with transitive alternance (*atea ireki da* 'the door opened'). In both cases the only argument itself initiates the event, and it is both the cause and the affected object of this event. Moreover, both require a subject with a certain kind of intrinsic property. Certainly, verbs such as *erori* 'fall', *apurtu* 'break', *hondatu* 'ruin' allow transitive counterparts and are compatible with practically any kind of argument, but then there is the Basque verb *hil* 'die/kill', a verb with transitive/unaccusative alternance, but whose subject/object is restricted by the fact that only living creatures die.

(30) a. Jon hil da.
Jon(ABS) die AUX-3sA
'Jon died.'

b. Lapurrak Jon hil du.
thief-DET-ERG Jon(ABS) kill AUX-3sA-3sE
'The thief killed Jon.'

In the unaccusative counterpart (30a), to die is a self-initiated process, just like to be born is in (28a). The two kinds of event differ with regard to the possibility of the event's being controlled: babies are born in nine months and one can not directly cause a baby to be born in two months. With respect to dying/killing, ruining or breaking persons or things, there is a natural process whereby living creatures and objects are brought respectively closer and closer to death, ruin or breakage, but these processes can be controlled by an external agent, which can decide to terminate the process. By contrast, an uncontrollable self-initiated process can be interrupted, but an external agent can not decide to complete the event.

Non controllable internal causes are incompatible with agents, that is, with interactive causes, but they are allowed with circumstantial causes. However, the possibility for this cause to fulfill the subject site appears to be an idiosyncratic lexical property of each verb in each language. For example, Spanish equivalents of Basque verbs such as *loratu* 'bloom', *herdoildu* 'rust' —i.e., the verbs *florecer* and *oxidar*— are avoided with a cause subject (Mendikoetxea 1999), whereas Basque frequently allows transitive counterparts of these verbs. In any case, the subject can never be an agent nor an instrument.

(31) a. Zuhaitza loratu da.
tree-DET(ABS) bloom AUX-3sA
'The tree bloomed.'

b. Beroak zuhaitza loratu du.
Heat-DET-ERG tree-DET(ABS) bloom AUX-3sA-3sE
'The heat bloomed the tree [i.e. caused the tree to bloom].'

Another example is the Spanish verb *hervir* 'boil', which allows unaccusative/transitive alternations, whereas the Basque equivalent *irakin* 'boil' is only allowed in intransitive constructions, specifically in unergative constructions (32).

(32) a. Esneak irakin du.
Milk-DET-ERG boil AUX-3sA-3sE
'The milk boiled.'

b. *Jonek esnea irakin du.
Jon-ERG milk-DET(ABS) boil AUX-3sA-3sE
'Jon boiled the milk.'

Certain verbs of growth similar to *loratu* 'bloom' exhibit a peculiar alternation in Basque: when used as transitives they are causative agentive verbs, expressing removal of the thing grown (Gràcia et al. 2000, Etxepare 2002). For example the verb *kimatu* 'bud/trim'. In (33b) the subject is an agent and not a circumstantial cause as in (33c). The unaccusative verb (33a) allows a circumstantial cause as adjunct (33c) but not as subject, because the ergative subject must be an agent.

(33) a. Zuhaitza kimatu da.
tree-DET(ABS) bud AUX-3sA
'The tree budded.'

b. Jonek zuhaitza kimatu du.
Jon-ERG tree-DET(ABS) trim AUX-3sA-3sE
'Jon trimmed the tree.'

c. Zuhaitza kimatu da, eguraldi onaren kausaz.
Tree-DET(ABS) bud AUX-3sA, weather good-GEN because of
'The tree budded because of the good weather.'

Internal cause verbs can also grammaticalize as unergatives (27b, 32a). In this case, the internal cause takes ergative, and no transitive counterpart is ever allowed. Circumstantial causes must always be expressed as adjuncts.

(34) a. Beirak distiratu du, eguzkiaren kausaz.
glass-DET-ERG shine AUX-3sA-3sE sun-GEN because of
'The glass shone because of the sun.'

b. *Eguzkiak beira distiratu du.
sun-DET-ERG glass-DET(ABS) shine AUX-3sA-3sE
'The sun made the glass shine/shone on the glass.'

Therefore we conclude that internal causes are arguments that self-initiate the event and are the only argument affected by this event. Internal causes include both controllable and uncontrollable causes. The former can take part in transitive constructions with agent subjects or circumstantial cause subjects. The latter take part only in intransitive constructions (unergatives or unaccusatives), or else in unaccusative constructions or transitive constructions with circumstantial cause subjects. But they avoid agent subjects, because they are uncontrollable.

Summarizing, an argument is an internal cause if it initiates the event itself and it is itself the only argument affected by this event. There are controllable and uncontrollable internal causes. Uncontrollable internal causes are incompatible with agents, but depending on the verb and on the language, they can take part in transitive constructions with circumstantial cause subjects. The subject of unergative constructions is always an uncontrollable internal cause

		Unergative constructions	Unaccusative constructions	Transitive constructions with agent subjects	Transitive constructions with circumstantial cause subjects
Internal causes +self-initiators	Controllable internal causes	–	+	+	+
+ self-affected	Uncontrollable internal causes	+	+	–	+

Natural processes are usually internal cause eventualities, which are grammaticalized as intransitive verbs. This is the case of the example in (6) repeated here as (35). The intrinsic property of genes is to bear information, and the eventuality expressed in (35) is an uncontrollable internal cause eventuality. Therefore it is not strange for this eventuality to have been grammaticalized as an unaccusative construction.

(35) Hormona batek agindua ematen dienean,
hormone a-ERG order-DET(ABS) give AUX-3sA-3pD-3sE-when,
geneak *adierazi* egiten dira
gene-DETpl(ABS) express do AUX-3plA
'When a hormone gives the order, gene expression takes place.'

4. Experiencer predicates

Psychological predicates can also be related to semantic features analyzed in previous sections. This kind of predicate is characterized by an experiencer argument. Following Belleti & Rizzi (1988), two major classes are distinguished: on the one hand we have verbs such as *temere,* in which the experiencer is the subject; on the other we have verbs such as *preoccupare* in which the experiencer is the object. Belleti & Rizzi (1988) and also Grimshaw (1990) attribute to both kinds of predicates an experiencer-theme argument structure. Belleti & Rizzi explained the striking behavior of the *preoccupare*-class verbs in respect to binding and control by claiming that the subject of these verbs is the internal argument (theme) of the verb. Pesetsky (1987, 1995) and Pustejovsky (1995), however, relate these experiencer predicates of the *preoccupare*-class with causative predicates such as *kill.*[7] The argument structure of these predicates should thus be cause-experiencer.

[7] Zabala (1993: 202-204) agrees with Pesetsky (1987) and attributes to the surface subject of Basque psych-verbs such as *beldurtu* 'frighten' the cause θ-role.

Predicates of the *preocupare* class such as Basque *kezkatu* 'worry', *beldurtu* 'frighten', *lotsatu* 'shame' often present inchoative/causative alternances and different types of subjects are allowed. The cause subject can be either animate (36a) or inanimate (36b, c). Furthermore, animate subjects allow either control or out of control readings (36a), while inanimate subjects allow either circumstantial causes (36b) or instruments (36c). Inchoative constructions are also found with this kind of verb (36d).

(36) a. Jonek haurra *beldurtu* du nahita/ nahi gabe
Jon-ERG child-DET(ABS) frighten AUX3sA-3sE voluntarily/involuntarily
'Jon frightened the child voluntarily/involuntarily.'

b. Ekaitzak haurra *beldurtu* du.
storm-DET-ERG child-DET(ABS) frighten AUX3sA-3sE
'The storm frightened the child.'

c. Makilak haurra *beldurtu* du[8].
stick-DET-ERG child-DET(ABS) frighten AUX3sA-3sE
'The stick frightened the child.'

d. Haurra *beldurtu* da (berez).
child-DET(ABS) get.frightened AUX-3sA (spontaneously)
'The child got frightened (spontaneously)'.

e. Haurra *beldur* da.
child-DET(ABS) fear is
'The child is frightened.'

Therefore in this kind of verb experiencers are internal arguments. When experiencers take part in a transitive construction they are affected objects: they suffer a change in their psychological state. Psychological state itself is expressed using the noun *beldur* 'fear' (36e) instead of the verb *beldurtu* 'frighten'. Furthermore, an experiencer can itself initate a change of psychological state and also be affected by this change (36d). Experiencers can thus be internal causes. Moreover, the experiencer role requires animacy (37a). Finally, experiencers are uncontrollable arguments (37b).

(37) a. *Atea *beldurtu* da.
door-DET(ABS) get.frightened AUX-3sA
'The door got frightened.'

b. *Haurra behartu dute [PRO *beldurtzen*]
child-DET(ABS) force AUX-3sA-3sE getting.frightened
'They forced the child to get frightened.'

Predicates of the *temere* class such as *maitatu* 'love', *miretsi* 'admire', *pairatu* 'suffer', *gustatu* 'like', *interesatu* 'interest', are always dyadic predicates. This is consistent with the

[8] *Jonek haurra beldurtu du makilaren bidez* 'Jon frightened the child with the stick' is also possible. The subject in (36c) is thus an instrument.

experiencer-theme argument structure, since the theme is usually obligatory. Some verbs of this kind case-mark the experiencer ergative (38a) and some others dative (38b).

(38) a. Jonek zure lana *miresten* du.
Jon-ERG your work-DET(ABS) admire AUX-3sA-3sE
'Jon admires your work.'

b. Joni zure lana *gustatzen* zaio.
Jon-DAT your work-DET(ABS) like AUX-3sA-3sD
'Jon likes your work.'

With this kind of verb, the experiencer also requires animacy and is uncontrollable. However, psychological verbs of this class always express psychological states and the absolutive theme is never affected.

Psych-verbs		Experiencer	External Cause	Theme
Preocupare-class *Beldurtu* 'frighten' *kezkatu* 'worry'	Transitives ERG-ABS	+ (ABS)	+ (ERG)	–
	Unaccusatives ABS	+(ABS) Internal Cause	–	–
Temere-class *miretsi* 'admire' *gustatu* 'like'	Dyadic ERG-ABS	+ (ERG)	–	+ (Non Affected)
	DYADIC DAT-ABS	+ (DAT)	–	+ (Non Affected)

5. Instruments as subjects and aspectual reading

In section 2 we argued that clauses with an instrument as subject bear an empty cause. This empty cause selects different types of predicates (states or events), but actions are avoided, since actions require agents as subjects. Belvin (1998) argues that predicates that select eventualities, such as causative predicates, "select for one of these three eventualities (states, events and actions) as their internal argument", and "there is a very close relation between event type and θ-role properties of a predicate".

This selection has syntactic consequences when instruments behave as subjects. The complement of the empty cause allows restricted kinds of interpretations: some constructions result in a reportive or in a futurate reading and some others have an ability reading. Giorgi & Pianesi (1997) call reportive reading the kind obtained in sentences of the type "In DP S" or narration of different kinds of events. Copley (2000) defines the futurate reading of a sentence as a future-oriented eventuality that is accept-

able with plannable eventualities, but is avoided with unplannable eventualities. We take these definitions from Alcázar (2002).[9]

Note that example (3) easily allows the "In DP S" construction (39a). Therefore a reportive reading is obtained. This reportive reading with instrumental subjects is characterized by aspectual restrictions: on the one hand, the perfective aspectual affix is avoided (39b); on the other, with the future aspectual affix (*-ko*), a futurate reading is obtained: the sentence instantiates the plan for arrows to be used as a symbol.

(39) a. Liburu honetan, *geziek* jarioaren noranzkoa adierazten dute.
Book this-in arrow-ERGp. flow-GEN direction-DET(ABS) express-IMP AUX
'In this book arrows indicate the direction of the flow.'

b. *Geziek jarioaren noranzkoa adierazi dute.
arrow-ERGp. flow-GEN direction-DET(ABS) express-PERF AUX
'Arrows indicated the direction of the flow.'

c. (Liburu honetan) geziek jarioaren noranzkoa adieraziko dute.
(book this-in) arrow-ERGp. flow-GEN direction-DET(ABS) express-FUT AUX
'In this book arrows will indicate the direction of the flow.'

As for the ability reading, it is obtained with subject instruments such as *giltza* 'key'. All aspectual affixes are allowed, and the reading can be paraphrased as 'X serves to/for S' (X = instrument subject). With the future aspectual affix, the sentence expresses a prediction about the ability of the key to open a certain door.

(40) a. Giltza honek atea irekitzen du.
Key this-ERG door-DET(ABS) open-IMPERF AUX
'This key opens the door.' (This key serves to open the door)

b. Giltza honek atea ireki du.
Key this-ERG door-DET(ABS) open-PERF AUX
'This key opened the door.' (This key has served to open the door)

c. Giltza honek atea irekiko du.
Key this-ERG door-DET(ABS) open-FUT AUX
'This key will open the door.' (This key will serve to open the door)

With another type of instrument, and so with another kind of eventuality, different kinds of aspectual suffixes are allowed. For example with the instrument *giltza* 'key', the

[9] Some Basque verbs bear synthetic and analytic forms, while others bear only analytic forms. Alcázar assigns to synthetic forms four available readings: progressive, reportive, habitual and futurate. Alcázar (2002) characterized aspectual interpretation of Basque verbs and claimed that the reportive and futurate are some of the interpretations of Basque synthetic forms of Basque trinko verbs: *Film honetan, espioiak eskaileretan gora doanean mikrofilmak ezkerreko poltsikoan dakartza* 'In that film, when the spy goes up the stairs, he brings the microfilms in his left pocket.'/*Mikel bihar dator* 'Mikel is coming tomorrow'.

eventuality *ireki* 'open' is an achievement, but the instrument *bolaluma* 'pen' in (41) is related to an activity *idatzi* 'write'. In this case, an ability reading is also obtained, but the perfective or future affixes are avoided, because the readings allowed by the empty cause are hard to obtain with an activity and those aspectual affixes.

(41) a. Bolaluma honek idazten du.
pen this-ERG write-IMPERF AUX-3sA-3sE
'This pen writes.' (This pen serves for writing)

b. Bolaluma honek ez du idazten.
pen this-ERG not AUX-3sA-3sE write-IMPERF
'This pen does not write.' (This pen does not serve for writing)

c. ?Bolaluma honek idatzi du.
pen this-ERG write-PERF AUX-3sA-3sE
'This pen wrote.'

d. ?Bolaluma honek idatziko du.
pen this-ERG write-FUT AUX-3sA-3sE
'This pen will write.'

6. Argument structure, syntactic instantiation of the verb and control of the event

Authier and Reed (1991) define control as "the possibility of canceling what is denoted by the predicate if the subject of this predicate decides to stop doing it", Is control determined only by the intentionality of the subject? Syntactic instantiation of the verb affects 'control', which suggests that the subject is not the only element involved in control. In section 3 we extended the concept of control to the possibility of intentionally causing the event to start or to be culminated. The nature of the direct internal argument of a verb is also relevant for determining the aspectual nature of the event. In particular, affected objects, paths and goals have been generally described as arguments which affect the delimitedness of the event and so its aspectual interpretation (Tenny 1988, 1989, 1994). An affected argument is an internal argument that undergoes some change of location, possessor or state. A path is a distance traveled. And a goal expresses the endpoint of an event. We argue that syntactic instantiation of these arguments also affect the interpretation of the event as controlled/non controlled.

(42) a. Jonek papera eraman du.
Jon-ERG paper-DET(ABS) carry AUX-3sA-3sE
'Jon carried the paper.'

b. Jonek papera eraman du etxean zehar.
Jon-ERG paper-DET(ABS) carry AUX-3sA-3sE house-INES through
'Jon carried the paper through the house.'

In (42) there is an agent and an affected object, since the paper changes location. In (42a) the event is delimited and the subject controls the action and the affecting of the

object. In (42b) there is a path and the subject must control the action along this path. The event is not delimited because there is not any information about the end of this action.

(43) a. Jonek papera eraman du bulegoraino.
Jon-ERG paper-DET(ABS) carry AUX-3sA-3sE office-ALAT
'Jon carried the paper to the office.'

b. Jonek papera eraman dio Mireni.
Jon-ERG paper-DET(ABS) carry AUX-3sA-3sD-3sE Miren-DAT
'Jon carried the paper to Miren.'

In (43a) and (43b) a goal and a beneficiary argument have been added respectively. Both arguments express the end of the path entailed by the event. Both delimit the event but only the beneficiary (43b) requires control of the action by the subject. Note that we can substitute the animate subject Jon by an inanimate subject such as *haize* 'wind' (44). In such cases, however, the dative argument can not be interpreted as a beneficiary, but must necessarily be interpreted as the possessor of the paper and thus the starting point of the event.

(44) a. Haizeak papera eraman du.
'The wind carried the paper.'

b. Haizeak papera eraman du etxean zehar.
'The wind carried the paper through the house.'

c. Haizeak papera eraman du bulegoraino.
'The wind carried the paper to the office.'

f. Haizeak papera eraman dio Mireni.
'The wind carried Miren's paper'

'The wind carried the paper to Miren.'

Minkoff (1997) claims that the oblique object of a verb is a beneficiary 'only if the argument responsible for causing the activity denoted believes that this object could (be seen to) acquire some power over the theme by receiving it'. Note that the office in (43a) is a goal and does not acquire any power over the paper. The subject is required to have a purpose in order for the oblique object to be a beneficiary. The verb *eraman* 'carry' allows the out of control adverb *nahi gabe* 'involuntarily' (45). As can be expected, the dative oblique object can not be interpreted as a beneficiary with the out of control construction. It is interpreted as a possessor or as a goal but not as a beneficiary, since the argument responsible for causing the activity lacks any intention or belief concerning this argument.

(45) Jonek papera eraman dio Mireni nahi gabe.
Jon-ERG paper-DET(ABS) carry AUX-3sA-3sD-3sE Miren-DAT involuntarily
'Jon carried Miren's paper involuntarily.'
'Jon carried the paper to Mary involuntarily.' [goal]

With the verb *joan* 'go', a dative with the possesor interpetation (46b) makes the event denoted by the verb *joan* 'go' a non controlled action.

(46) a. Jonen semea Amerikara joan da.
Jon-GEN son-DET(ABS) America-ALAT do AUX-3sA
'Jon's son went to America.'

b. Joni semea Amerikara joan zaio.
Jon-DAT son-DET(ABS) America-ALAT go AUX-3sA-3sD
'(It happened to Jon that) his son went to America.' 'Jon's son went to America.'

This fact appears clear when the sentence is inserted as the complement of a control verb.

(47) a. Jonen semea behartu dute [PRO Amerikara joaten]
Jon-GEN son-DET(ABS) force AUX-3sA-3pE America-ALAT going
'Jon's son has been forced to go to America.'

b. *(Jonen) semea behartu dute [PRO Joni Amerikara joaten]
(Jon-GEN) son-DET(ABS) force AUX-3sA-3pE JON-DAT America-ALAT going
'(Jon's) son has been forced to go to America (and this happened) to Jon.'

The subject has no control over the possession relation, and can not control the action denoted in (46b) and (47b). Nevertheless, there are verbs which affect just this possession relation and thus are able to control the event. In example (48a), the object is affected precisely because there is a change affecting the possessor:

(48) a. Jonek papera kendu dio Mireni.
Jon-ERG paper-DET(ABS) take away AUX-3sA-3sD-3sE Miren-DAT
'Jon took Miren's paper away from her.'

b. Jon behartu dute [PRO Mireni papera kentzen.]
Jon(ABS) force AUX-3sA-3pE Miren-DAT paper-DET(ABS) taking away
'Jon has been forced to take Miren's paper away from her.'

Natural self-initiated processes can not be controlled. This behavior can be tested with verbs such as *esleitu* 'assign' and *egotzi* 'attribute' mentioned for examples (4) an (5) in the introduction. Both verbs express voluntary actions and both require an animate subject. However *esleitu* 'assign' entails control over the oblique object, which becomes the beneficiary or possessor of the assigned object or characteristic, while the verb *egotzi* 'attribute' is an attitude verb which can not control the subject of a self-initiated process.

(49) a. Enpresari eskolaren zaharberritzea esleitu diote.
company-DET-DAT school-GEN remodeling-DET(ABS) assign AUX
'The remodeling of the school has been assigned to the company.'

b. Enpresari X izena esleitu diote.
Company-DET-DAT X name-DET(ABS) assign AUX
'The company has been assigned the name X.'

In (49b) the company becomes the possessor of the name, but this assignation does not entail any kind of activity for the dative object. In (49a), however, the oblique argument acquires some power over the theme by receiving it. This argument is considered able to carry out the action of remodeling the school. The action assigned to the oblique object can not be a natural process, because the subjects of this kind of events can not be controlled.

(50) a. Aurkitu duten substantziari X izena esleitu diote.
find AUX-REL substance-DET X name assign AUX-3sA-3sD-3pE
'The substance found has been assigned the name X.'

b. *Substantziari zahartzearen azkartzea esleitu diote.
Substance-DAT ageing-GEN acceleration-DET assign AUX-3sA-3sD-3pE
'The substance has been assigned the acceleration of ageing.'

By contrast, the verb *egotzi* 'attribute' is perfectly allowable in a sentence parallel to (49b) (see 50a). The attribution of a property is a voluntary action, but this action does not affect the object, because *egotzi* 'attribute' is an attitude verb (51).

(51) Substantziari zahartzearen azkartzea egotzi diote.
Substance-DAT ageing-GEN acceleration-DET attribute AUX-3sA-3sD-3pE
'They attributed to the substance the acceleration of ageing.'

7. The semantics of causation and derivation of deverbal words

In this section we try to show that the semantic features characterized in previous sections of this paper are relevant in the speaker's competence when deverbal adjectives and nouns are derived. Derivative affixes appear to compete in order to attach to verbs expressing different kinds of events. They also seem to compete in generating different types of deverbal words.

We compare the behavior of the following suffixes: *-kor, -garri, -gailu, -gaitz* and *-tzaile*. First we will look at the grammatical category of the derived words. Our first examples, *-kor* and *-gaitz*, only produce adjectives. The affixes *-gailu* and *-tzaile* produce nouns, although deverbal nouns with *-tzaile* are easily used as predicates or modifiers of a noun. In some cases, dictionaries attribute to words derived with the affix *-tzaile* the category adjective and thus we will consider them adjectives. Finally, the suffix *-garri* generates both nouns and adjectives. Concerning the valency of the verbs selected by these morphemes, the affix *-kor* selects monadic verbs or monadic instantiations of verbs with different kinds of valency (Oyharçabal 2001). The suffix *-gaitz* allows both monadic and dyadic verb instantiations. Finally *-garri, -gailu* and *-tzaile* require dyadic verbs or dyadic instantiations of verbal entries with more than one possible valency. Triadic verb instantiations are avoided with all these morphemes.

Morpheme	Valency	Nouns	Adjectives
-kor	1	–	+ *hauskor* 'fragile'
-garri	2 2	+ *lokarri* 'string, bond, cord'	+ *ulergarri* 'understandable' *hausgarri* 'breaking'
-gaitz	1/2	–	+ *ulergaitz* 'hard to understand'
-gailu	2	+ *lokailu* 'string, bond, cord'	–
-tzaile	2	+ *hiltzaile* 'killer'	+ *apurtzaile* 'breaking'

The affixes above seem to distinguish between different instantiations of verbs with different kinds of alternations (Oyharçabal 1996). Pustejovsky (1995) distinguishes two kinds of alternations: on the one hand, there are verbal alternations involving true arguments such as inchoative/causative alternations; on the other there are alternations involving an optional phrase (default arguments), such as material/product alternations, in which the expression of the material is optional.

With respect to inchoative/causative alternations, curiously enough the affix *-kor* necessarily gives us the inchoative interpretation, whereas any other affix in the list is necessarily related to the causative variant of the verb.

(52) a. Loreontzia *hautsi* da.
flower-vase-DET-AUX break AUX-3sA
'The flower-vase broke.'

b. Berriak Jonen bihotza *hautsi* du.
piece.of.news-DET-ERG Jon-GEN heart-DET(ABS) break AUX-3sA-3sE
'The news broke Jon's heart.'

The examples in (52) show the inchoative/causative alternation of the verb *hautsi* 'break'. The affix *-kor* gives us the inchoative interpretation of the verb (53a), while with the affix *-garri,* the adjective seems to have been derived from the causative variant of the verb. Derived adjectives with the affix *-kor* have been traditionally paraphrased as 'that has the tendency to' (Azkue 1923-25, Villasante 1974). In this sense, *hauskor* 'fragile' is the property of something that has the tendency to break. By contrast, the adjective *hausgarri* 'breaking' means the property of being the cause for something to break.

(53) a. Loreontzia *hauskorra* da.
flower-vase-DET(ABS) fragile is
'The flower-vase is fragile.'

b. Berri hori bihotz-*hausgarria* da
piece of news that(ABS) heart-breaking is
'That is heart-breaking news.'

With psychological predicates of the *preocupare*-class we find the same causative alternances and the same distribution of suffixes (54): *-kor* goes with the unaccusative variant and *-garri* with the causative variant.

(54) a. Haurra erraz beldurtzen da.
child-DET(ABS) easily get.frightened AUX-3sA
'The child easily gets frightened.'

→ haur *beldurkorra*
child fearful-DET
'fearful child'

b. Filmak haurra beldurtu du.
movie-DET-ERG child-DET(ABS) frightened AUX-3sA-3sE
'The movie frightened the child.'

→ film *beldurgarria*
movie frightening-DET
'frightening movie'

We also find the affix *-garri* with obligatory transitive verbs, but in this case, the adjective seems to absorb the theme argument (55). With psychological predicates of the *temere*-class such as *miretsi* 'admire', the affix *-garri* absorbs also the theme argument (55b).

(55) a. Ideia hau erraz *uler* daiteke. → ideia *ulergarria*
idea this (ABS) easily understand can idea understandable-DET
'This idea is easy to understand.' 'understandable idea'

b. Jonek zure lana *miresten* du. → lan *miresgarria*
Jon your work admire AUX-3sA-3sE work admirable-DET
'Jon admires your work.' 'admirable work'

The existence of two types of derived adjectives (54b/55b) with the affix *-garri* has usually been interpreted as the result of two *-garri* affixes: a passive *-garri* and an active *-garri* (see for example Azkarate 1990 and Azkarate & Gràcia 1995). Artiagoitia (1995), however, claims that there is a sole suffix *-garri*, which always externalizes an internal argument, and that consequently there is no active affix *-garri*. This Basque linguist, following Belleti & Rizzi (1988) relates the so called active *-garri* either with the theme argument of psych-predicates of the *preocupare*-class (*beldurgarri* 'frightening' in 54b) or with instrumentals (*apaingarri* 'decorative, ornamental'). As for instrumentals, Artiagoitia (1995) claims that they also should be internal subjects. This view is congruent with the analysis proposed for instrumentals in section 2 of this paper. Nevertheless, because

we accept the analysis of psych-predicates in section 4 of this paper, we agree with the traditional analysis, which maintains that there are two *-garri* affixes.

The affix *-kor* is incompatible with obligatory transitive verbs, but it is found with obligatory intransitives of either the unaccusative or the unergative class. We also find this affix with a small set of transitive verbs that allow for the object to be interpreted as generic. Finally we find the affix *-kor* with experiencers of the two classes of psych-predicates (56d, e).

(56) a. egoera *egonkor* [*egon* 'be with stage level predicates' UNACCUSATIVE]
situation stable
'stable situation'

b. beira *distirakor* [*distiratu* 'shine' UNERGATIVE]
glass shiny '
shiny glass'

c. lur *emankor* [*eman* 'give' TRANSITIVE]
land productive
'productive land'

d. haur *beldurkor*
child fearful
'fearful child'

e. gizon *mireskor*
man full.of.admiration

When *-kor* is attached to a transitive verb, the affix absorbs the external argument, and the internal argument is blocked. Compare the synthetic compound in (53b) with the impossibility to generate synthetic compounds with the affix *-kor* (57b).

(57) a. Lur honek patata onak ematen ditu.
land this-ERG potato good(ABSpl) give AUX-3pA-3sE
'This land produces good potatoes.'

b. *lur patata-*emankor*
land potato-productive

Our first conclusion is that the adjectival affix *-kor* only absorbs internal cause arguments. This is perfectly consequent with its semantics and with the fact that it only derives adjectives. Remember that for an argument to be an internal cause, the DP which saturates this cause must bear an intrinsic property that makes it able by itself to initiate the event denoted by the verb.

Turning to the suffix *-garri*, we agree with Azkarate (1990), Azkarate & Gràcia (1995) and Gràcia et al. (2000) in their distinguishing of two different suffixes. However, we do agree with Artiagoitia (1995) in one respect. The characterization of the affix *-garri* in (53b) as active is not the best choice, since the argument absorbed by this affix is always interpreted as a circumstantial cause or as an instrument. That is, non controller arguments are required, and animate entities are avoided as subjects of deverbal adjectives with *-garri* (Artiagoitia 1995). Note that the adjective *hilgarri* 'deadly' is incompatible with the human noun Jon, which requires the affix *-tzaile* (*hiltzaile* 'killer'). The label 'causative *-garri*' would be more appropriate for this suffix.

(58) a. Jonek/ substantzia horrek norbait *hil* du.
Jon-ERG/substance that-ERG someone kill AUX-3sA-3sE
'Jon/that substance killed someone.'

b. Jon *hiltzailea*/ **hilgarria* da.
Jon killer-DET/*deadly-DET is
'Jon is a killer.'

e. Substantzia hori **hiltzailea*/*hilgarria* da.
substance that(ABS) **killer*/ *deadly* is
'That is a deadly substance.'

When the affix *-garri* is attached to a verb that necessarily expresses controlled events, the derived adjective or noun is interpreted as an instrument. What is more, derived nouns with *-garri* are always interpreted as instruments. Note that *soken bidez* 'with strings in (59a) is an optional phrase and so a default argument, following Pustejovsky (1995). The presence/absence of the instrument is not related to a causative/inchoative alternance. The instrument requires a causative construction with a controller subject. 'Instrumental *-garri*' would be an appropriate label for this *-garri*. Furthermore, this use of the affix *-garri* overlaps only with that of the suffix *-gailu*, and we often find synonymous derived nouns with both affixes (59b).

(59) a. Jonek zapatak *lotu* ditu (soken bidez).
Jon-ERG shoe(ABSpl) tie AUX-3plA-3sE (string-GEN by.means.of)
'Jon tied his shoes (with strings).'

b. Zapaten *lokarriak*/*lokailuak*
shoe-GENpl tying-DETpl
'Laces'

As for adjectives derived with the affix *-gaitz*, we find antonyms of some adjectives derived with the affix *-kor* and some antonyms of adjectives with *-garri*. However, the affix *-gaitz* always absorbs either the θ-role of the internal argument of transitive verbs, or the only argument of unaccusative verbs. This affix is incompatible with unergatives and with absolute interpretations of transitive verbs.

(60) a. loreontzi *hauskor*/loreontzi *hauskaitz*
flower-vase fragile/ flower-vase *unbreakable*

b. ideia *ulergarri*/ ideia *ulergaitz*
idea *understandable*/idea hard.to.understand

c. substantzia *hilgarri*/substantzia **hilgaitz*
substance *deadly*/ substance **undeadly*

f. egoera *egonkor*/begi *egongaitz*
state *stable*/ eye *restless*

g. beira *distirakor*/beira **distiragaitz*
glass *shiny*/ glass *unshiny

h. lur *emankor*/ lur **emangaitz*
land *productive*/land unproductive

Finally, the suffix *-tzaile* is mostly attached to transitive verbs. Nevertheless, two exceptions are often mentioned in the literature (*egoile* 'inhabitant' and *joaile* 'emigrant').

In literary tradition, this affix appears mostly attached to verbs with human subjects, and the derived nouns often express types of trades (*idazle* 'writer', *epaile* 'judge', *aurkezle* 'presenter', *saltzaile* 'seller'). Other derived nouns express the subject of an event (*erosle* 'buyer', *igorle* 'sender', *hartzaile* 'receiver', *ikusle* 'viewer', *hiltzaile* 'killer', *jasale* 'sufferer, patient'). New derived words with *-tzaile* often refer to inanimate entities, and can often be paraphrased as 'something which intrinsically does __'. Therefore this suffix always absorbs the external argument which must be human or something characterized by doing the action denoted by the verb

(61) a. elektroi-*hartzaile*
electron-receptor

b. bizidun *fotosintetizatzaile*
organism photosynthesiser

c. espezie *adierazle*
species indicator

e. *disolbatzaile*
'solvent'

f. *kutsatzaile*
'pollutant'

The following table summarizes the behavior of the suffixes described in this section.

Affix	Verb-class	Argument absorbed	Examples
-kor	UNACCUSATIVES PSYCH-UNACCUSAT.	INTERNAL CAUSE INTERNAL CAUSE (EXPERIENCER)	*hauskor* 'fragile' *beldurkor* 'fearful'
	UNERGATIVES TRANSITIVES (the internal argument is blocked) PSYCH-TRANSIT.	INTERNAL CAUSE INTERNAL CAUSE INTERNAL CAUSE	*distirakor* 'shiny' *emankor* 'productive' *mireskor* 'full.of.admiration'
-garri	TRANSITIVES PSYCH-TRANSIT. (EXP.-THEME) (passive *-garri*)	THEME THEME	*ulergarri* 'understandable' *miresgarri* 'admirable'
	TRANSITIVES PSYCH-TRANSIT. (CAUS.-THEME) (causative *-garri*)	CAUSE CAUSE	*hausgarri* 'breaking' *beldurgarri* 'frightening'
	TRANSITIVES with instrument subjects (instrumental *-garri*)	INSTRUMENT	*lokarri* 'string'
-gailu	TRANSITIVES with instrument subjects	INSTRUMENT	*lokailu* 'string'
-gaitz	UNACCUSATIVES TRANSITIVES	INTERNAL CAUSE THEME	*egongaitz* ' restless' *hauskaitz* 'unbreakable' *ulergaitz* 'hard to understand'
-tzaile	TRANSITIVES	HUMAN SUBJECTS INANIMATE SUBJECTS CHARACTERIZED BY DOING THE ACTION DENOTED BY THE VERB	*hiltzaile* 'killer' *disolbatzaile* 'solvent'

References

Alcázar, A., 2002, *Aspectual interpretation in Basque.* MA Research paper, USC.

Artiagoitia, X., 1995, "*Garri* atzizkiaren izaera bikoitzaz: Zergatik den maitagarria bezain mingarria." *ASJU* 29, 2-3. 355-405.

Authier, M. & Reed, L., 1991, "Ergative Predicates and Dative Clitization in French Causatives", *Linguistic Inquiry* 22: 1, 197-204.

Azkarate, M., 1990, "Euskal *–garri* vs. latinezko *–bilis*, bi balio desberdin" in *Patxi Altunari Omenaldia*, Donostia. Mundaiz: 19-39.

— & Gràcia, Ll., 1995, "Agentivity and Modality in Deverbal Adjectives in Basque and in Catalan", *Rivista di Grammatica Generativa*: 39-50.

Azkue, R.M.ª, 1923-25, *Morfología vasca.* La Gran Enciclopedia Vasca. Bilbao.

Belleti, A. & Rizzi, L., 1988, "Psych Verbs and Theta Marking", *Natural Languages and Linguistic Theory* 6, 291-352.

Belvin, R., 1997, "The Semantic Hierarchy, Semantic Control and Eventivity in Nisgha" in Mendikoetxea & Uribe-Etxebarria (eds.), 35-54.

Berman, A., 1970, "Agent, Experiencer, and Contollability". Report Nº NSF-24. *Mathematical Lingusitics and Automatic Translation.* The Computation Laboratory of Harvard.

Chierchia, G., 1989, "Structured Meanings, Thematic Roles, An Control" in G. Chierchia, B. Partee & R. Turner (eds.) *Properties, Types, and Meaning.* Kluwer Academic Publishers. Dordrecht. The Netherlands: II, 131-166.

Copley, B., 2000, "On the Syntax and Semantics of Futurate Readings", *MIT Working Papers in Linguistics* 36 MITWPL. Cambridge.

Demirdache, H., 1997, "Out of Control' in Salish and Event (De)Composition" in Mendikoetxea & Uribe-Etxebarria (eds.), 97-144.

Etxepare, R., 2003, "Valency and argument structure in Basque verb" in J.I. Hualde & J. Ortiz de Urbina (eds.), *A grammar of Basque.* Mouton de Gruyter. Berlin.

Giorgi, A. & Pianesi, F., 1997, *Tense and Aspect.* Oxford University Press. New York.

Gràcia *et al.*, 2000, *Configuración morfológica y estructura argumental: léxico y diccionario.* The University of the Basque Country Press, Bilbao.

Grimshaw, J., 1990, *Argument Structure.* The MIT Press. Cambridge MA.

Jackendoff, R., 1993, "The Conceptual Structure of Intending and Volitional Action" in H. Campos & P. Kempchimsky (eds.), *Festschrift for Carlos Otero.* Georgetown University Press. Georgetown.

Kural, M., 1997, "Verb Incorporation and Causation Types" in Mendikoetxea & Uribe-Etxebarria (eds.), 231-254.

Laka, I. 1995, *A brief grammar of Euskara, the Basque language.* www.ehu.es/grammar

Lasnik, H., 1991, "Two Notes on Control and Binding" in Larson, R. et al. *Control and Grammar.* Kluwer Academic Press, Dordrecht: 235-251.

— & Fiengo, R., 1974, "Complement Object Delection", *Linguistic Inquiry* 5.4.

Levin, B., 1983, *On the Nature of the Ergativity.* Ph D. MIT.

— & Rappaport Hovav, M., 1995, *Unaccusatives: At the Syntax-Lexical Semantics Interface.* MIT Press. Cambridge. MA.

Mendikoetxea, A., 1999, "Construcciones inacusativas y pasivas" in I. Bosque & V. Demonte (eds.), *Gramática Descriptiva de la Lengua Española.* Madrid: II, 1576-1627.

— & M. Uribe-Etxebarria, (eds.), 1997, *Theoretical Issues at the Morphology-Syntax Interface.* Supplements of *ASJU*, San Sebastián.

Minkoff, S.A., 1997, "Argument Structure and Animacy Entailment" in Mendikoetxea & Uribe-Etxebarria (eds.), 285-324.

Odriozola, J.C. & Zabala, I., 2000, "Un diccionario bilingüe de verbos dirigido a usuarios del vasco técnico". In *Proceedings 1st International Conference on Specialized Translation. Universitat Pompeu Fabra. Barcelona. March 2000:* 175-178.

Oyharçabal, B., 1996, "Aditzen argumentu egituraz". *Berba* 20.
—, 2003, "*-Kor* atzizkiaz". In J.M. Makatzaga and B. Oyharçabal (eds.), *Euskal gramatikari eta literaturari buruzko ikerketak XXI. mendearen atarian. Gramatika gaiak,* Iker-14 (I), Euskaltzaindia, Bilbao, 357-383.
Pesetsky, D., 1987, "Binding Problems with Experiencer Verbs", *Linguistic Inquiry* 18.1, 126-140.
—, 1995, *Zero Syntax.* Cambridge MA: MIT Press.
Pustejovsky, J., 1991, "The Syntax of Event Structure", *Cognition* 41, 41-87.
—, 1995, *The Generative Lexicon.* Cambridge MA: MIT Press.
Shin, J., 1998, *Morfología y modularidad: la formación de verbos causativos en español.* Ph.D. Universidad Autónoma de Madrid.
Tenny, C., 1988, "The Aspectual Interface Hypothesis: the Connection between Syntax and lexical Semantics" in C. Tenny (ed.), *Studies in Generative Approaches to Aspect.* MIT Working Papers in Linguistics: 1-18.
—, 1989, *The Aspectual Interface Hypothesis* MIT Working Papers in Linguistics.
—, 1994, *Aspectual roles and the Syntax-Semantics Interface,* Kluwer Academic Publishers. Dordrecht.
Villasante, L., 1974, *Palabras vascas compuestas y derivadas.* Oñati.
Zabala, I., 1993, *Predikazioaren teoriak gramatika sortzailean (euskararen kasua).* Ph.D. University of the Basque Country.
—, 1995, "Aditzen hautapena euskara teknikoan", *Ekaia* 3, 123-134.

SUBJECT INDEX

CONTRIBUTORS

ADURIZ, Itziar
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
jiradagi@si.ehu.es

ALBERDI LARIZGOITIA, Xabier
Euskal Filologia Saila
Euskara Zerbitzua (Errektoregoa)
University of the Basque Country (EHU-UPV)
Sarriena auzoa
48940 Leioa
fvpallaj@lg.ehu.es

ALDEZABAL ROTETA, Izaskun
Departamento de Lingüística General
Universitat de Barcelona
Gran Via de las Corts Catalans, 585.
08007 Barcelona
jibalroi@si.ehu.es

ARANZABE URRUZOLA, Maria Jesus
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
maxux@si.ehu.es

ARRIOLA EGUROLA, Jose Mari
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
jmarriola@sf.ehu.es

ARTIAGOITIA, Xabier
Hizkuntzalaritza eta Euskal Ikasketak
Filologia eta Geografi-Historia Fakultatea
University of the Basque Country (EHU-UPV)
Vitoria-Gasteiz
fvparbex@vc.ehu.es

ARTOLA, Xabier
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
jiparzux@si.ehu.es

DIAZ DE ILARRAZA, Arantza
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
jipdisaa@si.ehu.es

GOENAGA, Patxi
Hizkuntzalaritza eta Euskal Ikasketak
Filologia eta Geografi-Historia Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Vitoria-Gasteiz
jipgogak@si.ehu.es

GOJENOLA GALLETEBEITIA, Koldo
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
jipgogak@si.ehu.es

ODRIOZOLA PEREIRA, Juan Carlos
Euskal Filologia Saila
Zientzia Fakultatea
University of the Basque Country (EHU-UPV)
644 PK.
48080 Bilbo
fvpodpej@lg.ehu.es

OYHARÇABAL, Bernard
IKER (CNRS)
28, Lormand
F-64100 Baiona / Bayonne
bernard.oyharcabal@univ-pau.fr

SARASOLA, Kepa
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
jipsagak@si.ehu.es

SOROA, Aitor
Lengoaia eta Sistema Informatikoak
Informatika Fakultatea
University of the Basque Country (EHU-UPV)
649 P.K.
20080 Donostia / San Sebastián
ccpsoeta@si.ehu.es

ZABALA UNZALU, Igone
Euskal Filologia Saila
Zientzia Fakultatea
University of the Basque Country (EHU-UPV)
644PK.
48080 Bilbo
fvpzauni@lg.ehu.es

XXV. JOSÉ Mª SÁNCHEZ CARRIÓN, *Un futuro para nuestro pasado,* 1991. 15 €.

XXVI. MANUEL AGUD - † ANTONIO TOVAR, *Diccionario etimológico vasco,* IV. *Egiluma-Galanga,* 1991. Out of print.

XXVII. JOSEBA A. LAKARRA - JON ORTIZ DE URBINA (eds.), *Syntactic theory and Basque syntax,* 1992. 18 €.

XXVIII. RICARDO GÓMEZ - JOSEBA A. LAKARRA (arg.), *Euskal dialektologiako kongresua (Donostia, 1991-ko irailaren 2-6),* 1994. 21 €.

XXIX. JOSÉ I. HUALDE - XABIER BILBAO, *A Phonological Study of the Basque Dialect of Getxo,* 1992. 8 €.

XXX. MANUEL AGUD - † ANTONIO TOVAR, *Diccionario etimológico vasco,* V. *Galani-Iloza,* 1992. 8 €.

XXXI. KARLOS OTEGI, *Lizardi: lectura semiótica de* Biotz-begietan, 1993. 18 €.

XXXII. AURELIA ARKOTXA, *Imaginaire et poésie dans* Maldan behera *de Gabriel Aresti (1933-1975),* 1993. 10 €.

XXXIII. MANUEL AGUD - † ANTONIO TOVAR, *Diccionario etimológico vasco,* VI. *Ilpiztu-Korotz,* 1993. 8 €.

XXXIV. JOSÉ I. HUALDE - GORKA ELORDIETA - ARANTZAZU ELORDIETA, *The Basque dialect of Lekeitio,* 1994. 18 €.

XXXV. GEORGES REBUSCHI, *Essais de linguistique basque,* 1997. 18 €.

XXXVI. XABIER ARTIAGOITIA, *Verbal projections in Basque and minimal structure,* 1994. 12 €.

XXXVII. MANUEL AGUD - † ANTONIO TOVAR, *Diccionario etimológico vasco,* VII. *Korpa-Orloi,* 1994. 8 €.

XXXVIII. PATXI GOENAGA (ed.). *De grammatica generativa,* 1995. 18 €.

XXXIX. ANTONIO CID, *Romancero y balada oral vasca (literatura, historia, significado).* In preparation.

XL. AMAYA MENDIKOETXEA - MYRIAM URIBEETXEBARRIA (eds.), *Theoretical issues at the morphology-syntax interface,* 1997. 21 €.

XLI. BERNHARD HURCH - Mª JOSÉ KEREJETA, *Hugo Schuchardt - Julio de Urquijo: Correspondencia (1906-1927),* 1997. 21 €.

XLII. JOSE I. HUALDE, *Euskararen azentuerak,* 1997. 15 €.

XLIII. RUDOLF P. G. DE RIJK, *De lingua vasconum: selected writings,* 1998. 15 €.

XLIV. XABIER ARTIAGOITIA, PATXI GOENAGA, JOSEBA A. LAKARRA, *Erramu Boneta:* Festschrift for Rudolf P. G. de Rijk, 2002. 30 €.

XLV. JOSEBA A. LAKARRA, *Ikerketak euskararen historiaz eta euskal filologiaz.* In press.

XLVI. BERNARD OYHARÇABAL, *Inquiries into the lexicon-syntax relations in Basque,* 2003.