

ANUARIO DEL SEMINARIO DE FILOLOGÍA
VASCA «JULIO DE URQUIJO»

International Journal of Basque Linguistics and Philology

XLI-2

2007



Gipuzkoako Foru Aldundia
Diputación Foral de Gipuzkoa



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Anuario del Seminario de Filología Vasca “Julio de Urquijo”
International Journal of Basque Linguistics and Philology
(ASJU)

Sortzaileak / Fundadores / Founded by
Manuel Agud (†) - Luis Michelena (†)

Zuzendariak / Directores / Directors
Joseba A. Lakarra (JUMI, UPV/EHU) - Ibon Sarasola (JUMI, UPV/EHU)

Argitaratzailea / Editor
Joseba A. Lakarra (JUMI, UPV/EHU)

Idazkaritza / Redacción / Board

Gidór Bilbao (JUMI, UPV/EHU)
Iñaki Camino (JUMI, UPV/EHU)
Ricardo Gómez (JUMI, UPV/EHU)
Joaquín Gorrochategui (JUMI, UPV/EHU)
Iván Igartua (JUMI, UPV/EHU)

Miren Lourdes Oñederra (JUMI, UPV/EHU)
Javier Ormazabal (JUMI, UPV/EHU)
Iñigo Ruiz Arzalluz (JUMI, UPV/EHU)
Blanca Urgell (JUMI, UPV/EHU)
Koldo Zuazo (JUMI, UPV/EHU)

Idazkaritza arduradunak / Coordinadores / Editor's Adjunts
Iñigo Ruiz Arzalluz (JUMI, UPV/EHU) / Blanca Urgell (JUMI, UPV/EHU)

Aholku Batzordea / Consejo Asesor / Advisory Board

Jacques Allières (Toulouse) (†)
Patxi Altuna (Deustu, Donostia) (†)
Jesús Antonio Cid (Complutense)
Lyle Campbell (Utah)
Joan Coromines (Barcelona) (†)
María Teresa Echenique (Valencia)
Ricardo Etxepare (Iker, CNRS)
Jon Franco (Deustu, Bilbo)
Jean Haritschelhar (Bordele III)
José Ignacio Hualde (Urbana, Illinois)
Bernard Hurch (Graz)
Jon Juaristi (Alcalá)
Itziar Laka (UPV/EHU)

Jesús María Lasagabaster (Deustu, Donostia)
Francisco Oroz Arizcuren (Tübingen)
Jon Ortiz de Urbina (Deustu, Bilbo)
Beñat Oyharçabal (Iker, CNRS)
José Antonio Pascual (Carlos III)
Georges Rebuschi (Paris III)
Rudolf P. G. de Rijk (Leiden) (†)
Patxi Salaberri (UPNA)
Iñaki Segurola (Orotariko Euskal Hiztegia)
Larry Trask (Sussex) (†)
Juan Uriagereka (Maryland)
Myriam Uribe-Etxebarria (UPV/EHU)
Charles Videgain (Iker / UPPA)

ASJU 1954an sortutako Euskal Linguistika eta Filologiako nazioarteko aldizkaria da, “Julio Urkixo” Euskal Filologi Mintegi-Institutuak (JUMI) argitaratua, eta iker-eremu horietatik edo horietarako ere interesgarri izan daitezkeenatik goi mailako artikulu, ohar eta liburu-iruzkinak argitaratzen ditu. Urtero bi zenbaki ateratzen dira (orotara 750 bat orrialde). ASJU-k badu, orobat, *GEHIGARRI* sail bat, non artikulu formatuaz gorako lanak argitaratzen diren, epe jakinik gabe (ikus zerrenda zenbakiaren amaieran).

Originaleri buruzko harremanetarako ikus bitez bukaerako EGILEENTZAKO OHARRAK. Harpidetza eta eskarietarako: Euskal Herriko Unibertsitatea, Argitalpen Zerbitzua, Leioa (luxedito@lg.ehu.es).

ASJU es una revista internacional de Lingüística y Filología Vasca fundada en 1954 y publicada por el Instituto-Seminario de Filología Vasca “Julio Urquijo” (JUMI). Se publican en ella artículos, notas y reseñas sobre los campos mencionados y afines, y otros de interés para los mismos. En la actualidad es de periodicidad semestral (unas 750 páginas). Sin regularidad preestablecida, ASJU publica en sus ANEJOS trabajos de formato superior al de un artículo (véase la lista al final del volumen).

Para correspondencia relacionada con los originales véase la INFORMACIÓN PARA LOS AUTORES al final del número. Para suscripciones y pedidos: Servicio de Publicaciones, Universidad del País Vasco (UPV/EHU), Leioa (luxedito@lg.ehu.es).

ASJU is an International Journal of Basque Linguistics and Philology founded in 1954, and published by the “Julio de Urquijo Seminar of Basque Philology” Institute (JUMI). It publishes highquality papers, notes, squibs and reviews about the above mentioned and other related topics in two issues per year (up to a total of 750 pages). Longer works are published as SUPPLEMENTS to the regular issues of the ASJU (see the list at the end of the issue).

For correspondence about papers see the INFORMATION FOR AUTHORS at the end of this issue. For subscriptions and order: Servicio de Publicaciones, Universidad del País Vasco (UPV/EHU), Leioa (luxedito@lg.ehu.es).

ANUARIO DEL SEMINARIO DE FILOLOGÍA
VASCA «JULIO DE URQUIJO»

International Journal of Basque Linguistics and Philology

XLI-2

2007



Gipuzkoako Foru Aldundia
Diputación Foral de Gipuzkoa



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

© “Julio Urkixo” Euskal Filologia Mintegia / Seminario de Filología Vasca «Julio de Urquijo»

ISSN.: 0582-6152. Depósito legal / Lege gordailua: S.S.-400/1967
Fotocomposición / Fotokonposizioa: Ipar, S. Coop. - Zurbarán, 2-4 (48007 Bilbao)
Impresión / Inprimatzea: Itxaropena, S.A. - Araba kalea, 45 (20800 Zarautz-Gipuzkoa)

PREFACE

BIDE 2005 would not have taken place without the help and support of many colleagues. We would like to begin this special volume by acknowledging them all and hoping that they will continue to support our conference.

First, we want to thank the scientific committee for their help in the review process. These are Asier Alcázar (University of Southern California), Ellen Broselow (Stony Brook University), Mónica Cabrera (Loyola Marymount University), Alvaro Cerón-Palomino (University of Southern California), Urtzi Etxebarria (University of the Basque Country/HiTT), Ricardo Etxepare (CNRS/HiTT), Dan Finer (Stony Brook University), Leyre Goitia (University of Deusto), Carolina González (Florida State University), Rodrigo Gutiérrez Bravo (CIESAS-Mexico City), Nina Kazanina (University of Maryland), Abe Kazemzadeh (University of Southern California), Heejeong Ko (Stony Brook University), Alazne Landa (University of the Basque Country/HiTT), Thomas Leu (New York University), Luisa Martí (University of Tromsø), Franc Marušič (University of Nova Gorica), Elixabete Murguía (University of Deusto-Bilbao), Iván Ortega-Santos (University of Maryland), Leticia Pablos (University of Reading), Lara Reglero (Florida State University), Itziar San Martín (University of the Basque Country/HiTT), Carmen Silva-Corvalán (University of Southern California), Luis Vicente (University of Leiden) and Masaya Yoshida (University of Maryland).

For financial and logistic support in hosting the conference, we are thankful to the School of Philosophy and Letters and the CIDE Program at the University of Deusto, HiTT, the Basque Government, *Bilbao Iniciativa Turística*, and the Getxo Tourism Office.

We want to express a special gratitude to Joseba Lakarra from ASJU Press and his team for making possible the publication of this volume and leading us through the publishing procedure.

Finally, our most sincere and warm thanks go to Jon Franco and Jon Ortiz de Urbina from the University of Deusto. They are the driving force behind BIDE and have been vital in the organization of this conference. We are grateful for having them as part of the BIDE team.

Irene Barbería
Rebeka Campos-Astorkiza
Susana Huidobro

INTRODUCTION

Irene Barbería

University of Deusto

Rebeka Campos-Astorkiza

The Ohio State University

Susana Huidobro

Stony Brook University / University of the Basque Country

This volume contains the proceedings of BIDE 2005, the second International Conference of Students in Linguistics, held in June 2005 at the University of Deusto, Bilbao. Its first edition, BIDE 2004, received a warm welcome from the international linguistic community and this paved the way for a second edition under the same basic premise: bring together young scientists to the University of Deusto, a Basque center of linguistic research, thus boosting research and cooperation. In the same spirit as BIDE 2004, BIDE 2005 was organized and run by a group of former Deusto students, who are currently pursuing their linguistic careers in different international institutions.

It is not accidental that this International Conference of Students in Linguistics takes place at the University of Deusto in Bilbao, since this University plays an instrumental role in the formation of linguists. There is a long tradition of Deusto faculty members encouraging and helping their students to pursue graduate degrees in linguistics at some of the best universities. This tradition is best illustrated by the plenary speaker of BIDE 2005, Ricardo Etxepare. Professor Etxepare obtained his doctorate from the University of Maryland, under the supervision of a former Deusto student, Professor Juan Uriagereka.

Ricardo Etxepare's excellence in research is proven by his current position at the 'Centre National de la Recherche Scientifique (CNRS)', where he holds the title of *permanent main researcher*, since October 2000. Etxepare belongs to a handful of linguists whose interests and research focus mainly on two languages, namely Spanish and Basque, but also on some other Romance languages such as French. It is relevant to note here that Rikardo Etxepare has contributed to the recently published 'A Grammar of Basque' (2003, Hualde and Ortiz de Urbina (eds.)) with several chapters on syntax. This grammar has become an influential publication, filling a gap in the field. Consequently, Etxepare's contribution places him as one of the few experts in the area of Basque linguistics worldwide. Furthermore, his broad and exceptional work in Basque is comparable to his great and varied work on Hispanic Linguistics. Professor Etxepare is one of the most active and productive scholars both in Hispanic and Basque linguistics and his publications can be found in some of the most prestigious journals such as *Probus* and the *International Journal of Basque Linguistics (ASJU)*.

Professor Etxepare belongs to the second generation of Basque linguists that succeeded in getting their Ph.D.s from an overseas university. They were motivated, on the one hand, by professors from the University of Deusto and on the other hand, by the experiences of previous students from the Basque Country, who successfully finished their studies abroad. Etxepare is a very valuable member of the linguistic community in the Basque Country, and he has been a keystone in the formation of new students not only from the Basque Country but also from France and Spain. His regular one-semester courses at the University of the Basque Country have always inspired students due to his excellence in teaching, his dedication to the students and his enthusiasm for linguistics. Undoubtedly, Rikardo Etxepare is responsible for the new generation of promising linguists in the Basque Country. Moreover, he is a constant reference for students from different parts of the world, and this is reflected in the fact that he is a member of several dissertation committees inside and outside the Basque Country.

The presence of Rikardo Etxepare in BIDE 2005 has added to the conference's richness by allowing participants not only to learn about Etxepare's latest research but also to discuss their own research with him. In fact, BIDE has emerged as an important forum for students in linguistics, where they can meet scholars with similar interests and get feedback from their peers and professors. Several BIDE attendees have been able to develop some of the new ideas obtained at the conference into papers and dissertation chapters.

BIDE offers a unique opportunity in Spain for international students with an interest in generative linguistics. The linguistic tradition in the Basque Country has greatly benefited from this line of research. However, there was a need to create an international conference within this community, where students could share their work. BIDE has filled this void and has attracted not only generativists but also researchers from other frameworks. For this reason, BIDE plays a pivotal role in helping build ties between the Basque linguistic community and researchers in other parts of the world, and also in creating a bond among the different linguists in the Basque region and Spain. This is especially useful for students and young scientists who are beginning to create their circle of collaborators and connections. For example, BIDE has helped the professional research group HiTT develop and work towards its objective of organizing events for the discussion of current linguistic developments.

HiTT (Hizkuntzalaritza Teorikorako Taldea) or *the Basque Research Group of Theoretical Linguistics*, is a group of language researchers in different disciplines, among them, syntax, semantics, phonology, phonetics, pragmatics and sign language. The researchers in this group are affiliated with the University of Deusto (Bilbao), the University of the Basque Country (Vitoria-Gasteiz, Bilbao and San Sebastian), and Le Centre Nationale de la Recherche Scientifique (CNRS, France). It should be noted that valuable members of HiTT belong to the organizing committee of BIDE as it is the case of Jon Franco, Jon Ortiz de Urbina and Susana Huidobro. Similarly, Rikardo Etxepare, the invited speaker in BIDE 2005, is part of this group. These researchers have been involved in several HiTT projects, such as 'The architecture of language: Multidirectional architecture of the linguistic interfaces', 'Methodological foundations for the development of a Basque-Spanish tutor for computer-assisted teaching', and 'The structures of events: Tense and aspect and phrase structure', among others. HiTT members have also collaborated in the organization of different

events such as the 14th Conference of the Student Organization of Linguistics in Europe, the course *Giza Hizkuntzaren Natura* (Human Language Nature), and the 7th LEHIA International Workshop in Linguistics.

Papers in this volume

The articles in these volume are selected papers based on presentations given at BIDE 2005. As in the previous edition, a diverse range of linguistic subfields and topics are represented in this selection: syntax, semantics, phonology, computational linguistics and language acquisition. Here, an overview of the volume's contents is presented, together with a brief description of each paper and its relevance within current linguistic research.

The first paper is Etxepare's "Aspects of Quotative Constructions in Iberian Spanish", where the author observes that in colloquial speech, main declarative clauses in Iberian Spanish can be headed by an overt complementizer. His paper develops the idea that such structures in Spanish involve an extra speech eventuality, and that this speech eventuality is syntactically mapped into the structure of the sentence as a complex verbal predicate. This complex verbal predicate is composed of a light verb GO and an aspectual projection, which takes as complement an utterance denoting expression. This complex predicate is akin to what in other languages are called "*Quotative Verbs*", introducing direct or semi-direct speech (see Lord 1993, Güldemann 2001). Etxepare shows that this verbal predicate shares properties of Speech Act operators, in the sense of Krifka (2001), and of ordinary lexical verbs. Several types of evidence are brought to bear on issues related to these constructions, such as the semantic primitives involved in the quotative predication, and the syntactic configurations giving rise to the complex predicate.

Moving on to the phonology section of BIDE 2005, Huber's "On the interaction of velars and labials" presents data from a wide variety of languages to show a two-folded generalization. On the one hand, there is pervasive direct interaction between labials and velars to the exclusion of coronals. This interaction is exemplified through cases of changes where the elements involved are velars and labials. On the other hand, the motivation behind these phenomena is the presence of labiality in labials and the lack of any place specification in velars. Thus, the data discussed in this paper further support the view that velars lack place specification (Huber 2004), against standard views that consider coronal consonants as the placeless elements by default (Paradis and Prunet 1991). Huber's proposal that velars lack place information is developed within the formal framework of Government Phonology. One of the main contributions of Huber's paper is that his typology shows that the different phenomena where labials and velars interact are all in fact phonologically conditioned and absolutely regular, rather than unprincipled changes. On this basis, the paper sets up a new typology of the phenomena, which better captures the phonological conditions underlying them.

The area of computational linguistics is growing within the linguistics community. Several research programs take advantage of different computational tools to achieve their goals. An increasingly employed method is the use of corpora for linguistic analysis, i.e., corpus linguistics. Several papers in this volume look at different

ways in which corpora can be adapted to linguistic studies, and how these research tools can be most efficiently used. Another important area within computational linguistics is machine translation, which is explored in a paper by Gábor and Héja.

In “Consumer Corpus: towards linguistically searchable text”, Alcázar develops the possibility of using the Consumer corpus as a linguistic research tool. This corpus is built from a monthly online magazine published in Spain. The articles are originally written in Spanish and later translated to three other languages: Basque (a language isolate), Catalan and Galician (two Romance languages). The topics discussed in the magazine are related to consumers’ issues. Alcázar aims at making this corpus linguistically searchable, so that the search can be formulated in linguistic terms and at different levels of linguistic interest. The first step is sentence alignment, adopting Moore’s alignment tool (2002), which facilitates cross-linguistic comparison. The alignment allows for the possibility of comparing search results across the four languages of the corpus since the search results for a query in one language may be accompanied by its translations to the other three. Second, Alcázar has applied a part-of-speech tagging procedure to the Spanish portion of the corpus. The result is an annotated corpus that offers the advantage of searching for parallel words or constructions in four different languages.

In “Complements and Adjuncts in Machine Translation”, Gábor and Héja focus on automated syntactic analysis in relation to machine translation. The aim of the paper is to present a specific method for automatically differentiating between complements and adjuncts, with the purpose of building a Hungarian verbal argument structure database suited for machine translation. The authors examined Government and Binding theory (GB) and Lexical Functional Grammar (LFG) paying attention to their description of argument structure and their representation of surface argument structure. They find that neither the GB nor the LFG treatment of arguments and adjuncts proves satisfactory for Hungarian. The Hungarian surface order cannot be used for distinguishing complements and adjuncts. However, the rich morphological system can serve as a basis for the task at hand. Thus, rather than using configurational information, Gábor and Héja use morphology, namely case marking, as an indicator of the syntactic role. Their proposal is that not every occurrence of an NP with a case suffix is lexically subcategorized by a verb: some of them are added to the sentence by productive rules.

In “Extracting Information from Participial Structures”, Héja and Gábor aim at increasing the efficiency of a rule-based information extraction (IE) system by enhancing it with further grammatical knowledge. The NewsPro IE system was developed and tested on a corpus of short business news. In this IE system, the sentence’s event is identified through the main verb. The arguments and adjuncts of the main verb are correlated with the participants and circumstances of the event. The authors’ goal is to extract the information within NPs formed with non-finite verbs, such as participles. Thus, they propose a rule-based system to transform participle structures into sentences with a finite verb. The main challenge is to differentiate between participle structures that result in ill-formed finite sentences and those that do not. The authors’ solution is based on the fact that there is an adjective/participle homonymy in Hungarian. Consequently, those transformations that contain adjectives produce ill-formed sentences, and the structures containing participles render grammatical fi-

nite forms. Consequently, it is necessary to distinguish between adjectives and participles. Héja and Gábor show that the following three criteria are sufficient to make the right classification: It is a participle if (1) at least one of the base verb's complements is present, (2) at least one of the base verb's adjuncts is present, and (3) at least a preverb is present.

Two papers explore different aspects in first and second language acquisition, supporting their conclusions with experimental data. Huarte studies syntactic competence in first language learners, while Yanguas argues for a new model of second language acquisition that includes the role of motivation.

Huarte's "The acquisition of Basque ergative case" investigates an intriguing problem related to the acquisition of Basque and the ergative system in this language. This study focuses on the production and comprehension of the ergative case marker in Basque children. According to the literature, Basque children tend to omit the ergative case marker for a period of five months since they start producing case markers in general until they master the ergative case marking. 24 Basque bilingual children were under study. The results of this study show that comprehension of Basque case marking, more precisely, of the ergative case precedes the production of it.

In "A Look at Second Language Learners' Task Motivation", Yanguas studies the relationship between task motivation and linguistic variables in a written production task, following Dönyei's Process Model of motivation (Dönyei 2000, Dönyei & Ottó 1998). This model has proved to be successful in accounting for L2 performance in oral argumentative tasks and Yanguas adapts this model to L2 performance in written tasks. The linguistic variables are based on the model for measuring frequency, accuracy and complexity by Wolfe-Quintero et al. (1998). The author chooses to measure five linguistic complexity variables in the L2 learners' written task: number of words, number of t-units, proportion of error-free t-units, number of words per t-unit and lexical variety. He aims at answering two questions: (1) is there a correlation between task motivation scores and any of the five linguistic variables? and, since motivation tasks happen to divide the subjects of the experiment into two groups, (2) are there differences in performance in any of the five variable measurements across both groups? The results show that task motivation is in direct correlation with the linguistic variables investigated; also, the high motivation group outperforms the low motivation group.

The work on interfaces is nowadays of much importance for linguistic theory, and two papers presented at BIDE05 deal with the syntax-semantics interface. Irurtzun's "The Structure of Pair-List Answers" analyzes the properties of the answers to multiple-Wh questions. The author argues that in these constructions we find a split focal structure that leads towards having a pair of elements as being the actual focus at *logical form*. This analysis provides us with a natural understanding of the question-answer pairings since all the material that stands for a variable in the question is taken to be focal in nature. Thus, treating these answers as instances of split foci, we can dispense with the theoretical primitive of 'contrastive topic' and gain in understanding of the interface phenomena observed crosslinguistically.

Gallego and Irurtzun's "Consequences of Pair-Merge at the Interfaces" explores the semantic and syntactic nature of traditional VP modifiers. In the first part of this study, they find arguments in favor of adjuncts not having to undergo computational

licensing, as a consequence of their particular phrase structure status (that is, the thesis that they occupy a separate plane; cf. Chomsky 2004). The remainder of the paper concentrates on different semantic issues which concern adjuncts: adicity, theta-roles, licensing, and possible readings in adjunct clustering. Following Martin & Uriagereka (2000) and Uriagereka (2003), the authors adopt the idea that adjuncts display two types of readings, which they call Markovian and non-Markovian: under the former one, adjuncts are interpreted as independent predicates of the event (the traditional approach stemming from Davidson 1967), whereas under the latter one, adjuncts create a framing (scopal) effect which blocks the expected entailment patterns.

Within the subfield of syntax, Penka's and de Cuba's papers investigate two different aspects of negation, based on data from several languages. In "A Crosslinguistic perspective on n-words", Penka takes a new perspective on n-words by analyzing negative concord together with two different phenomena that n-words give rise to in non-negative concord languages, namely scope splitting in German and distributional restrictions in the Scandinavian languages. These three phenomena suggest that n-words should not be analyzed as negative quantifiers but rather as morpho-syntactic markers of sentential negation. The fact that n-words show negative concord indicates that they are semantically non-negative. That n-words refer to sentential negation is manifested in the phenomenon of scope splitting. The distributional restrictions of n-words in the Scandinavian languages confirm that n-words are subject to licensing conditions that are syntactic in nature. Penka's analysis is based on the assumption that n-words are semantically non-negative and must be licensed by a (possibly abstract) negation. According to her proposal, n-words are basically of the same nature cross-linguistically and variation between languages regarding their behavior are due to parametric variation. Previous analyses fail to give a unifying account to the three phenomena discussed by Penka. However, under the author's proposal, these three phenomena are all manifestations of the same underlying nature of n-words: n-words themselves are semantically non-negative and must be syntactically licensed by negation.

De Cuba's "Negative polarity licensing, factivity, and the CP field" investigates a pair of asymmetries between the sentential complements of what have been traditionally called *factive* and *non-factive* predicates in the literature: (a) the availability of non-local Negative Polarity Item (NPI) licensing in sentences embedded under negated non-factives, but not under negated factives; and (b) the presupposition of truth in sentences embedded under factives, but not under non-factives. De Cuba argues that these asymmetries are the result of a syntactic difference in the CP field of sentential complements selected by the different classes of predicates. The article's main proposal is that there is an extra syntactic projection in the CP field that is associated with non-factive verbs like *believe*. This projection is not present under factive verbs like *regret*. The extra projection houses an operator that licenses NPIs when embedded under a matrix negative verb or negated non-factive predicate. In addition to licensing NPIs, this operator is necessary to separate the speaker from responsibility for the truth content of the embedded sentence. de Cuba provides cross-linguistic evidence from English, Basque, Mainland Scandinavian and Hungarian that this extra structure is optional, therefore, the (non-)factivity resides not in the lexical se-

mantics of the matrix predicates, but in whether or not the operator structure is selected.

Currently, there is an active group of researchers studying the syntax of Hebrew. This line of investigation is represented in this volume by two papers that analyze different syntactic structures in Hebrew and their consequences for grammatical theory in general. In “Adjectival Passives and Adjectival Decausatives in Hebrew”, Meltzer goes over a well-known distinction between adjectival and verbal passives in Hebrew claiming that a closer look at this distinction is needed. The author’s revision of the facts reveals that Hebrew adjectival passives have to be divided into two groups: on the one hand, *adjective passives* and on the other hand, *adjective decausatives*. This division is based on their interpretation: while in adjective passives there is an implicit argument in their interpretation, in adjective decausatives the external argument of the transitive verb is not part of its semantics, behaving in this respect as unaccusative verbs. Therefore, Meltzer proposes that there is a parallelism between the divisions of adjectival passives and the verbal system. Meltzer’s main proposal is then that the operations that form these adjectives are the same as the operations that form unaccusative and passive verbs. The novelty of this analysis is that no additional operations need to be stipulated in order to account for passive formation.

In “Argument Mapping and Extraction”, Preminger proposes a unified account for argument mapping and islandhood in the verbal domain. Furthermore, his proposal brings new light to the notion of external argument, as well as to the interaction between case and argument mapping. Preminger begins by examining external arguments and, focusing on object-experiencer verbs, claims that no existing framework correctly predicts which argument and when will be external. Similarly, there is no explanation as to what is special about external arguments’ syntactic mapping. Preminger further shows that some internal arguments behave syntactically as external. To address these issues, the author proposes a system in which both types of syntactic merger assumed in minimalist syntax (*set-merge* and *pair-merge*; Chomsky 2004) are used for the merger of verbal arguments. The type of merger determines the islandhood of the argument at its base position. In addition, he argues that the interaction of *pair-merge* and accusative case determines which (if any) of the arguments will be external. Choice of the type of merger is governed by the feature composition of the thematic role assigned to an argument, using the thematic feature system developed by Reinhart (2000). This approach has clear empirical advantages, when compared to existing frameworks. In addition, it provides answers for previously unresolved questions about argument externality.

BIDE 2005 has attracted a number of researchers working on Romance languages. For instance, the following two papers look at the nominal phrase structure in Spanish and Romanian, respectively. In “Prenominal and postnominal demonstratives in Spanish: A [\pm Deictic] approach”, Taboada presents an analysis for pronominal and postnominal demonstratives in Spanish that accounts for the difference in meaning and in structure between these two constructions. Her proposal also explains the complementary distribution of the article and the demonstrative in pronominal positions. Taking as a starting point Bernstein’s (1997) observation that the postnominal constructions have a deictic meaning that the pronominal

ones lack, Taboada argues that the presence of a feature [+deictic] in Dem^o triggers the raising of the demonstrative to D^o, in order to check the [+Ref] feature present in this position. The postnominal demonstrative is marked [-deictic], and this prevents it from moving, and forces the appearance of the expletive article in D^o as a last resort operation. The author claims that the two features, [± deictic] and [± Referential] must be related, since the presence of the [± deictic] feature can check the [+Ref] one, and it is decisive for the appearance of the expletive article or the movement of the demonstrative. Taboada further extends her proposal to other constructions containing a demonstrative: Postnominal demonstratives without an article and postnominal demonstratives with a place adverb can be captured with the [±Deictic] approach.

Mardale's study on "Case Marking and Prepositional Marking" analyzes the alternation between DPs morphologically marked for Genitive and PPs headed by the preposition *DE* in Romanian. Previous studies have given a unitary approach to this alternation, based on ideas about semantic similarity and free substitution of one construction by the other. However, the author proposes to treat them differently, due to a number of constraints that suggest that both types of constructions behave differently with respect to the nature of their second argument. The author shows a correlation between syntactic categories (DP versus NP), case-marking (morphological versus prepositional) and semantic type (<e> versus <e, t>).

Gallego presents another paper on Romance languages. "Phase Theory, Case, and Relative Clauses" puts forward a minimalist analysis of Spanish relative clauses that builds on Pesetsky & Torrego's (2001, 2004) claim that Case is an unvalued tense feature. Assuming Kayne's (1994) head-raising analysis of relative clauses, the paper focuses on two well-known restrictions of Romance: a) relative pronouns must be introduced by prepositions (e.g., *El chico *(con) quien habló* 'The boy (to) who(m) he talked (to)'), and b) relative clauses do not allow so-called "complementizer deletion" (e.g., *El chico *(que) vi* 'The boy (that) I saw'). The author reviews (and rejects) Bianchi's (1999) *Left Peripheral account*, and argues for a *T-to-C analysis* (see Pesetsky & Torrego 2001) consistent with the well-grounded and old intuition within the GB literature that subjects show A-bar properties in Romance. In particular, entertaining Pesetsky & Torrego's (2001) hypothesis that the Case feature of subjects can be used for checking purposes in the CP, Gallego argues that Romance behaves differently because nominative Case is assigned within the v*P phase: if correct, that explains why the Case feature of subjects becomes computationally inert when the CP cycle is activated.

A fruitful approach to linguistics involves cross-linguistic comparison in order to obtain analyses with explanatory power. This approach has been adopted by several of the papers included in this volume to explain phenomena such as clitic syncretism, the behavior of adjectives and proper nouns, and *en*-prefixation.

Pescarini's "Types of syncretism in the clitic systems of Romance" discusses the hypothesis according to which every clitic system has an elsewhere item, i.e., a non-specific clitic. This elsewhere clitic can be inserted in those cases where more specific items are ruled out the Subset Principle (Halle & Marantz 1993). The author presents data from different Italo-Romance varieties to support his thesis. Synthetic clusters, or contextual syncretisms, are sequences of clitics with a mismatch between

their morphological form and their syntactic functions. In these syncretisms, two identical clitics cannot occur together due to an OCP markedness constraint. Pescarini claims that the Subset Principle explains which clitic will be inserted to satisfy the OCP constraint, namely the elsewhere clitic given that it is the least specific in the system. Assuming Pescarini's claim that the elsewhere clitic is the best candidate for syncretism, then this predicts a relation between the process of contextual syncretism and that of absolute syncretism within a clitic system. This implies that the same clitic is involved in both types of syncretism. In fact, the author's typology shows that in all the reported varieties the clitic used in contextual and absolute syncretism is the same. Finally, Pescarini points out to some cases where the process of absolute syncretism does not involve the elsewhere clitic. In these cases, the author claims that the inserted clitic is in fact the result of a phonological development that led to homophony between two historically distinct clitics.

Giurgea's "Adjectives and Proper Nouns in Romance and English", examines the relationship between determiners and proper nouns in English and Romanian. The author describes the contrasts between English and Romance languages with respect to nominal structures containing proper nouns (PN), when adjectival restrictive modifications apply. In English the same structure, i.e., Adj+N, is used for common and proper nouns, and when the structure is restricted, *the* is inserted before the adjective. In Romance languages PN+*the*+Adj type of constructions are preferred, obligatorily in the case of Romanian. In order to account for this contrast, Giurgea proposes that Romance language use the PN + *the* + Adj type of construction with a particular type of restrictive modification, involving a selection of PNs with a familiarity condition. Moreover, in Romanian, this selection is obligatory due to morphological reasons. The analysis of the structure falls directly from a semantic rule of proper noun to common noun conversion, from which all the properties of the construction derive.

Padrosa's article entitled "Argument Structure and Morphology: the Case of *en*-Prefixation Revisited" examines a number of *en*-prefixed words in Catalan and English. She claims that they also follow the Right-hand Head Rule (RHR, Williams 1981), unlike previous analyses that considered them counterexamples to this rule. She proposes that adjective and noun conversion to verb is motivated by an *en*-suffix attachment, which occurs before prefixation, thus not violating RHR. Furthermore, Padrosa claims that the *en*-prefixation is responsible for the [-c-m] role in the case of *en*+N Vs with a locative meaning. Differences between Catalan and English are then accounted for by the use of the prefix: Catalan requires the presence of the prefix in order to account for locative Vs and English does not require it any longer, which is explained by the disassociation of the [-c-m] role from the prefix and its re-association with its base N.

References

- Bernstein, J., 1997, «Demonstratives and Reinforcers in Romance and Germanic Languages», *Lingua* 102: 87-113.
- Bianchi, V., 1999, *Consequences of Antisymmetry. Headed Relative Clauses*. Berlin: Mouton de Gruyter.

- Chomsky, N., 2004, «Beyond explanatory adequacy». In A. Belletti (ed.), *Structures and beyond: The cartography of syntactic structures*, vol. 3. Oxford: Oxford. U.P., 104-131.
- Davidson, D., 1967, «Reply to Martin, The Logical Form of Action Sentences». In N. Rescher (ed.), *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press [Reprinted in D. Davidson. *Essays on Actions and Events* (2. ed.), Oxford/New York: Oxford U.P., 129-148].
- Dörnyei, Z., 2000, «Motivation in action: Towards a process-oriented conceptualisation of student motivation», *British Journal of Educational Psychology* 70: 519-38.
- & Ottó, I., 1998, «Motivation in action: A process model of L2 motivation», *Working papers in Applied Linguistics*. London: Thames Valley University, 43-69.
- Güldemann, T., 2001, *Quotative constructions in African languages: a synchronic and diachronic survey*. Habilitationsschrift. Leipzig University.
- Halle, M. & Marantz, A., 1993, «Distributed Morphology and the Pieces of Inflection». In K. Hale & S. J. Keyser (eds.), *The View from Building 20*. Cambridge, MA: MIT Press, 111-176.
- Huber, D., 2004, «On the theoretical status of epenthesis». Presented at the *12th Manchester Phonology Meeting*, 20-22 may, 2004, Manchester.
- Kayne, R., 1994, *The Antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Krifka, M., 2001, «Quantifying into question acts», *Natural Language Semantics* 9-1, 1-40.
- Lord, C., 1993, *Historical Change in Serial Verb Constructions*. John Benjamins.
- Martin, R. and J. Uriagereka, 2000, *Collapsed Waves*. Manuscript: U. Tsukuba & UMD.
- Moore, R. C., 2002, «Fast and Accurate Sentence Alignment of Bilingual Corpora». In *Machine Translation: From Research to Real Users* (Proceedings, 5th Conference of the Association for Machine Translation in the Americas, Tiburon, California), Heidelberg: Springer-Verlag, 135-244.
- Paradis, C. & Prunet, J. F. (eds.), 1991, *The Special Status of Coronals: external and internal evidence*. San Diego: Academic Press.
- Pesetsky, D. & E. Torrego, 2001, «T-to-C Movement: Causes and Consequences». In M. Kenstowicz (ed.), *Ken Hale: A Life in Language*. Cambridge (MA): MIT Press, 355-426.
- 2004, «Tense, Case, and the Nature of Syntactic Categories». In J. Guéron & J. Lecarme (eds.), *The Syntax of Time*. Cambridge (MA): MIT Press, 495-537.
- Reinhart, T., 2000, «The theta system: syntactic realization of verbal concepts», *OTS Working Papers in Linguistics* 00, 01/TL.
- Uriagereka, J., 2003, *Pure Adjuncts*. Manuscript, UMD.
- Williams, E., 1981, «Argument Structure and Morphology», *The Linguistic Review* 1: 81-114.
- Wolfe-Quintero, K.; Inagaki, S. & Kim, H.-Y., 1998, *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawaii, Second Language Teaching & Curriculum Center.

CONSUMER EROSKI PARALLEL CORPUS¹

Asier Alcázar

University of Missouri-Columbia

Abstract

This paper introduces the Consumer Eroski Parallel Corpus, a collection of articles originally written in Spanish and later translated to three languages also spoken in Spain: Basque, Catalan and Galician. The articles have been correlated in the four languages at the sentence level automatically using Moore's bilingual sentence alignment tool (2002). The Spanish section is also annotated morphosyntactically for parts of speech using SVMtool (Giménez and Márquez 2004). The Basque, Catalan and Galician sections may be annotated in a future release with the collaboration of Computational Linguistics Groups in Spain. To my knowledge, the Consumer Eroski Parallel Corpus is the first resource to exist that encompasses a substantial body of parallel text from these four languages spoken in Spain. I would like to thank the Eroski Foundation for granting permission to share the corpus in the public domain. Making this resource public will provide additional opportunities to test, train and develop natural language processing tools in the computational linguistics community. It may also help translators as a reference. With the addition of an advanced search interface, currently under development, the corpus may be consulted by Basque and Romance linguists interested in cross-linguistic research.

1. Introduction

The Consumer Eroski Parallel Corpus (henceforth CEPAC for short) is a database consisting of all the articles published in the online version of the Consumer Eroski magazine (*revista.consumer.es*).

¹ I would like to thank the BIDE 2005 organizing committee (Irene Barbería, Rebeka Campos, Susana Huidobro and Leticia Pablos), Jon Ortiz de Urbina and, last but not least, Jon Franco, for their invitation to present at the conference, and the audience for comments and feedback. Thanks also to Joseba Abaitua and everyone at the DELi Computational Linguistics group at the University of Deusto for discussion and encouragement to pursue this project. This paper has also benefited from a presentation at and attendance to *Ordenagailuz Lagundutako Itzulpena* (Basque for Computer-Assisted Translation; Summer Courses of the University of the Basque Country, Miramar Palace, May 30-June 2, 2005). I am grateful to Josu Waliño Jr. and the Elhuyar Foundation, the presenters and audience for helping me understand the place of the Consumer Eroski Parallel Corpus among the current existing resources for Basque, Catalan, Galician and Spanish. I am indebted to Itziar Otegi for getting me in touch with Ricardo Oleaga, Iker Merchán, and Ainara Zarraga from the Consumer Eroski magazine, all of whom enthusiastically supported this project.

The Consumer Eroski magazine is a free publication produced by the Eroski Foundation (www.fundaciongrupoeroski.es). It exists online since January 1998, although the Eroski Foundation has published this magazine on paper under different names for the last 30 years. When the magazine initiated its life online, it did so under the name Consumer, which was eventually changed to Consumer Eroski. Consumer was initially published entirely in Spanish (January 1998) and gradually started to be translated to Basque (November 1998), then Catalan (January 2000) and finally Galician (April 2000). Since the current name of the magazine is Consumer Eroski, I find it fitting to name the corpus CEPAC.

The Consumer Eroski magazine has been published online for 8 years now. During this time the magazine has produced a new issue every month, with the exception of the months of July and August, which make up a joint number, for a total of 11 issues a year. Each issue contains 14 different sections of varying length that cover diverse topics (see table 1), the signature of the magazine being its consumer reports.

Table 1

List of Sections in Consumer Eroski Magazine (Spanish)

Tema de portada	Psicología	Medio ambiente	“Lo más práctico”
Análisis de productos	Miscelánea	Alimentación	Consejos
Informe	Economía doméstica	Nuevas tecnologías	Sentencias
Salud	Entrevista		Consultorio legal

Since the year 2000, the Spanish articles have been translated to the three other languages and the magazine has been published simultaneously in the four languages. Thus, the Consumer Eroski magazine has been published effectively as a multilingual edition since April 2000.

According to the Eroski Foundation, revista.consumer.es receives an average of 300,000 monthly visits. These numbers, added to the paper version, make the magazine an important publication in Spain.

At the time of submitting this paper, the current collection contained in CEPAC encompasses all the issues from January 1998 to October 2005. The amount of text for each language is substantial (as there are 14 sections times 11 issues times 8 years). All articles exist in Spanish, for this is the language in which the magazine is originally written. Consequently, the Spanish section of the corpus is the largest (1,078 articles). It is closely followed by Basque (991), Catalan (855) and Galician (806). The word count in each section of the corpus exceeds a million. The reader is referred to table 2 at the end of the next section, which details the current size of CEPAC in the different languages.

The remainder of this paper is structured as follows. Section 2 explains the rationale for CEPAC and the steps I took in building it. Section 3 provides an in-

formal overview of the currently available monolingual and bilingual corpora and places CEPAC in the context of these collections. Section 4 introduces a search interface, currently under development, that permits consulting the corpus beyond simple queries and explores how best to integrate advanced linguistic searches that are research oriented. The last section concludes with the release of CEPAC and future updates.

2. Building CEPAC

This section introduces anecdotically my first contact with the online magazine, the driving force behind a project like CEPAC, and the steps that I took to build the corpus.

2.1. First contact and reason d'être for CEPAC

The first time that I came across the Consumer Eroski Magazine I was browsing the World Wide Web in search for web pages that could be used as a reference for health care terminology in Spanish speaking countries. This endeavor was part of a collaborative effort with *Hablamos Juntos* (www.hablamosjuntos.org), a non-profit organization in the United States. Our goal was to gather a set of websites that would guide the development of a Spanish Glossary for health care terms, one that would pay attention to existing cultural differences in the use of terminology. This glossary would assist translators in the difficult task of rendering health care terms to a diverse Spanish-speaking population that are often particular to the United States health care system. The Consumer Eroski magazine stood out for its consistency in the use of health related terminology among the Spanish websites surveyed in Spain.

The Consumer Eroski magazine also presented a rare opportunity to a linguist and computational linguist like me. The magazine is published in three Romance languages (Catalan, Galician, and Spanish) and a language isolate (Basque). The parallel nature of the text, consisting of full translations of the Spanish original to three languages, called for processing this wealth of materials into a database for research use.

Although it would have been wisest to share the programming load of turning these web pages into a database with other computational linguists, it has not always been the case that parallel corpora, rather than monolingual corpora, has been the primary objective in the field. More recently, however, with the success of statistical approaches to machine translation (Knight & Marcu 2005), the search for and processing of parallel text into parallel corpora has received greater attention. For example, the European Constitution has been turned into a multilingual parallel corpus that exists for the most widely spoken languages among the member countries (<http://logos.uio.no/opus/>). It is unfortunate that this project does not yet include Basque, Catalan and Galician, among many other languages spoken in the European Union. In computational linguistics there is less of an interest in minority languages as compared to widely spoken languages that could result into profitable applications. In contrast, in linguistics and related disciplines, linguistic diversity enriches

our perspectives on language as a cognitive phenomenon. It is in gaps of this kind that CEPAC finds its niche.

2.2. From *revista.consumer.es* to CEPAC

What follows is a brief account of the steps and software involved in the processing of the Consumer Eroski magazine into CEPAC.

The first step consisted in devising a strategy for the automatic download and grouping of the original articles with their translations. To this end, I wrote a spider program in Python that would crawl *revista.consumer.es* and harvest the articles. This was an inappropriate way to access the contents of the magazine and I apologize to the Eroski Foundation for having procured the entire magazine without their permission. I initiated this project as a research activity and did not realize that in time the project could be of use to the research community. My gratitude here goes to the Eroski Foundation for their understanding and their generosity in allowing me to share this resource freely in the public domain.

The second step was to strip the articles from anything but plain text. At this stage I also developed my own programs in Python that would achieve this task. It was important to carefully study the formatting of the web pages to exploit some of these tags as sentence and paragraph dividers for use in the later stages of text processing. Some of these tags were critical in the later processing of the text. For example, the tags that specifically divide sequences of headlines or numbered/bulleted lists might be the only way to effectively divide chunks of text that have no terminating punctuation. Furthermore, later stages of processing attain more satisfactory results if they are fed the appropriate chunks of text. Understandably, a part of speech tagger will better resolve the morphosyntactic labels for a headline (an element that may be a phrase rather than a sentence) if this headline has not been put together with the following or preceding main/secondary/tertiary... headline, possibly creating a long sentence without a verb.

A third step involved breaking the plain text files into sentences with the added complexity of the inconsistency in punctuation that results from manipulating vast amounts of texts. The difficulty in this task was increased by the need to process four different languages, with overlapping yet not identical punctuation conventions. I needed to devote much time to this stage to maximize the results of the two third party applications (the aligner and the tagger) that would complete the current stage of the database. During this process I was able to treat a high number of exceptions in a systematic way, yet I also made certain case specific provisos to rescue a few hundred exceptional cases. The resulting corpus is by no means perfect but it has benefited from extensive human supervision.

To align the different language pairs, I used Robert Moore's alignment tool (2002), written in Perl, and refer the reader to his paper for the implementation details. It is of interest to note that the success of the alignment tool in the Spanish-to-Catalan and Spanish-to-Galician pairs has been higher than the Spanish-to-Basque pair. In the former two, the alignment exceeds 92% of the text, while in the latter the alignment falls to 84%.

There are at least two potential reasons to account for this difference in the results. One is to assume that the Basque translation has fewer 1-to-1 correspondences with respect to the original text. However, it is difficult to establish this fact with utmost certitude without incurring into a manual verification. All things being equal, the quality of the translation to the three languages, and the agreement observed for Catalan and Galician, suggests that over 90% of correspondences should be expected for Basque as well. This leads us to consider the other possible reason, namely that the agglutinative nature of the Basque language makes the automatic building of a dictionary of word correspondences sparser (this is an intermediate step in the alignment process), as multiple correspondences may be established between Spanish and Basque words due to the different inflectional endings. Indeed, Basque has a rich declensional system with 17 different cases, most of which have four different number forms, and there exist six alternative case forms for animates (see Zubiri and Zubiri 2000). Without a lemmatizing tool for Basque that separated the ending from the stem, it is not possible to overcome this morphological obstacle. This second reason seems a better candidate to explain the lower percentage in the absence of manual verification. It is plausible that a new alignment with Basque lemmatized text would yield more pairs of aligned sentences.

That said, it is perhaps more interesting to bring attention to the fact that the alignment from Spanish to Basque was highly successful regardless. The alignment program had to face that Basque is agglutinative (meaning fewer words: *etxe-ra* 'to the house (house-to.the)'; see table 2 in the next subsection) and a data sparsity problem arising from the rich morphology of Basque (one to many words in the Spanish-to-Basque dictionary). The usability of the alignment program attests to the robust technology employed in alignment systems and Moore's design in particular. Incidentally, the Spanish-to-Basque character ratio approximates 1.0, as seen for English, French and German (Church & Gayle 1993), so character-length based approaches could also fare well in principle.

Finally, to annotate the Spanish section with part of speech information, I used the C++ version of the part of speech tagger by Giménez and Márquez (2004), based on Support vector Machines, with the models for Spanish based on the LEXESP corpus (5.5 million annotated words). A technical description of this tool can be found at www.lsi.upc.edu/~nlp/SVMTool. I have not corrected manually any of the errors that the tagger may have done. As stated above, repeated revisions of the success of sentence dividers visibly improved the performance of the tagger, particularly in headlines and cases where sentence integrity had been compromised.

2.3. CEPAC in numbers

This subsection offers a numeric view of the corpus.

The Consumer Corpus consists of 1078 articles written in Spanish (January 1998 to October 2005). Of these, 804 have been translated to Basque, Catalan and Galician. The table below (2) shows the number of sentences, words and characters for the fully multilingual version of the corpus. As noted earlier, because Basque is an agglutinative language, it has fewer words (0.87 million) than its non-agglutinative

neighbors (between 1.09–1.19 million). Note that the number of characters is similar (6.9 million) compared to Catalan (6.9 too) or Galician (6.7).

The numbers in the table are contingent. The Consumer Eroski magazine continues to be published and is well worth periodical updates that will cause the numbers to increase. Future revisions in sentence division may modestly alter the numbers too.

Table 2

804 articles in four languages in CEPAC

	Spanish	Basque	Catalan	Galician
Sentences	59,111	56,531	56,765	57,027
Words	1.19 m	0.87 m	1.14 m	1.09 m
Characters	7.29 m	6.90 m	6.90 m	6.70 m

The actual number of sentences in each section of the corpus is higher (Spanish: 79,982; Basque: 70,496; Catalan: 60,697; Galician: 57,147). As noted in the introduction, translation started at different times. For this reason, there are pockets of articles that have not been translated to Galician, or to Galician and to Catalan. The first 10 issues of the online version of the magazine were not translated either.

3. CEPAC in the context of other existing monolingual and parallel corpora

The goal of this section is to orient the reader as to the niche that CEPAC occupies among the existing monolingual and parallel corpora. It is by no means intended to be an exhaustive list of corpora available in Spain or elsewhere. For further references, see the Linguistic Data Consortium (www ldc.upenn.edu).

The characterizing property of CEPAC as a parallel corpus is that it contains the same text in four languages, presenting unusual pairs like Basque-to-Galician, Galician-to-Catalan, Catalan-to-Basque, etc.

With the exception of the European Constitution, if it is eventually processed for all the languages spoken in the European Union, there is little chance to find parallel text of this nature. Some companies aim to gain market share in Spain and dedicate part of their budget to better market their products via linguistic localization. However, the bulk of these materials are likely to be advertising texts and product manuals (e.g. telecommunications). While literary resources may be found that are translated in the four languages, literary classics, to name one such source, it is unlikely that these texts will be localized in one place.

As we narrow our focus to particular languages or specific language pairs, we find that substantive and sometimes vast collections of text are available for Basque, Catalan, Galician and Spanish. Many of these can be accessed online, although full access to these resources is generally not possible. The following is a random and informal walk through some of these resources.

The *Real Academia de la Lengua Española* (Spanish for Royal Academy of the Spanish Language) or RAE has two large collections of literary, journalistic and oral texts online divided into two monolingual corpora: CREA for contemporary Spanish (1976 to present date) and CORDE, a historical corpus that extends back to the 19th century (www.rae.es). These texts contain information relating to their author, topic, publisher, publication year, and country of origin (Spain, Portugal, Latin America, United States, Philippines). This information provides new opportunities for studies in language variation and sociolinguistics. For example, Mayoral Hernández surveyed the position of frequency adverbials in Spanish in a recent study that made use of data from CREA (2004). The downside of CREA and CORDE is their limited access. Results are constrained to 1000 per query, these matches being either paragraphs and paragraph chunks or units smaller than a sentence.

Given that CREA does not contain the Spanish section of CEPAC, it may one day be added to its press section. While CEPAC alone cannot be used for sociolinguistic purposes (author information is not included), it provides a different opportunity to study linguistic variation in translation or across particular language pairs. To facilitate this task, it is necessary that the search interface provided for users allows more flexibility in its advanced searches (support for part of speech search, for example), and the convenience to return sentences. In any case, the distribution of CEPAC as a free resource will grant access to circa 250,000 sentences.

The monolingual corpus for Basque provided by Euskal Herriko Unibertsitatea (Basque for University of the Basque Country) contains approximately 18 million words (<http://www.ehu.es/euskara-orria/euskalareduzkoa/araka.html>), 8.7 million in the press section. CEPAC would be worth including to bring this section closer to the 10 million barrier. The Basque reference corpus is fairly limited in its temporal scope 2000-2005 compared to CREA. Its user interface is more convenient though in that it returns full sentences and allows searches complemented by part of speech information.

The largest parallel corpus for Basque and Spanish is LegeBi or ‘Official bilingual gazettes from the Basque Administration (1994-2004)’ collected by the DELi Computational linguistics Group (www.deli.deusto.es/AboutUs/Resources/LegeBi). Similarly to the European Constitution or the UN proceedings, LegeBi is another example that the administrative domain is a frequent and abundant source of parallel text. Like LegeBi, CEPAC also provides the opportunity to compare the same text in the two languages with the convenience of doing so at the sentence level. The Basque reference corpus and CEPAC more closely resemble the everyday written standard Basque.

Regarding Catalan, CucWeb (http://ramsesii.upf.es/cucweb/about.en_US.htm) is an attractive example that serves to illustrate a different type of corpus, and one that will be increasingly available. It consists of over 200 million words collected from web pages written in Catalan and, like the Basque reference corpus, can be consulted with the aid of part of speech information. To its favor, CucWeb and similar corpora possibly outweigh all other collections in its size. On the other side of the coin, CucWeb is an all-encompassing collection of texts, not an organized and annotated collection like CREA or CORDE. For this reason, this type of corpus is not a reference for the written standard.

Finally, for matters relating to Galician monolingual and parallel corpora, it is best to refer the reader to CLUVI (sli.uvigo.es/CLUVI) and references therein.

4. CEPAC as a reference tool

This section illustrates the value of the corpus as a reference tool and introduces an advanced search interface project that aims to bring queries to a higher level of abstraction.

It is little wonder that one of the advantages of parallel corpora is to present the same sentence in more than one language to bilingual speakers (translators, journalists, second language learners). By way of example, I may have come across the Spanish acronym ONG for non-profit organization and may wonder how to express such concept in Basque. Querying the corpus for ONG in Spanish, I find out that GKE is the corresponding acronym and that it stands for ‘Gobernuz Kanpoko Erakundea’.

Table 3

Sample sentences containing ‘ONG’ and their Basque counterparts

Spanish	Son cada vez más frecuentes las noticias sobre malas gestiones de ONGs y los actos reprochables de las específicamente que trabajan en proyectos de desarrollo.
Basque	Gero eta sarriago agertzen dira GKEen gestio txarrei buruzko albisteak eta garapen proiektuetan lan egiten dutenen jokaera gaitzesgarriak.
Spanish	Específicamente, la labor de las ONG es la práctica de ayuda humanitaria, es decir, la asistencia de las sociedades vulnerables y vulneradas de la que nos ocupamos sobre el terreno.
Basque	Gobernuz kanpoko erakundeon eginkizuna giza laguntza ematea da, hau da, eraso-garri eta eraso jasanak diren gizarteetan asistentzia ematea.

In time, advances in computational linguistics may provide sophisticated ways to query databases that enable linguistic research beyond its most widespread trends today (authorship, concordances, language variation, sociolinguistics, etc). However, for that matter texts should be linguistically searchable. By linguistically searchable, I mean that the search should ideally be formulated in linguistic terms and at different levels of abstraction. I continue to develop a search interface in Python that explores these new avenues (see table 4).

At this point the search for Spanish can be abstracted to parts of speech because this section of the corpus is annotated. It is important to provide the capability to search beyond particular words and we have seen several resources in the earlier section that offer this feature. For my interface, I have defined an extensive list of linguistically relevant tags that abstract the particular part of speech tags used by the tagger (reduced Parole tagset). For example, the verbs are annotated with 18 different tags: auxiliary, semi-auxiliary and main verb have different tags for the infinitive, gerund, participle, indicative, subjunctive and imperative forms. I group these 18 tags into VERB, and lower levels of abstraction like AUX for all auxiliary forms (also S-AUX, MV), IND for all indicative forms, NFVB for all non-finite verb forms, etc. This allows for searches like ‘siempre’ followed by VERB followed by INF (any infinitive form). Similar levels of abstraction are defined for the remain-

ing part of speech tags. All lower levels tags, for example VMI (main verb indicative form), are directly accessible as well.

For example: power search all questions, find those with interrogative pronouns with POS tag search, then filter results to those with, say, auxiliaries or indicative verbal forms with Verb form search.

Table 4

Overview of search levels, available categories and sample results

Level	Categories	For ex.: intended result
Morpheme	Prefix, infix, suffix and part of speech	A collection of Spanish nouns ending in <i>-ción</i>
Word	Hundred+ linguistic categories of various levels of abstraction (main verb participle > main verb non-finite > non-finite > verb)	A collection of Spanish sentences where <i>siempre</i> precedes a non-finite form
Phenomenon	All of the above plus specific search modes	A collection of sentences with clitic doubling
Construct	— Verb search — Part of speech search — Morpheme Search	A collection of absolute participial clauses
Sentence	— Regular expression search — Power search (a predefined set) — Operator search	A set of questions with interrogative pronouns
Paragraph	— Chain search (all of the above)	A set of paragraphs with overt subject pronouns

On the linguistic side, the alignment provides the possibility to compare search results across three Romance languages and a language isolate. In effect, the search results for a query in a language may be accompanied by its translations to the other three. For example, a regular expression search for clitic doubling in Spanish may find interesting companions in the Catalan and Galician translations. A power search for absolute constructions in Basque, which are ambiguous between absolute participial and gerundival clauses, may find revealing correlations in the Spanish original, which uses distinct non-finite forms. The translations to Catalan and Galician, in turn, offer the possibility to do a case study on non-finite non-complement clauses in Romance. These data have helped my joint research with Mario Saltarelli on participial clauses.

5. The Publication of CEPAC

Thanks to the permission granted by the Eroski Foundation, CEPAC will be placed in the public domain for research or reference. Future versions of CEPAC may have the Basque, Catalan and Galician sections of the corpus annotated for parts of speech and a slightly better Spanish-to-Basque alignment.

References

- Gale, W. and Church, K., 1993, «A Program for Aligning Sentences in Bilingual Corpora», *Computational Linguistics* 19:1, 75-102.
- Giménez, J. and Márquez, L., 2004, *SVMTool: A general POS tagger generator based on Support Vector Machines*. Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04). Lisbon, Portugal.
- Knight, K., D. Marcu, 2005, *Machine Translation in the Year 2004*, Proceedings of ICASSP.
- Mayoral Hernández, R., 2004, «Importance of Weight and Argumenthood on the Ordering of Adverbial Expressions», *Proceedings of the 23rd West Coast Conference on Formal Linguistics (WCCFL)*, edited by V. Chand, A. Kelleher, A. Rodríguez and B. Schmeiser. Somerville, MA: Cascadilla Press, 569-582.
- Moore, R. C., 2002, «Fast and Accurate Sentence Alignment of Bilingual Corpora». In *Machine Translation: From Research to Real Users* (Proceedings, 5th Conference of the Association for Machine Translation in the Americas, Tiburon, California), Springer-Verlag, Heidelberg, Germany, 135-244.

NEGATIVE POLARITY LICENSING, FACTIVITY AND THE CP FIELD

Carlos de Cuba
Stony Brook University

1. Introduction¹

There are a number of asymmetries in the behavior of sentential complement clauses embedded under non-factive versus factive (Kiparsky & Kiparsky 1971) predicates.² The primary goal of this paper is to examine one of these asymmetries; the licensing of Negative Polarity Items (NPIs) in sentential complements when they are embedded under these different classes of predicates with matrix negation.³ The NPI *a red cent* needs to be licensed by negation, as shown in (1).

- (1) (a) *Jon has **a red cent** to his name.
(b) Jon doesn't have **a red cent** to his name.
- (2) (a) I don't **believe** that Jon has **a red cent** to his name.
(b) *I don't **regret** that Jon has **a red cent** to his name.

In (2a), matrix negation licenses *a red cent* in the embedded clause of non-factive *believe*. However, in (2b) matrix negation fails to license *a red cent* in the clause embedded under factive *regret*. This difference is puzzling, given that the only apparent difference between the sentences is in the choice of verb.

In addition to the NPI licensing facts above, a semantic difference is also found when complement clauses are embedded under factive versus non-factive predicates. Only under factives are the complement clauses presupposed to be true. This is illustrated in (3) and (4).

- (3) (a) George **believes** [that there are WMDs in Iraq]
(b) George doesn't **believe** [that there are WMDs in Iraq]

¹ Many thanks to Pablo Albizu, Xabier Artiagoitia, Urtzi Etxeberria and Nerea Madariaga for providing Basque data and judgements, and to Enikő Tóth and Barbara Ürögdi for data and judgements from Hungarian. Thanks also to the participants of the 6th Annual CUNY/SUNY/NYU Miniconference, where an earlier version of this paper was presented, and to the organizers and participants of BIDE05.

² Note that I use the terms *factive* and *non-factive* for ease of exposition. In fact, a predicate classification along the lines of *stance verbs* as presented in Cattell (1978), and modified by Hegarty (1992) is more accurate. In this paper, when I use the term *factive* I am referring to *response-stance* and *non-stance* predicates, while *non-factive* refers to *volunteered-stance* predicates. See also footnote 3.

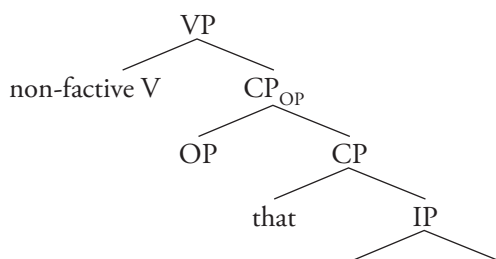
³ Two other asymmetries, namely the availability of embedded verb second in Mainland Scandinavian, and factive island constraints, will be discussed in sections 3 and 4 respectively.

- (4) (a) #George **regrets** [that there are WMDs in Iraq]
 (b) #George doesn't **regret** [that there are WMDs in Iraq]

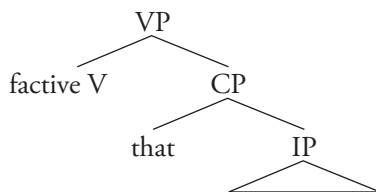
If we assume that there are no weapons of mass destruction (WMDs) in Iraq, the sentences in (3) are fine, while those in (4) are out. Non-factive predicates do not presuppose that their complements are true, while factive predicates do.

In this paper, I argue that these two asymmetries, (a) in NPI licensing, and (b) in presupposition, can be explained with a unified analysis. The main proposal involves the difference in structure shown in (5).

- (5) (a) **non-factives** (*believe, think, assume, claim, deny, doubt*)



- (b) **factives** (*regret, resent, hate, realize, forget, notice*)



I propose that there is an extra projection in the CP field selected by non-factive verbs like *deny* and *believe*, and that this projection is not present under factive verbs like *regret*.⁴ The extra projection, present in the non-factive structure in (5a) but not in the factive structure in (5b), is headed by an operator that licenses NPIs when embedded under a matrix negative verb or negated non-factive predicate, as in (2a). The lack of the operator under factives, as in (2b), leaves the NPI without a local licenser, crashing the sentence.

In addition to licensing NPIs, I propose that the operator is necessary to serve the important function of separating the speaker from responsibility for the truth content of the embedded sentence. This is in line with work by Nichols (2001), who proposes an 'assertive operator' associated with non-factive predicates, and Progovac (1994), who proposes an operator in Comp that is licensed by 'unfixed truth-values'. The proposed operator in the non-factive structure in (5a) allows for the non-factive interpretations in (3), while its absence in (4) ensures a factive reading. The present

⁴ The idea of CP-recursion being possible under non-factive predicates as in (9a) is not new. Holmberg (1986), Platzack (1986) Iatridou & Kroch (1992) and de Cuba (2002), among many others, propose CP-recursion to account for embedded verb second phenomena in many Germanic languages. See further discussion in section 3.

proposal differs from Nichols and Progovac in arguing that the operator creates syntactic structure.

The paper is organized as follows. In section 2, I present Laka's (1990) negative complementizer analysis of non-local NPI licensing. It is superficially similar to the current proposal but, as will be seen, it faces certain problems not faced by the operator analysis. Section 3 presents syntactic and semantic motivation for the proposed structure. In section 4, I briefly present an analysis of factive island phenomena, exploiting the presence of extra structure in non-factives to open an escape hatch for adjunct extraction. In section 5, I argue that the proposed operator and its associated syntactic projection are sometimes optional, and that the so-called negative complementizer in Basque can be decomposed into two separate morphemes, with the second being associated with the operator. Section 6 presents more data from Basque, examining two types of factive complementation. Section 7 is the conclusion.

2. Laka's Negative Complementizer Analysis

Laka (1990) argues that NPIs in sentential complement clauses are licensed by a negative complementizer. She gives the data in (6) to provide evidence that there is an intermediate licenser available to license NPIs interclausally.

- (6) (a) *The witnesses denied **anything**
 (b) I deny [**that**_{NEG} the witnesses denied **anything**] (Laka 1990: 169)

In (6a) the NPI *anything* fails to be licensed by the negative verb *deny* in its own clause, but in (6b) *deny* selects a negative complementizer that in turn licenses *anything* in the embedded clause. Laka shows that in Basque, unlike English, negative complementizers differ morphologically from their declarative counterparts. In (7a) the declarative complementizer (*e*)*la* appears, while the negative complementizer (*e*)*nik* appears under the negative verb *deny*. (7b) also shows that the NPI *anyone* is licensed interclausally, just like English *anything* in (6b).

- (7) (a) [Galapagoak muskerrez beterik daudela] diote
Galapagos lizards-of full are-that say-they
 'They say that the Galapagos are full of lizards'
 (b) Amaiak [**inork** gorrotoa dionik] ukatu du
Amaia anyone hatred has-that_{NEG} denied has
 'Amaia denied that anybody hated her' (Laka 1990: 204-5)

While at first blush Laka's analysis seems to account for the data, problems arise when we look more closely at English. First, complementizers are optional under non-factive verbs like *believe*.

- (8) (a) *I believe [(that) Jim **slept a wink** last night]
 (b) I **don't** believe [(that) Jim **slept a wink** last night]

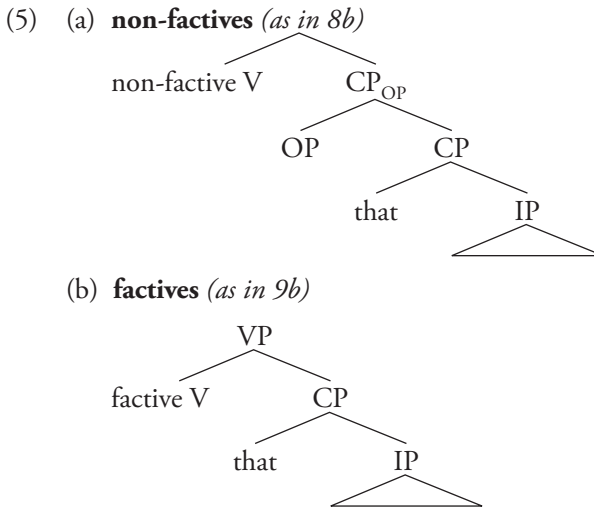
(8a) confirms that the NPI *slept a wink* is unlicensed in the absence of negation, while in (8b), it is grammatical in the presence of matrix negation. The grammaticality of (8b) is not affected in the absence of *that*. This is unexpected under Laka's

analysis, as for her the negative complementizer is the licenser of polarity items in embedded clauses.

The above problem may be solved with a PF deletion or null complementizer analysis, but a second, more serious problem arises in the complements of factive verbs in English. The NPI licensing that seems to occur interclausally in sentences like (6b) and (8b) does not take place in their factive counterparts.

- (9) (a) *I regret [(that) Jim **slept a wink** last night]
 (b) *I don't regret [(that) Jim **slept a wink** last night]

Under a Laka-style analysis, we would expect (9b) to be grammatical, with the NPI *slept a wink* licensed by a negative *that*_{NEG} selected by the negated matrix verb, as in (8b). The fact that (9b) is ungrammatical brings the negative complementizer analysis into question. Given this problem, I argue for a modification to the negative complementizer analysis that maintains the attractive points of Laka (1990), and moreover, accounts for the difference between (8) and (9). The structures in (5), repeated below, provide a difference in the syntax, with (8b) corresponding to non-factive (5a), and (9b) corresponding to factive (5b).



Crucially, the operator is a separate entity from the complementizer. Only in non-factive (5a) is there an operator available to license the NPI in the embedded clause.

3. Motivation for the Extra Structure and Operator

The clausal/non-clausal asymmetry in NPI licensing by inherently negative verbs like *deny* and *doubt*, was illustrated in (6). There is no such asymmetry induced by overt negation, as illustrated in (10).

- (10) (a) The witnesses **didn't** say **that**_{NEG} **anybody** left the room before dinner.
 (b) The witnesses **didn't** say **anything**. (Laka 1990: 179)

However, Laka's analysis of (10a) is the same as (6b). The negative complementizer is selected by the negated matrix verb, and the NPI *anybody* is licensed by the complementizer. In Basque, matrix negation also licenses an NPI in a non-negative embedded clause, as in (11). As in (10a), *anybody* is licensed by the complementizer.

- (11) **Ez** du Zurinek [**inor** etorriko **denik**] esan
no has Zurine anybody come will AUX-that_{NEG} said
 'Zurine has not said that anybody will come' (Laka 1990: 209)

Laka's proposal follows Progovac (1988, 1994) in arguing that the syntax of sentential clauses embedded under inherently negative verbs differs from the syntax of those embedded under non-negative verbs. While Laka proposes that a different complementizer is selected under negated or negative matrix verbs, Progovac argues for an operator in the head of Comp, as in (12).

- (12) I doubt [_{CP} [_C that **OP** [_{IP} **anyone** has come.]]] (Progovac 1994: 67)

For Progovac, this operator is licensed in a clause whose truth-value is not set positively. The operator also appears in other contexts with unfixed truth-values, as in (13-16), which are all non-negative polarity contexts. The NPIs in these sentences are all licensed by the operator in the absence of negation.

- (13) Yes/no questions:
 [_{CP} [_C Has **OP** [_{IP} **anyone** come?]]]
- (14) Conditionals:
 [_{CP} [_C If **OP** [_{IP} **anyone** comes]]], let me know.
- (15) Universal Quantifiers:
 [_{NP} Every man [_{CP} who [_C has **OP** [_{IP} read **anything** by Chomsky]]]]
 will attend the lecture.
- (16) Counterfactual Conditionals:
 [_{CP} Had **OP** [_{IP} **anyone** misbehaved], we would have left.]
 (Progovac 1994: 67)

Progovac argues against a Downward Entailing (DE) analysis of NPIs (Ladusaw 1980), pointing out that yes/no questions like (13) license NPIs without being DE environments. In embedded contexts the operator must be selected by the matrix predicate, as in (12), or by a quantifier, as in (15). Progovac provides further motivation for the existence of this operator. With the proper intonation, a question without Subject Auxiliary Inversion (SAI) is possible, as in (17a).

- (17) (a) He complained about his salary?
 (b) ?*He complained about **anything**?
 (c) Did he complain about **anything**? (Progovac 1994: 76-7)

If we suppose that SAI is triggered by the operator in Comp, then the contrast between (17b) and (17c) falls out: (17b) is out because there is no operator there to trigger movement, and if there is no operator in the structure, there is no licenser for *anything*.

3.1. Embedded Verb Second

The main claim of this paper is that there is an extra syntactic projection in the CP field under non-factive predicates. More evidence for this projection comes from cases of embedded verb second (EV2) in Germanic. EV2 is possible under a complementizer in the Mainland Scandinavian languages, among others (Vikner 1995: 66). This is shown in the Swedish examples in (18), where optional EV2 takes place in (18b), with the finite verb moving above negation.

- (18) (a) Rickard sa att han **inte var** hemma [Swedish]
*Rickard said that he **not was** home*
 (b) Rickard sa att han **var inte** hemma
*Rickard said that he **was not** home*
 ‘Rickard said that he was not home.’

Vikner follows many others in arguing for a recursive CP structure to accommodate the presence of an overt complementizer (in the higher CP) and verb second movement (in the lower CP). It is notable that CP-recursion is only possible under non-factive predicates.⁵ This is illustrated in (19), where EV2 is not available under factive *regret*.

- (19) (a) Rickard ångrade att han **inte var** hemma [Swedish]
*Rickard regretted that he **not was** home*
 (b) *Rickard ångrade att han **var inte** hemma
Rickard regretted that he **was not home*
 ‘Rickard regretted that he was not home.’

The proposed structures in (5) accommodate these facts easily, as there are two CP layers under non-factives to accommodate both the overt complementizer (in the upper CP) and EV2 (in the lower CP) in (18b), but only one CP available under factives, ruling out (19b).

4. Factive Islands

Nichols (2001) examines the syntax and semantics of propositional attitude reports, focusing on extraction facts. Adjunct extraction is allowed from under a non-factive predicate like *believe*, but not from under a factive like *regret* (20).

- (20) (a) How do you think that you behaved *t* ?
 (b) *How do you regret that you behaved *t* ?

⁵ Vikner (1995: 71-2) gives lists of verbs that allow and don't allow EV2 in Danish and German. Vikner found no properties shared by all the verbs in either of the lists. However, Hegarty (1992) shows that the lists can be easily categorized in terms of a slightly modified version of Cattell's (1978) *stance verbs*. The verbs that allow EV2 are *volunteered-stance*, and those that do not allow EV2 are either *response-stance* or *non-stance*. These classifications fit the data much more precisely than the standard *factive/non-factive* distinction.

Nichols argues for the special status of non-factives as opposed to factives, and that there is an operator associated with non-factives that is not present under factives. She states:

A consideration of the semantic properties of the factivity classes in terms of the character of evaluation sets of worlds reveals that the factivity problem as currently stated (e.g. “Why is extraction blocked out of factive complements?”) has been conceptualized the wrong way around, essentially backwards.

(Nichols 2001: 121)

Nichols proposes that there is an ‘assertive operator’ associated with non-factive verbs. The contribution of the operator is summarized briefly in (21).

- (21) (a) CPs have associated context variable sets $C \langle \text{speaker (source), time, world} \rangle$ needed for interpretation, as in (Schlenker 1999).
 (b) value $\langle + \text{ current speaker} \rangle$ – the actual world is necessarily included in the evaluation set – main clauses.
 (c) Factives – do not supply a $\langle \text{speaker} \rangle$ value to the context variable set – the default value is specified $\langle + \text{ current speaker} \rangle$.
 (d) Non-factives – have an assertive operator that can supply a different value for $\langle \text{speaker} \rangle$.

Nichols assigns no position in the syntax to the assertive operator. For her, syntactic differences derive from semantic properties.⁶ Factive islands like (20b) are considered the norm—in other words, adjunct movement is not allowed in normal circumstances. Only under the special condition in which the assertive operator is present, changing the $\langle \text{speaker} \rangle$ value in the evaluation set of the embedded clause, do we get an extension of the domain of movement. For Nichols, this is what allows adjunct movement in (20a) as opposed to (21b) where there is no domain extension.

A full discussion of factive islands is beyond the scope of this paper. However, I will briefly show that there is a syntactic alternative to the Nichols (2001) semantic domain extension analysis presented above. The additional CP structure I proposed in non-factive (5a) plays a crucial role in allowing for adjunct extraction from non-factive complements. McCloskey (2005), following Chomsky (1986), proposes the ‘Adjunction Prohibition’, banning adjunction to any lexically selected phrase.

- (22) Adjunction to a phrase which is s-selected by a lexical (open class) head is ungrammatical.

Adjunction to a lexically selected phrase is argued to interfere with the selectional relationship between the selecting verb and its complement. Following the Adjunction Prohibition, adjunction to a CP that is directly selected by a matrix verb (or adjective) is impossible, while adjunction to a CP selected by a functional head is possible.

⁶ Nichols (2001) does not analyze the cases of non-local NPI licensing that are presented in this paper, or cases of embedded verb second under non-factives in Germanic, two cases in which clear syntactic differences appear to go along with the semantic differences in predicate classes. I takes these cases as evidence for the structural difference proposed in (5).

sible. A CP-recursion structure like that in non-factive (5a) opens up the lower CP for adjunction, as the lower CP is not lexically selected. In the present analysis, CP-recursion is only available in non-factives, and this adjunction position provides the escape hatch for adjunct extraction. Factives, which I argue are lexically selected, as in the factive structure in (5b), do not have this adjunction possibility, and are thus islands for adjunct extraction.⁷

The work of Laka (1990) and Progovac (1994), in addition to the Mainland Scandinavian EV2 facts discussed above, provides evidence that there is a syntactic component to NPI licensing in non-negative contexts, as opposed to a purely semantic treatment. The present analysis follows the analyses of Laka and Progovac in proposing an operator that facilitates NPI licensing across a CP boundary, but departs from them by arguing that this operator creates syntactic structure. In the next section, I argue that the proposed operator and its associated syntactic projection are sometimes optional.

5. Optional extra structure?

The availability of a factive/non-factive reading correlates with syntactic structure cross-linguistically. In fact, some normally non-factive verbs can allow a factive reading of their complement, and some normally factive verbs can allow a non-factive reading. I propose that what is crucial is not whether or not the verb itself is factive or non-factive, but whether or not the extra structure is present. Basque, English and Hungarian all show syntactic and semantic effects that provide evidence that the optional interpretations are due to the presence or absence of the proposed extra structure in non-factive (5a), not to the semantics of the particular verb.

5.1. Basque

Basque shows a very interesting complementizer alternation with syntactic and semantic effects relevant to the present discussion. Laka (1990) presents a pair of sentences that are identical except for the choice of complementizer.

- (23) (a) *Iñigok ez du sinisten [lurrak eztanda egingo duela]*
Iñigo no has believed earth explode do will AUX-that
 ‘Iñigo does not believe that the earth will explode’
- (b) *Iñigo ez du sinisten [lurrak eztanda egingo duenik]*
Iñigo no has believed earth explode do will AUX-that_{NEG}
 ‘Iñigo does not believe that the earth will explode’ (Laka 1990: 211)

In (23a) the declarative complementizer (*e*)*la* is present, while in (23b) the negative complementizer (*e*)*nik* appears. Laka describes the semantic difference between the two in the following way: in (23a), *that the earth will explode* is taken to be a fact,

⁷ An argument in the specifier of CP does not affect selection the way an adjunct does. The idea I am pursuing here is that argument and adjunct movement are essentially different, with arguments moving through spec CP while adjuncts re-adjoin to CP. The methods of attachment to CP differ, with conditions on adjunct extraction being more restrictive (due to the Adjunction Prohibition). I leave a more detailed spelling out of this hypothesis to future work.

one that *Inigo* happens not to believe. In (23b), *that the earth will explode* is not taken to be a fact; it could be true or false. I argue that this is evidence for the optionality of the operator, and that when it is not present, even under a non-factive verb like *believe*, a default factive reading results. In (23a) there is no operator, while in (23b) the operator is present, resulting in the non-factive reading.⁸

In an investigation of the syntax and semantics of unselected embedded questions, Adger & Quer (2001), following Laka (1990, 1994) and Uribe-Etxebarria (1994), argue that the Basque negative complementizer can be decomposed into two constituents, as in (24).

- (24) $-(e)n + ik$
 C Partitive (Adger & Quer 2001: 116)

The first is a bound C morpheme that appears in several complementizer uses (relative clauses, embedded questions, etc.), while the second corresponds to what Basque grammars traditionally label as partitive case marking. This proposal can be straightforwardly adopted to the present analysis if we take *ik* in (24) to be associated with the proposed operator.⁹ When it is absent in (23a), a factive reading results, and when it is present in (23b) a non-factive reading results.

5.2. English

A similar example to (23) can be found in English when non-factive *believe* is stressed, as in (25).

- (25) (a) I don't believe [that Liverpool won last night].
 (b) I don't BELIEVE [that Liverpool won last night].

As in (23), the sentences in (25) use the same traditionally non-factive verb *believe*. The truth of the complement clause in (25a) need not be determined, but (25b) forces a factive reading. The fact that complements of the same verb can have two different semantic interpretations provides more evidence that factivity is not provided by the verb alone.

5.3. Hungarian

Hungarian embedded clauses also exhibit two different patterns, one for non-factives and one for factives (de Cuba & Ürögdi 2001).

- (26) (a) Azt hiszem hogy Mari okos.
 it-ACC I-think Comp Mary smart-is
 'I think that Mary is smart.'

⁸ Laka analyzes the difference in meanings in (23) as a result of *(e)nik* needing to be interpreted under the scope of the negation that selects it, while *(e)la* is interpreted outside the scope of matrix negation. Sentences headed by *(e)nik* remain in the scope of matrix Infl and V, while those headed by *(e)la* undergo Quantifier Raising at LF.

⁹ I am exploring a different line of analysis than Adger & Quer, who analyze the partitive case marker in (24) as a polar sensitive determiner like English *any*.

- (b) (**Azt*) sajnálom hogy Mari okos.
it-ACC I-regret Comp Mary smart-is
 'I'm sorry that Mary is smart.'

In (26), the pronominal element *azt* can be argued to come from the lower clause, since it represents the object of the matrix verb, which is the lower CP itself.¹⁰ This pronoun is only present in cases where the matrix predicate is non-factive.¹¹ The fact that *azt* bears accusative case provides evidence that it comes from below the verb. One could imagine that *azt* is the overt realization of the operator that I am proposing, present in the non-factive case but missing with factives. When *azt* is not present in non-factive context, a factive reading results, as in (27).

- (27) (a) *Azt* mondta Péter, hogy későn kezdődik a meccs.
that-ACC said Peter Comp late begins the match
 'Péter said that the match will begin late' (but we don't know if it's true)
 (b) *Mondta* Péter, hogy későn kezdődik a meccs.
said Peter Comp late begins the match
 'Péter told (me) that the match will begin late' (and in fact it will)

These facts are consistent with those in Basque and English in (23) and (25). If the operator is not present, even under a non-factive verb, a factive reading results.

The semantic effects of the pronominal element in Hungarian can also be seen with some factive verbs. The pronominal *Úgy* (*so*) shows similar semantic effects to *azt* in Hungarian (Enikő Tóth, Barbara Ürögdi, p.c.).¹² When *úgy* appears with a factive verb like *know*, as in (28b), a non-factive reading results.

- (28) (a) *Tudja* János, hogy Mari okos,
knows John that Mary smart-is
 'John knows that Mary is smart'
 (fully factive reading)
 (b) *Úgy* tudja János, hogy Mari okos.
so knows John that Mary smart-is
 'John knows that Mary is smart'
 (to the best of John's knowledge, Mary is smart)

The presence of *úgy* in (28b) removes the factive interpretation of the embedded clause, while in the absence of *úgy*, the default factive reading results (28a). I take the facts from Basque, English and Hungarian in this section to provide evidence that

¹⁰ For an analysis along these lines, see Lipták (1998), as discussed in Kiss (2002: 234-5).

¹¹ If the sentences have neutral intonation, then factive predicates don't allow *azt*, while non-factives do. However, if *azt* is in focus position and heavily stressed, it then becomes grammatical, as in (i) (Enikő Tóth, Barbara Ürögdi, p.c.).

(i) *AZT* sajnálom, hogy Mari megbukott a vizsgán.
 'It's that Mari failed the exam that I'm sorry for.'

At present I have no account for this. I leave this case of focused *azt* to future research.

¹² Kiss describes *úgy* as an alternative to the demonstrative pronoun *azt*, serving the semantic function of expressing a reservation concerning the truth of the subordinate proposition (Kiss 2002: 233).

the proposed operator is optional under some verbs. The observed semantic differences in factivity are due to the presence or absence of the operator, not simply the lexical semantics of the verb.

5.4. NPI Licensing in Optional Cases

The analysis presented thus far predicts that NPI licensing should only take place when the operator is present, which is signaled in Basque by (*e*)*nik*. Confirmation of this is in found in (29).

- (29) (a) *Íñigok **ez** du sinisten [**ezerk** eztanda egingo duela]
 *Íñigo **no** has believed **anything** explode do-will AUX-**that**
 ‘Íñigo does not believe that anything will explode’
 (b) Íñigok **ez** du sinisten [**ezerk** eztanda egingo duenik]
 Íñigo **no** has believed **anything** explode do-will AUX-**that**
 ‘Íñigo does not believe that anything will explode’ (Laka 1990: 211)

As in (23), the only difference between the two sentences in (29) is in complementizer choice. Under the present analysis, this difference in NPI licensing possibility results from the lack of an operator in (29a) and its presence in (29b).¹³ In non-factive cases where there is no matrix negation or inherently negative verb, the operator is present, but has no phonological realization, as is the case in English.

More support for the analysis in this section comes from English, where the Basque NPI licensing facts in (29) also seem to carry over to (30).

- (30) (a) I **don’t** believe [that Jon smokes **anymore**.]
 (b) *I **don’t** BELIEVE [that Jon smokes **anymore**.]

Recall from example (25), that when stressed, *believe* forces a factive interpretation of the embedded clause. The present analysis predicts that the operator is responsible for both the non-local licensing of NPIs, and the availability of a non-factive interpretation. The ungrammaticality of (30b) is thus expected, as there is no operator available to license the NPI *anymore*, even though *believe* is typically non-factive.

The data in this section provides evidence that semantic effects of the operator on truth-value evaluation go along with visible differences in the syntax, in the form of NPI licensing in Basque and English, and extra morphosyntax in Hungarian. I argue that these syntactic licensing and semantic interpretation differences are a result of the presence or absence of the proposed operator and its related structure in (5).

6. Factive Cases in Basque

As was shown in section 2, in English, non-local NPI licensing is available in non-factive contexts, but not in factive ones. This was illustrated in (8b) and (9b) above, repeated below in (31a) and (31b) respectively.

¹³ This is essentially the same as Laka’s (1990) analysis, where the negative complementizer (*e*)*nik* licenses the NPI in (29b). My analysis differs in that the operator and the complementizer are separate, accounting for the factive/non-factive NPI licensing asymmetry in English in (2).

- (31) (a) I don't believe [(that) Jim **slept a wink** last night]
 (b) *I don't regret [that Jim **slept a wink** last night]

In Basque, 'true factive' verbs (*regret, resent, hate*) don't take finite complements, but a nominalization construction similar to the English NP-gerund, as in (32).¹⁴ Constructions using the complementizers (*e*)*la* or (*e*)*nik* under true factives are ungrammatical in Basque.

- (32) (a) Zuriñek Jon joan **izana** deitoratu du
Zuriñe-ERG Jon gone have-ART regret AUX
 'Zuriñe regrets that John left' (lit: John having left)
 (b) Zuriñek ez du Jon joan **izana** deitoratu
Zuriñe-ERG no AUX Jon gone have-ART regret
 'Zuriñe doesn't regret that John left' (lit: John having left)

Unlike true factives, 'semi-factives' do take finite complements (*notice, realize, forget*), as in (33).¹⁵

- (33) (a) Zuriñe [Jon joan **dela**] ohartu da.
Zuriñe Jon go AUX-that notice AUX
 'Zuriñe has noticed that John has already left'
 (b) Zuriñe ez da konturatu [gaur astelehena **dela**]
Zuriñe no AUX realize today Monday AUX-that
 'Zuriñe hasn't realized that today is Monday'
 (c) Zuriñek ez du ahaztu [gaur bere egun-a **dela**]
Zuriñe no AUX forget today her day-ART AUX-that
 'Zuriñe hasn't forgotten that today is her birthday'

All of the grammatical sentences in (33) use the complementizer (*e*)*la*. However, if the complementizer is switched to (*e*)*nik*, as in (34), the sentences all become very awkward, if not totally out.¹⁶

- (34) (a) ?Zuriñe ez da ohartu Jon joan **denik**.
Zuriñe no AUX notice Jon go AUX-that
 'Zuriñe has not noticed that John has already left'
 (b) *?Zuriñe ez da konturatu [gaur astelehena **denik**]

¹⁴ *True factives* are also referred to as *emotive factives* in the literature.

¹⁵ It is interesting to note the different behavior here of so-called *emotive* or *true factives* like *regret, resent, and hate* vs. *semi-factives* like *realize, forget* and *notice*. Icelandic has a related phenomenon.

(i) Ég hata að Jón skuli hafa barið Maríu (ii) Ég veit að Jón hefur barið Maríu
I hate that John should have hit Mary I know that John has (ind.) hit Mary
 (iii) Ég tel að Jón hafi barið Maríu
I believe that John has (subj.) hit Mary

The true factives, which use the nominalization structure in Basque, correspond to true factives in Icelandic, which take complement clauses with the modal *skuli* (i). Semi-factives in Icelandic take complement clauses in the indicative mood (ii), while non-factives take the subjunctive mood (iii) (Thráinsson 1979: 211-13).

¹⁶ p.c. Xabier Artiagoitia and Nerea Madariaga.

- Zuriñe no AUX realize today Monday AUX-that*
 ‘Zuriñe hasn’t realized that today is Monday’
- (c) *?*Zuriñek ez du ahaztu [gaur bere egun-a denik]*
Zuriñe no AUX forget today her day-ART AUX-that
 ‘Zuriñe hasn’t forgotten that today is her birthday’

The fact that the (*e*)*la* examples in (33) are fine, while the (*e*)*nik* examples in (34) are degraded conforms to what we would expect given the present analysis; the (*e*)*nik* examples in (34) are all out because (*e*)*nik* cannot appear in a factively evaluated CP.¹⁷ Finally, (35) illustrates the expected result that an NPI should not be licensed under factive *realize*, regardless of the complementizer chosen.

- (35) (a) **Zuriñe ez da konturatu [inor etorriko denik]*
Zuriñe no AUX realize anybody come-FUT AUX-that
 ‘Zuriñe hasn’t realized that anybody will come’
- (b) **Zuriñe ez da konturatu [inor etorriko dela]*
Zuriñe no AUX realize anybody come-FUT AUX-that
 ‘Zuriñe hasn’t realized that anybody will come’

In (35a), (*e*)*nik* cannot be selected by a factive verb like *realize*, while in (35b), non-local NPI licensing is not possible in the absence of the operator, signaled by the choice of (*e*)*la*.

7. Conclusion

In this paper, I have argued that there is an extra syntactic projection in the CP field that is associated with what have traditionally been called non-factive verbs. This extra structure houses an operator that when under matrix negation licenses negative polarity items non-locally. The operator is also responsible for licensing unfixed truth-values in embedded clauses by allowing for a change of the <speaker> (source) value in the evaluation set, allowing the actual world to be excluded. I have provided cross-linguistic evidence that this extra structure is sometimes optional—when it is missing, a factive interpretation results, and when it is present a non-factive interpretation results. The presence or absence of the structure also affects the availability of non-local NPI licensing.

References

- Adger, D. and J. Quer, 2001, «The syntax and semantics of unselected embedded questions», *Lg* 79.4, 2001.
- Cattell, R., 1978, «On the source of interrogative adverbs», *Lg* 54.1:61-77.
- Chomsky, N., 1986, *Barriers*. Cambridge: MIT Press.
- de Cuba, C., 2002, «A Phase-derivational Analysis of Limited Embedded V2 in Mainland Scandinavian». Qualifying Paper. Stony Brook University, NY.

¹⁷ In some Basque dialects however, (*e*)*nik* is possible under factives (Urtzi Etxeberria, p.c.). At this point I will only consider the dialects that disallow factive (*e*)*nik*, and leave these cases to future research.

- and B. Ürögdi, 2001, *The Role of factivity in the syntax and semantics of embedded clauses*. Unpublished manuscript: Stony Brook University, NY.
- Hegarty, M., 1992, *Familiar complements and their complementizers: On some determinants of A'-locality*. Unpublished manuscript: University of Pennsylvania.
- Holmberg, A., 1986, *Word order and syntactic features in the Scandinavian languages and English*. Doctoral Dissertation, Department of General Linguistics: University of Stockholm.
- Iatridou, S. and A. Kroch, 1992, «The Licensing of CP-recursion and its relevance to the Germanic verb-second phenomenon», *Working Papers in Scandinavian Linguistics* 50: 1-25.
- Kiparsky, P. and C. Kiparsky, 1970, «Fact», ed. M. Bierwisch and K. Heidolph. In *Progress in Linguistics*. The Hague: Mouton.
- Kiss, K. É., 2002, *The syntax of Hungarian*. Cambridge University Press, UK.
- Lipták, A., 1998, *A magyar fókuszemelések egy minimalista elemzése*. In Büky and Maleczki (eds.), 93-116.
- Ladusaw, W. A., 1980, *Polarity sensitivity as inherent scope relations*. Garland, New York.
- Laka, I., 1990, *Negation in syntax: On the nature of functional categories and projections*. Doctoral Dissertation, MIT, Cambridge, Mass.
- 1994, *On the syntax of negation*. New York and London: Garland.
- McCloskey, J., 2005, «Questions and questioning in a local English». To appear in *Cross-Linguistic Research in Syntax and Semantics: Negation, Tense and Clausal Architecture*. R. Zanuttini, H. Campos, E. Herburger and P. H. Portner (eds.), Georgetown U. P., expected 2005.
- Nichols, L., 2001, *The syntax and morphosyntax of propositional attitude complementation*. Ms. Harvard University, Cambridge, MA.
- Progovac, L., 1994, *Negative and positive polarity: a binding approach*. Cambridge U. P.
- Platzack, C., 1986, «Comp, infl, and Germanic word order.» ed. L. Hellan and K. Koch Christensen. *Topics in Scandinavian Syntax*: 185-234.
- Schlenker, P., 1999, *Propositional attitudes and indexicality; a cross-categorial approach*. PhD. MIT.
- Thráinsson, H., 1979, *On Complementation in Icelandic*. Garland Publishing.
- Uribe-Etxebarria, M., 1994, *Interface licensing conditions on PSIs: A theory of polarity and tense interactions*. University of Connecticut dissertation, Storrs, CT.
- Vikner, S., 1995, *Verb Movement and Expletive Subjects in the Germanic Languages*. Oxford U. P.

ASPECTS OF QUOTATIVE CONSTRUCTIONS IN IBERIAN SPANISH

Ricardo Etxepare

Iker, CNRS

1. Introduction¹

In colloquial speech, main clauses in Iberian Spanish can be headed by an overt complementizer (Spitzer 1942; Porroche Ballesteros 1995; García 1996; Etxepare 1998), which does not seem to be linked to any other term:

- (1) a. Oye, el Barça ha ganado la Champions
Listen, the Barcelona has won the Champions League
b. Oye, que el Barça ha ganado la Champions
listen that the Barça has won the Champions League
- (2) a. Si viene mi madre, el tabaco es tuyo
if comes my mother the tobacco is yours
b. Si viene mi madre, que el tabaco es tuyo
if comes my mother that the tobacco is yours

The apparent optionality of the complementizer masks an important semantic difference between the (a) and (b) cases. Consider (1): as a typical declarative sentence, (4a) constitutes an assertion, whose propositional content is that a given soccer team (Barcelona) has won the Champions League. When compared with (1a), (1b) contributes the additional meaning that someone else (who is not the speaker) said (1a), such that the (speaker's) utterance of (1b) constitutes a report of what has been said. (1b) is thus reported speech (Coulmas 1986), unlike (1a), which is an ordinary assertion. The two sentences would be produced in quite different settings: (1a) could be uttered for instance by a person who has been to the finals match, with the purpose of spreading the news. In such a setting, (1b) would be definitely odd. (1b), on the other hand, would be appropriate if I were listening to the radio and

¹ Different parts of this research have been presented at the Symposium Syntactic Functions-Focus on the Periphery, held in Helsinki (2003), at the 2003 meeting of the project *Architecture de la Phrase (Axe CP)* of the Fédération Typologie et Universaux (CNRS), at the Spring Linguistics Meeting of the Spanish and Portuguese Department at the University of California Santa Barbara (2004), at the Second Vasconian Meeting on Semantics and Pragmatics in Sara (2004), at the 9th International Pragmatics Conference (2005) in Garda Lake, Italy and in BIDE2005, which provided the basis for this paper. I want to thank all the audiences. I also acknowledge financial support from the project "Architecture de la Phrase" of the Fédération Typologie et Universaux at CNRS, directed by Alain Kihm and Hans Obenauer, as well as the project BFF2002-04238-C02-01, from the MCYT, led by Miriam Uribe-Etxebarria.

heard the news that Real Madrid won the Champions League. Then I could chose to report on the news by employing the comp-initial sentence. In that case, I would be implying that I got the news from someone else's saying, as was the case.

Now consider (2). Imagine the following situation: two teenagers are secretly smoking in a room. Suddenly, fearing that his mother could show up and find out, one tells the other (2a): "si viene mi madre, el tabaco es tuyo". By saying that, the speaker asks the other person to act as if the tobacco was his or hers, if mother comes. By saying (2b), the speaker asks something more than just pretense: he or she asks the other person to *say* that the tobacco is his or hers. If the roommate doesn't *say* so, he or she will not be complying with the speaker's request. In both cases, the semantic contribution of the clause initial complementizer is that of adding (or referring to) an explicit speech event.

Taken together, (1)-(2) clearly show that the phenomenon of main clause complementizers in Spanish must be kept separate from clause typing phenomena, where the clause initial Comp marks a sentence as declarative (cf. Gascon, Rohlf's 1977; Campos 1992): the presence of the sentence initial complementizer doesn't *make* a sentence declarative (the sentences in (1) and (2) are all declarative). Despite the apparent "hearsay" interpretation of (1b), it should also be distinguished from evidentiality phenomena, where the assertoric force of the sentence is modified by particles that indicate the source of the information (Givon 1982; Chafe and Nichols 1986; Rooryck 2001a,b; Dendale and Tasmowski 2001): hearsay particles would be decidedly odd in contexts such as (2b), where the assertoric force is not in question, but rather an explicit saying is requested from the hearer. Finally, sentence initial complementizers in Spanish do not have an emphatic function (cf. Arabic *?inna*, Shlonsky 2000). No particular emphatic function is associated to the complementizer *que* in these cases. The presence of the root complementizer in those contexts seems to be related to the presence of a speech eventuality in the logical form of the sentence, which is otherwise absent.

This paper is a preliminary analysis of root complementizer constructions in Spanish. I will defend the view that root complementizer constructions in Spanish involve a speech eventuality which is mapped in the grammatical representation of the sentence. I will suggest that the speech eventuality is represented as an indefinite description contributing an existential quantification over a variable which ranges over utterances, adapting a proposal of Lahiri's (2002) for "quotative" dependents in Spanish. This indefinite description can enter a more complex structure, consisting of the indefinite description plus a light verb. This complex verbal structure is analogous to what in other languages are called "quotative verbs" (see Lord 1993; Frajzingier 1996; Güldemann 2001; Amberber 1996, among many others): it introduces a report and frames constituents which may show properties of direct or semi-direct discourse. The paper is divided as follows: it starts by describing some basic properties of quotative constructions in Spanish (section 2), showing that the underlying speech eventuality has a grammatical representation and that it seems to possess properties of speech act operators, in the sense of Krifka (2001). Section 3 analyses the internal event configuration of the quotative VP. Section 4 discusses how the dependent clause relates to the quotative structure. Section 5 discusses the temporal anchoring of the quotative construction. It is shown that features other than Tense, such as vocatives or locative demonstratives, containing deictic indices, help anchor

the quotative construction in the discourse. In this regard, quotative constructions in Spanish seem to behave as verbal structures in languages which have no grammatical Tense (see Ritter and Wiltschko 2005). Section 6 provides a technical solution to the absence of overt Tense and Person morphology in reduced quotatives.

2. Basic properties of root complementizer constructions

2.1. Adverbial modification and pronominal anaphora

A standard test to determine which kind of underlying abstract object we are dealing with in a given construction is to find out under which semantic context anaphoric reference to that abstract object is possible (Asher 1993; Ormazabal 1995). Consider in this regard the following contrasts:

- (3) a. Si viene mi madre, el tabaco es tuyo, #y rápidamente/educadamente
if comes my mother the tobacco is yours and rapidly/politely
'#If my mother comes, the tobacco is yours, and politely/quickly'
- b. Si viene mi madre, que el tabaco es tuyo, y rápidamente/educadamente
if comes my mother that the tobacco is yours, and rapidly/politely
'If my mother comes, you say that the tobacco is yours, and you say it politely/quickly'

Manner adverbs such as *rápidamente* "rapidly" or *educadamente* "politely" modify events. Whereas in (3a) the sentence cannot be followed by a conjunct that contains an event-modifying adverb, in (3b) that same continuation is possible, under the interpretation that the modified event corresponds to the speech event associated to the complementizer. That is, what must be rapid or polite is the saying that the tobacco is yours. The adverbial modifiers can also precede the complementizer, directly modifying the speech event:

- (4) a. #Si viene mi madre, educadamente/rápidamente el tabaco es tuyo
if comes my mother politely/quickly the tobacco is yours
- b. Si viene mi madre, educadamente/rápidamente que el tabaco es tuyo
if comes my mother politely/quickly that the tobacco is yours

A similar test can be devised with pronominal anaphora:

- (5) a. Si vienen a buscarlo, [está fuera]_i
if they-come to look-after-cl, he-is away
#Pro_i apenas te costará.
barely cl take-an-effort-fut
'(#)If they come after him, he is away. It will barely take you any effort.'
- b. Si vienen a buscarlo, [(tú) que está fuera]_i
if they-come to look-after-cl, (you) that he-is dead long ago
Pro_i apenas te costará.
barely cl take-an-effort-fut
'If they come after him, (you) say_i he is away. It_i will barely take you any effort.'

The null pronoun is the subject of the verb *costar* “take you an effort/cost”. This is a typical insertion context for event denoting anaphors (Petersen 1982). As shown by the oddness of (5a), the pretense main clause can not provide the right type of antecedent. In (5b), the pronoun can refer to the speech event associated to the complementizer, and the sentence is good.

2.2. Lexical content, thematic structure and complement selection

Events typically have participants, and clause initial complementizers in Spanish support the presence of nominal expressions which play a thematic role in the speech event. Consider for instance the following cases:

- (6) a. Tu padre que cuándo vas a ir a visitarle
 your father that when you-are-going to visit him
 “Your father is saying: ‘when are you going to visit me?’”
 b. Si viene mi madre, tú a ella que el tabaco es tuyo
 if comes my mother, you to her that the tobacco is yours
 ‘If my mother comes, you say to her that the tobacco is yours’

In (6a) the nominal expression preceding the complementizer is interpreted as the agent in the speech event. In (6b) we have an agent and a goal. The presence of thematic material in (6) indicates the presence of a tacit verbal structure. This structure must then involve a lexical feature, one on which a verbal scaffolding can be constructed. This lexical feature, I will call [linguistic communication], a term I adopt from Ross’s performative hypothesis (1970). The necessity of invoking such an abstract feature (instead of a more elaborate one, akin to ordinary verbs of saying) is illustrated by cases such as (7), where the understood eventuality is not, strictly speaking, speech-like:

- (7) Etxepare, 100 veces en la pizarra y con buena letra
 Etxepare, 100 times in the blackboard and with good writing
 que no tirarás nada a tus compañeros
 that neg you-will-throw anything to your classmates
 ‘Etxepare write 100 times and with good handwriting that you will not throw anything to your classmates’

This lexical feature, present also in (7), only selects speech act complements, unlike ordinary verbs of saying. Consider in this regard a typical verb of saying such as *decir* “say/tell”:

- (8) a. Pedro ha dicho quién viene
 Pedro has said who is-coming
 ‘Pedro said who is coming’
 b. Pedro ha dicho que quién viene
 Pedro has said that who is-coming
 ‘Pedro said: who is coming?’

As other verbs of speech, such as *preguntar* “ask” and *responder* “answer”, the verb *decir* “say” can take complements of different semantic dimensionality: it can take sets

of propositions, such as interrogative or exclamative *wh*-complements, or it can take utterances or speech act complements (Plann 1982; Brucart 1992; Suñer 1993, Lahiri 2002). In the latter case, the verb introduces a dependent which is interpreted with its own illocutionary force. Note that only dependents of the (8b) type can be understood as questions (as requests for information). This difference is keyed to the presence of the complementizer *que*. Unlike ordinary verbs of saying, the tacit verb of quotative constructions can only select dependents with their own illocutionary force: sentences analogous to (8a) are impossible in the quotative construction:

- (9) a. *Juan, tu padre Ø quién viene
 Juan, your father who comes
 ‘(Intended meaning) Juan, your father is saying who is coming’
 b. Juan, tu padre Ø que quién viene
 Juan, your father that who is-coming
 ‘Juan, your father is saying: who is coming?’

2.3. Quantification over speech eventualities

A straightforward way of checking whether some particular construction involves an underlying eventuality or not is to see whether the purported eventuality can be quantified over (Parsons 1990). Consider the following sentences:

- (10) a. Tú siempre que qué bonito es aquello, ya estoy harto
 You always that how beautiful is that already I-am sick-of-it
 ‘You are always saying: “how beautiful that is” I am sick of it’
 b. Tú siempre que cuándo viene
 You always that when he-is-coming
 ‘You are always saying “when is he coming?”’

In (10a,b) a universal quantifier precedes the complementizer and takes scope over the speech event. The result, as shown in the translations, is a universal quantification over speech events. Not all quantifiers are acceptable in root complementizer structures, though. Negative adverbs, for instance, are out.²

² This constitutes evidence that the tacit speech event is not simply an elided verb *say* or some other verb of communication. Well known elision processes, such as gapping, are insensitive to the nature of the surrounding quantifiers:

- (i) a. Siempre me dices que haga esto pero nunca [me dices] que haga lo otro
 always cl tell that do-subj. this but never cl tell that do-subj. D other
 ‘You always tell me that I should do this, but never [] that I should do that’
 b. Siempre me dices cuándo llegan pero rara vez [me dices] cuántos vienen
 always cl tell when they-arrive but seldom cl tell how-many come
 ‘You always tell me when they are coming but seldom how many they are’

The same happens with quantificational subjects:

- (ii) Todo dios dice cuándo va a venir, pero nadie [e] cuánto tiempo se va a quedar
 everyone says when he/she-is coming, but noone how long he/she-is-going to stay

- (11) a. *Tu nunca que cuántos vienen
 You never that how many come
 ‘You never say “how many are coming?”’
 b. *Tu nunca que qué bonito es
 You never that how beautiful is
 ‘There’s never a saying of the sort “How beautiful it is”’
- (12) a. *Tu rara vez que cuántos vienen
 You seldom that how-many come
 ‘You seldom say “how many are coming?”’
 b. *Tu rara vez que qué bonito es esto
 You seldom that how beautiful is this
 ‘You rarely tell me “how beautiful this is”’

This restriction on negative quantifiers extends to the subject of the quotatives:

- (13) a. Aquí todo dios que cuándo les van a subir el sueldo
 here everyone that when cl they-are-going to raise the salary
 ‘Everyone is saying: “When are they going to give us a raise?”’
 b. *Aquí nadie que cuándo les van a subir el sueldo
 Here noone that when cl they-are-going to raise the salary
 c. *Aquí poca gente que cuándo les van a subir el sueldo
 here few people that when cl they-are-going to raise the salary

Those restrictions on the kind of quantifier that can precede the tacit speech event are reminiscent of the intervention effects that have been observed on split and *in situ* operator constructions (Beck 1996; Honcoop 1998; Pesetsky 2000; Mathieu 2002, among others). Consider for instance the contrast in (14a,b), (apud Beck 1996):

- (14) a. *Was glaubt niemand wer da war?
 What thinks no one who there was
 ‘Who does no one think was there?’
 b. Was glaubt jeder wen Karl gesehen hat?
 What thinks everyone who Karl saw?
 ‘Who does everyone think Karl saw?’

Whereas a split construction cannot have an intervening downward entailing quantifier, it accepts a universal quantifier. This would suggest that we treat the tacit, underlying speech eventuality as a quotative operator, undergoing LF movement across the quantifier (as proposed for split or *in situ* operator constructions in Pesetsky 2000, or Mathieu 2002). This quotative operator would then be similar to the one proposed for quote structures by Collins (1997). There are good reasons however not to proceed that way. When analysed closely, the set of quantifiers which induce intervention effects and those which can not precede quotative *que* are not the same: universal quantifiers are interveners in so called split constructions, but they can precede quotative *que*. The intervention effect of universal quantifiers in split constructions, unlike that of negative quantifiers, is only apparent at the interpretive level (Beck 1996): they disallow the wide scope reading of the *in situ* operator. Take again the contrast between (15a) and (15b):

- (15) a. *Was glaubt niemand wer da war?
 What thinks noone who there was
 ‘Who does no one think was there?’
 b. Was glaubt jeder wen Karl gesehen hat?
 What thinks everyone who Karl saw?
 ‘Who does everyone think Karl saw?’

Beck (1996: 20) credits Pafel (1993) for observing that although grammatical, (15b) lacks a reading in which the *in situ* operator is interpreted as having wide scope. It must be obligatorily interpreted under the scope of the universal quantifier, eliciting a pair-list answer. Given that *in situ* operators obligatorily undergo LF movement to an A' position, Beck is forced to argue that universal quantifiers such as *every*—unlike negative quantifiers such as *no one*—undergo QR to a position that c-commands the LF landing site of the *in situ* operator. This analysis extends to other quantifiers such as *meisten* ‘most’, whose intervention effect is also only apparent at the interpretive level. Other possible quantificational expressions such as indefinites or *only-DP* nominals avoid the intervention effect by adopting a ‘referential’ interpretation, instead of a quantificational one. Beck’s conclusion concerning universal quantifiers is reinforced by Honcoop’s (1998) analysis of intervention effects as an instance of more general restrictions on the construction of discourse referents. The blocking effect of universal quantifiers is then immediately evident in cross-sentential anaphora:

- (16) Todo dios_i tiene un coche. #pro_i es demasiado caro
 everyone has a car it-is too expensive

Unlike the complex array of scope interaction facts that we observe in split and *in situ* operator constructions, the quantificational restrictions in the domain of quotative constructions are disarmingly simple: only the universal quantifier can quantify over the underlying speech event. Other quantifiers which can get around intervention configurations are simply impossible:

- (17) a. *Aquí la mayoría que el Madrid ha ganado la Champions
 here most that the Madrid has won the Champions
 b. *Aquí sólo Juan que el Madrid ha ganado la Champions
 here only Juan that the Madrid has won the Champions
 c. *Aquí algunos que el Madrid ha ganado la Champions
 here some that the Madrid has won the Champions
 d. *Aquí alguien que el Madrid ha ganado la Champions
 here someone that the Madrid has won the Champions
 e. *Aquí muchos que el Madrid ha ganado la Champions
 here many that the Madrid has won the Champions

The special behavior of universal quantifiers vis-à-vis the rest of the quantifiers sets apart the quantificational restrictions on quotative constructions from those observed in any known typology of intervention effects.

If the quantificational restrictions operating in quotative constructions can not be traced back to those noted in standard intervention configurations, what alternative do we have? Krifka (2001, 2003) has recently argued that the illocutionary force

of the sentence is semantically represented by a speech act operator and that speech acts can be quantified over under restrictions which turn out to be identical to the ones operating on quotative constructions. Krifka shows that certain logical operations, such as disjunction or negation, are hardly applicable to speech acts. This is so because according to him the (denotation) domain of speech acts does not constitute a boolean algebra, but at most a semi-lattice. In that domain, certain operations such as conjunction are well defined, whereas disjunction and negation are not. Consider for instance the following assertion (from Krifka 2001: 16):

- (18) Al made the pasta and Bill made the salad
 a. I assert: Al made the pasta and Bill made the salad
 b. I assert: Al made the pasta, and I assert: Bill made the salad

The conjunction operator, as shown in (18a,b), can be interpreted either as conjoining the asserted propositions or as conjoining two acts of assertion. Unlike conjunction, disjunction is only interpreted at the propositional level:

- (19) Al made the pasta or Bill made the salad
 a. I assert: Al made the pasta or Bill made the salad
 b. #I assert: Al made the pasta, or I assert: Bill made the salad

Disjunction at the speech act level amounts to canceling the illocutionary force of the sentence. Speech acts also lack negation as a general operation (20). As Krifka notes, it is not clear what the complement of a speech act could possibly be.

- (20) a. #I don't assert: Al made the pasta
 b. #Noone asserts: Al made the pasta

Quotative constructions, like speech act operators, reject disjunction and accept conjunction:

- (21) Tu padre que se está haciendo tarde y/#o
 your father that cl is getting late and/or
 tu madre que no os espera más
 your mother that neg cl wait-for anymore
 'Your father is saying: "it is getting late" and/??or your mother: "I am not waiting for you anymore"'

Krifka extends his theory of speech acts to the analysis of pair-list readings in question-quantifier interactions. As shown by Chierchia (1993), only (non-negative) universal quantifiers give rise to pair-list interpretations. The pair-list reading induced by universal quantifiers directly follows from a semantic representation where the universal quantifier takes scope over a question act:

- (22) Which dish did every boy make?
 ↔ For every boy x: Which dish did x make?
 ↔ Which dish did Al make, which dish did Bill make, and which dish did Carl make?

If speech act operators participate in scope interactions, then the fact that only universal quantifiers give rise to pair list answers follows from the fact that only uni-

versal quantifiers can take scope over speech acts. And this is so because only universal quantifiers are generalized conjunctions. As Keenan and Faltz (1985) show, the universal quantifier is logically equivalent to a series of conjunctions, unlike the rest of the quantifier types:

- (23) a. Every boy came \leftrightarrow Al came **and** Bill came **and** Carl came...
 b. Some/A boy came \leftrightarrow Al came **or** Bill came **or** Carl came
 c. No boy came \leftrightarrow **Not**: Al came **or** Bill came **or** Carl came
 d. Most boys came \leftrightarrow Al came and Bill came, **or** Al came and Carl came,
or Bill came and Carl came

Quantifiers other than the universal one would amount to logical disjunction of the terms of the partition induced by the question. But the terms of the partition are, as shown, speech acts, and speech acts do not allow disjunction. If this is the right approach to the quantificational and boolean restrictions on the occurrence of quotative constructions, we are led to conclude that the tacit verb of quotative constructions, which supports thematic material and only selects speech act dependents, is akin to Krifka's speech act operators.

2.4. Aspectual auxiliaries

Krifka's insight into the nature of speech act quantification opens the way to account for another intriguing asymmetry in quotative constructions. Quotative constructions can have overt aspectual auxiliaries (frequentative *andar* "walk", locative-be *estar* "to be in a location", and *empezar* "start"), in which case they are inflected with Tense and Agreement features:

- (24) a. Tus padres siempre andan que cuándo iremos a visitarles
 your parents always they-walk that when we-will-go to visit them
 'Your parents are always saying: "when are you coming to visit us?'"
 b. Tu hijo siempre estaba que cuándo podría montarse en el tiovivo
 your son always was that when he-could get in the carroussel
 'Your son was always saying: when will I get in the carroussel?'
 c. El niño empezó en el coche que cuánto faltaba para llegar
 the child started in the car that how long it-would-take to get there
 'The child started saying in the car: "when are we going to get there?'"
 d. Tu padre sigue que cuándo vamos a ir a visitarles
 your father keeps that when we will-go to visit them
 'Your father keeps saying: "when are you coming to visit us?'"

Not all aspectual verbs are allowed in this context, though:

- (25) a. *Tus padres suelen que cuándo vamos a ir a visitarles
 your parents use-to that when we-will go to visit them
 b. *Tus padres terminaron/acabaron que porqué no íbamos a visitarles
 your parents ended up/finished that why neg we-went to visit them
 c. ??El crío va en el coche que cuándo vamos a llegar
 the child goes in the car that when we-will get there

- d. *El niño continúa que cuándo vamos a llegar
the child continues that when we-are-going to et there

The asymmetry suggests, first, that the cases in (24) are not instances of a general process of coercion (one which would raise the type of a *que*-clause —a propositional entity— into a higher semantic type —that of speech acts— under the context [Asp __]). Then, they also show that whatever distinguishes between the possible and the impossible cases does not have to do with the form of an eventual elided constituent: all cases would be good with an overt verb of saying, and in both (24) and (25) we seem to elide either gerunds or infinitives. Rather, the relevant feature seems to be the special habitual/iterative character of the good cases. Both *andar* “walk”, by itself a frequentative aspectual auxiliary, and *estar* “be” and *empezar* “start”, when they are combined with a gerund or an infinitive, support iterative readings. *Terminar* “end up” and *acabar* “finish” don’t. *Ir* “go” and *continuar* “continue” do not license iterative readings of the event they embed: consider in this regard the contrast between *andar/empezar* on the one hand (26) and *ir/continuar* on the other (27).

- (26) a. El niño anda corriendo continuamente
the child walks running continuously
‘The child stops and starts running continuously’
b. El niño empieza a correr continuamente
The boy starts to run continuously
‘The boy stops and starts running once and again’
- (27) a. El niño va corriendo continuamente
the boy goes running continuously
‘Once and again, the boy goes running’
b. El niño continúa corriendo todo el rato
the boy continues running all the time
‘The boy continues running all the time’

Whereas in (26), the aspectual auxiliaries allow (and in the case of *empezar*, force) a discontinuous reading of the event denoted by the lexical verb, this is not the case for *ir* “go” and *continuar* “continue”, where the adverb only modifies the process of running. We could consider iterative aspect as an instance of generalized conjunction at the event level. That is, instead of partitioning a domain of individuals, the aspectual auxiliaries would partition the event domain. Since in this case, partition involves speech eventualities, the same restriction applies.

Ir and *continuar* become better with quotative dependents when they are conjugated with imperfective aspect:

- (28) a. Los niños iban en el coche que cuándo íbamos a llegar
the children went-imp in the car that when we-were-going to get there
‘The children went saying: “when are we going to get there?”’
b. ?Los niños continuaban que cuándo íbamos a llegar
the children continued that when we-were-going to get there
‘The children continued saying: “when are we going to get there?”’

objects (the clausal dependents (but see section 3.2)), and they admit manner, aspect and agent oriented adverbial modifiers:

- (29) a. Tú rápidamente que el tabaco es tuyo
 you rapidly that the tobacco is yours
 ‘You say quickly that the tobacco is yours’
 b. Aquí todo dios constantemente que se trabaja demasiado
 here everyone constantly that cl works too much
 ‘Here everyone is saying constantly that people work too much’
 c. Aquí todo dios de motu propio
 here everyone voluntarily
 que quiere trabajar más, es lo nunca visto!
 that he-wants to work more, is the never seen
 ‘Here everyone voluntarily that he/she wants to work more, it is un-
 heard-of!’

The adverbial evidence suggests that the tacit verbal structure projects a complex VP. Assuming recent work on the internal syntactic structure of complex events (Harley 1995; Kratzer 1996; Travis 2000; Borer 2005 among many others), the tacit VP would be composed of a light verb *v* and a sister VP:

- (30) ..._[AspP] Asp⁰ [_{vP} DP v⁰ [_{VP} V⁰...CP]]]

For the moment being (we will come back to the internal structure of the *vP* in section 3.2.), we can take *V* to be the locus of the lexical feature [linguistic communication], *v* the locus of agentivity. Indirect objects would occupy the specifier of VP (Larson 1988; Baker 1996).

If we follow Cinque (1999) in the idea that adverbs are (inner) specifiers of functional projections (or alternatively, that different subsets of them are associated to given functional domains (Ernst 2002; Tenny 2000)), we are led to propose further functional structure above the VP. Verbal quotatives admit temporal modifiers:

- (31) a. Tu padre ayer que no quería venir
 your father yesterday that neg he-wanted to-come
 ‘Your father was saying yesterday that he didn’t want to come’
 b. Juan, tu padre hoy que no quiere venir
 Juan, your father today that neg he-wants to-come
 ‘Juan: your father is saying today that he doesn’t want to come’
 c. Ya verás, tu padre mañana que no quiere venir
 you’ll see, your father tomorrow that neg he-wants to-come
 ‘You will see, your father will be saying tomorrow that he does not want to come’

In the three cases, the temporal adverb is understood as modifying the underlying speech eventuality. We amend accordingly the structure in (31) to (32) (see section 4 for a revision):

- (32) [_{TP} T⁰ [_{AspP} Asp⁰ [_{vP} DP v⁰ [_{VP} DP V⁰ [CP]]]]]

It turns out that modifiers of a higher type, such as modal (root or epistemic) or factive adverbs, can not modify the quotative clause:

- (33) a. (Tu padre) *sorprendentemente/*supuestamente/*probablemente/*quizá
 Your father surprisingly/allegedly/probably/perhaps
 (tu padre) que cuándo venís
 everyone that when you-are coming
 ‘Your father surprisingly/allegedly/probably is saying: When are you coming?’

Truth-functional operators are also excluded from quotative constructions:

- (34) a. (*Sí) tu padre (*sí) que cuándo viene
 yes your father yes that when he-is-coming
 ‘Your father does say/says indeed: “when is he coming?”’
 b. (*No) tu padre (*no) que cuándo viene
 neg your father neg that when he-is-coming
 ‘Your father does nor say: “when is he coming?”’

Auxiliary-less quotative constructions therefore, seem to have a reduced clausal structure. They have Tense (phonologically realized with overt aspectual auxiliaries, hidden with temporal and aspectual adverbs), but offer no room for higher syntactic projections belonging to a CP domain (in the sense of Rizzi 1997; also Cinque 1999). The latter is confirmed by the fact that Case-marked topics (35a), question words (35b) and contrastive foci (35c), which must move into the CP-domain (see Torrego 1984, for wh-words; Uriagereka 1995; Laka 1990; Etxepare and Uribe-Etxebarria 2005; Campos and Zampini 1991, for contrastive foci; Cinque 1977 for Case-marked topics) are out in the quotative clause:

- (35) a. *A Juan, tu padre que están esperando
 to Juan your father that they-are waiting
 ‘To John, your father is saying that they are waiting’
 b. *Quién [] que viene? c. *PEDRO [] que viene
 Who that comes Pedro that comes
 ‘Who says he is coming?’ ‘PEDRO is saying that he is coming’

The ban extends to yes/no questions and exclamatory sentences, when they target the whole quotative construction:

- (36) a. *Tu padre [e] que viene? (sí o no?)
 Your father that he-is-coming (yes or no)
 ‘Is your father saying that he is coming?’
 b. *¡Tu padre [e] que viene!
 Your father that he-is-coming
 ‘Your father says that he is coming!’

Quotative constructions in Spanish thus seem to come in two types: a full one, showing an overt auxiliary; and a minimal one, deprived of any overt morphosyntactic material beyond the event participant. The alternance between the full and the simpler quotative constructions, when approached with a comparative eye, immediately recalls a well known phenomenon in languages which have quotative construc-

- d. Así gritó (él)/?Lo gritó e. Así piensa él/Lo piensa
 thus he-screamed/cl he screamed thus he-thinks/cl he-thinks
 'He screamed thus/??He screamed it' 'He thinks thus/he thinks it'

All verbs of speech can make anaphoric reference to their understood dependent through the manner demonstrative *así*.⁶ *Lo* is also available, but not always (cf. Examples (c) and (d)).⁷ *Lo* and *así* on the other hand, do not make reference to identical elements: *así* anaphorically refers to speech act dependents; whereas *lo* does not seem to be fit for that function.

⁶ It is often cited that we can also ask about the content of a verb of speech by *how*, rather than by *what*:

- (i) a. Cómo has dicho? b. Qué has dicho?
 How have-you said What have-you said?
 'How did you say?' 'What did you say?'

This possibility extends to other verbs of speech:

- (ii) a. Tú cómo preguntaste? b. Cómo respondiste?
 You how asked? How you-answered
 'How did you frame your question?' 'How did you answer?'

Verbs of communication which can frame a quotative dependent but which do not strictly involve speech, are less good with *how*:

- (iii) a. ??Cómo pensaste? b. ??Cómo has escrito?
 How you-thought How you-have written
 'How did you write?'

So the distribution of *así* and of *how* is not identical. Another case where a manner wh-pronoun seems to introduce dependents of verbs of speech is reportive *como* "how":

- (iv) Pedro contó/dijo como de pequeño iban a bañarse al río
 Pedro told how when a child they-went to swim in the river
 'Juan told us) how when he was a child, they went to swim in the river'

Reportive *como* is also good with verbs of thinking, when the dependent is introduced by a preposition:

- (v) Juan pensó *(en) cómo iban a bañarse al río de pequeños
 Juan thought in how they-went to swim-infinitive to the river when children
 'Juan thought of how they used to go swimming to the river'

Manner of speech verbs, however, are not good with reportive *como*:

- (vi) *Juan respondió/gritó como no quería ir al río
 Juan answered/shouted how neg he-wanted go-infinitive to the river
 'Juan answered/shouted how he didn't want to go to the river'

I will leave aside the possible commonalities and differences between anaphoric *así* "thus" and wh-pronoun *cómo* "how". I add, in this regard, that manner demonstratives are a very common source for the grammatical category of complementizer (see for instance Heine and Kuteva 2002: 273-274).

⁷ *Lo* is impossible precisely in those cases where the verb of speech seems to only select for speech act dependents. Plann (1982) suggested that the verbs that select a double complementizer/quotative dependent are the same which can introduce quotations. Interestingly, quotations can not be introduced by clitic *lo*:

- (i) Juan (??lo) dijo: "cuándo venís?"
 Juan cl said: "when are you coming?"

- (41) a. Así dijo él, que cuándo íbamos a reunirnos
 thus he-said he, that when we-would meet
 ‘He said thus, (namely) when we would gather together’
 b. ??Lo dijo él, que cuándo íbamos a reunirnos
 cl he-said he that when we would meet
 ‘He said it, namely when we would gather together’
- (42) a. Que cuándo íbamos a reunirnos, así dijo él
 that when we-would gather, so he said
 ‘When we would finally gather. Thus he said’
 b. ?Que cuándo íbamos a reunirnos, lo dijo él
 that when we-would gather, cl he-said he
 ‘When we would finally gather, he said it’

The choice between *así* and *lo* does not only reflect the nature of the dependent: it is also associated to different aspectual structures. Verbs of saying can denote an aspectually complex eventuality, projecting both a process and a resultant state (in the sense of Pustejovsky 1991; Tenny 1994, 2000; Levin and Rappaport 1998; Borer 2005 and many others). The resultant state, which delimits the complex event, can be explicitly measured by prepositional phrases headed by *hasta* ‘till’ and locative *en* ‘in’:

- (43) a. Juan dijo **hasta la saciedad** que no le gustaba esa solución
 Juan said till exhaustion that neg cl he liked that solution
 ‘Juan said to exhaustion⁸ that he didn’t like that solution’
 b. Juan dijo **en (muy) pocas palabras** que no le gustaba esa solución
 Juan said in very few words that neg cl he-liked that solution
 ‘Juan said in very few words that he didn’t like that solution’
 c. Juan dijo **en parte** que no le gustaba esa solución
 Juan said in part that neg cl like that solution
 ‘Juan said in part that he didn’t like that solution’

PPs like *hasta la saciedad*, *en pocas palabras* and *en parte* are incompatible with a durative adverbial phrase such as *durante horas* ‘for hours’:

- (44) Juan dijo en pocas palabras/en parte que no le gustaba (*durante horas)
 Juan said in few words/in part that neg cl he-liked for hours

⁸ “Exhaustion” here is interpreted not as the subject’s exhaustion (he became exhausted by repeating his discontent), but as the exhaustion of the saying event: no further event of expressing his disagreement can be performed. Unlike other *hasta*-PPs, *hasta la saciedad* here can not be topicalized:

- (i) *Hasta la saciedad, Juan dijo que...
 Till exhaustion, Juan said that...

This should be related to the similar contrast in English:

- (ii) a. He will bore them to death
 b. *To death, he will bore them

Verbs of saying also license restitutive *again*, which according to Tenny (2000) modifies a resultant state. Restitutive *again*, unlike non restitutive *again*, can be informally paraphrased by “two times”. The difference between the two readings of *again* naturally arises with an indefinite object:

- (45) a. Otra vez, Juan dijo algo b. Juan dijo algo otra vez
 again, Juan said something Juan said something again

Whereas (45b) can mean that Juan said one same thing twice; (50a) only means that Juan spoke again.

The possibility of measuring or modifying a resultant state is not indifferent to the *así/lo* alternation: only the presence of a D-clitic *lo* licenses a resultant state that can be measured. Bare *así* complements don't:

- (46) a. (*Así) dijo (*así) hasta la saciedad, que...
 (thus) he-said (thus) to exhaustion, that...
 b. (??Así) dijo (??así) en pocas palabras, que...
 (thus) he-said (thus) in few words, that...
 c. (*Así) escribió otra vez, que...
 (thus) he-wrote again, que...
- (47) a. Lo dijo hasta la saciedad c. Lo dijo otra vez
 cl he-said to exhaustion cl he-said again
 b. Lo dijo en pocas palabras
 cl he-said in few words

If quotative dependents, which are anaphorically referred to by manner demonstratives, are not regular objects but manner dependents, we expect measuring the saying event to be impossible in quotative constructions. The prediction is borne out:

- (48) a. Tu padre dijo hasta la saciedad quién era esa persona
 your father said to exhaustion who was that person
 b. Tu padre dijo (??hasta la saciedad) que quién era esa persona
 your father said to the exhaustion that who was that person
 ‘Your father said to exhaustion: “who is that person?”’
- (49) a. Tu padre dijo en parte/en muy pocas palabras quiénes iban a venir
 your father said in part/in very few words who were coming
 ‘Your father said in part/in very few words who were going to come’
 b. Tu padre dijo (*en parte/en muy pocas palabras) que quienes iban a venir
 ‘your father said in part/in very few words: “who are coming?”’

We can summarize our findings as follows: verbs of speech can denote aspectually complex or simplex events. Complex events contain a resultant state, and have propositional dependents which can only be referred to by a determiner clitic. Simplex events are bare processes, and have quotative dependents which can only be referred to through *así* “thus”, a manner demonstrative.

If the tacit verb of speech is a quotative verb that frames quotative dependents, we expect it to behave as a simplex verb of saying. That this is so is shown by the sharp ungrammaticality of those cases where we try to delimit the speech eventuality:

- (50) a. Ayer tu padre (*en parte) que no pensaba asistir
 yesterday your father in part that neg think attend
 ‘(lit) Yesterday your father that he was not thinking of attending’
 b. Ayer tu padre (*hasta la saciedad) que se aburría
 yesterday your father to exhaustion that cl be-bored
 ‘(lit) Yesterday your father to exhaustion that he was bored’
 c. Ayer tu padre (*en muy pocas palabras) que estaba decepcionado
 yesterday your father in very few words that he-was upset
 ‘(lit) Yesterday your father in very few words that he was upset’

Overt modification of the tacit speech eventuality is otherwise possible. Modifiers of process subevents are perfectly admissible:

- (51) a. Ya verás, tu padre mañana a voz en grito que no quiere ir
 you’ll see, your father tomorrow shouting that neg wants to-go
 ‘(lit) You’ll see, your father tomorrow in a shouting manner that he does not want to go’
 b. (Handing you the phone:)
 Tu padre en vasco/medio afónico que vayamos
 your father in Basque/half-voiceless that we should-go-there
 ‘Your father says in Basque/half-voiceless that we should go there’

The tacit speech eventuality in Spanish quotative constructions is therefore a bare process verb, with no inner resultant state, and with a manner component that is occupied by the quotative dependent. As Amberber (1996) suggests for Amharic quotatives, the quotative verb comes very close to an unergative verb of communication such as *hablar* “speak/talk”. Unergative verbs do not, by themselves, license resultant states (Hale and Keyser 1993; Levin and Rappaport 1998; Tenny 2000; Ritter and Rosen 1998, among many others), do not support measures of the type mentioned,⁹ and use a manner anaphor to refer to their dependent:

- (52) a. *Juan habló en parte c. Juan (*lo) habló así
 Juan spoke in part Juan cl spoke thus
 b. *Juan habló en pocas palabras
 Juan spoke in few words

⁹ Unlike with *decir* “say” or the tacit quotative verb, *hasta la saciedad* can combine with *hablar* “speak”:

- (i) Juan habló hasta la saciedad
 Juan spoke to exhaustion

However, unlike with *decir*, the PP-modifier here is compatible with durative adverbs:

- (ii) Juan habló hasta la saciedad durante horas

This suggests that, despite appearances, the PP does not delimit the event in this case. It should be taken as modifying the process of speaking. In this sense, it just an intensifier, as *a lot* in English “he talked a lot (for hours)”.

4. The quotative dependent

4.1. Manner CP-s

At first glance, the quotative dependent does not appear in the appropriate form to be anaphorically referred to by a manner demonstrative in Spanish: it is not a prepositional, participial, gerundive or adverbial phrase. The unexpected manner reading associated to the clausal dependents, however, is not peculiar to quotative constructions: manner readings of CPs in Spanish are otherwise well-attested in modifying functions (see Alvarez 1999; Demonte and Masullo 1999):

- (53) a. Juan hablaba [que no callaba]
 Juan spoke that neg shut-up-past
 'Juan spoke in such a way as he wouldn't shut up'
 b. Juan llegó a la meta [que no se tenía en pie]
 Juan arrived to the winning-line that neg cl stand up
 'Juan arrived to the winning post in such a way that he would not stand up by himself'
 c. Juan canta [que da gusto]
 Juan sings that it-gives pleasure
 'Juan sings in a very pleasurable way'

The manner-CPs can be taken to modify different parts of the events involved: with unaccusatives, they can modify a resultant state (54); with unergatives, they must modify the process part (55a,b):

- (54) a. Juan llegó a la meta que tuvo que entrar en el botiquín
 Juan arrived to the line that he-had to come inside the first-help box
 'Juan arrived in such a state that he had to come to the first-help box'
 (55) a. *Juan habló que perdió la voz
 Juan spoke that he lost his voice
 'Juan spoke in such a way that he lost his voice'
 b. Juan habló que no calló
 Juan spoke that neg he-shut-up
 'Juan spoke in such a way that he would not shut up'

Since quotative verbs are unergative verbs, sentences corresponding to structures like (55b) should be possible, and they are:

- (56) Tu padre otra vez que no calla
 Your father once again that neg shuts-up
 '(Intended meaning) Your father is talking once again in such a way that he will not shut up'

In (56) the CP-dependent does not express what the subject (your father) says, but how he does it, modifying the process subevent in the speech eventuality.

Another typical occurrence context for CP-clauses is as attributive predicates (Demonte and Masullo 1999):

- (57) a. Tu padre está que no puede más
 your father is that neg he-can more
 'Your father is in such a state that he can't do more'
 b. Tu padre anda que no puede más
 your father walks that neg they-can more
 'Your father is usually in such state that he can't do more'

In this case, the dependents can embed features of (semi-)direct speech:

- (58) Tu padre está/anda que cuándo venís
 your father is/walks that when you-are-coming
 'Your father is saying: "when are you coming?"'

But (58) looks identical to our quotative constructions. It is then natural to ask whether the structure in (58), involving an attributive CP-predicate, and the quotative constructions are the same structure, and whether the relation between the CP and the rest of the clause is identical in both cases. There are reasons to think that it is not, and that both cases, despite their obvious similarities, are not exactly the same structure. As I will show however, there is much to learn from (57) regarding the relation between the CP and the verbal structure in quotative constructions.

4.2. Manner CP-s and quotative dependents

Alvarez (1999), summarizing the traditional descriptive work on the subject, concludes that the structures in (58) are consecutive modifiers from which an antecedent term of degree has been omitted:

- (59) Tu padre está (tan mal/tan cansado) que no puede más
 your father is so bad/so tired that neg he-can more
 'Your father is so ill/tired that he can't do more'

Under this view, *que* is a term of relation, a linker between a degree expression and a proposition. *Que* can also relate a noun and a proposition in a complex degree expression. These are the so-called "consecutive-relatives":

- (60) Me hicieron un recibimiento que para ti lo quisieras
 cl they-did a welcome that for you it you-would-like
 'They made me such a welcome that you would have wanted it for yourself'

Consecutive-relatives form a syntactic constituent, as shown by constituency tests:

- (61) a. [Un recibimiento que para ti lo quisieras], me hicieron
 a welcome that for you it you-would-like, cl they-did
 'They did to me a welcome that you would like it for you'
 b. [lo que me hicieron] fue [un recibimiento que para ti lo quisieras]
 cl that cl they-did was a welcome that for you cl you-would-like
 'What they did to me was a welcome that you would have wanted it for yourself'

I will adopt the traditional view of consecutive predicates as containing a silent degree operator (62a). In the case of consecutive-relatives, the degree operator overtly modifies a noun (62b):

- (62) a. [_{CP} Degree Op_i que [[Clause] t_i]]
 b. [_{CP} [_{DegreeP} Op Degree⁰ [_{DP} un recibimiento]]_i que [[Clause] t_i]]

(62a,b) are null operator constructions, of the sort we find in relative clauses, with the only difference that the operator is not an argument but a predicate. Underlying consecutive constructions there is a predication relation between a clause and a degree, one that tells us that a given proposition (that your father can't do more, in 59), should be interpreted as a degree. This is the classical configuration of complex noun phrase constructions as analyzed by Stowell (1981). Now, the structures in (62) pose an immediate syntactic problem for *que*-clauses as attributive predicates, (as noted by Demonte and Masullo 1999): there is substantial evidence that CP-clauses in Spanish are nominal structures (see recently Picallo 2002). But nominal attributive predicates with *estar/andar* are impossible in Spanish:

- (63) *Tu padre está/anda un loco
 your father is/usually-is a fool

The only way a noun phrase can contribute to an attributive predication is with the support of a preposition:

- (64) Tu padre está/anda como un loco
 Your father is/walks like a fool
 'Your father lately is like crazy'

I will propose that predicational relations like (57) are mediated by a null preposition.

4.3. Evidence for a null aspectual preposition

Let me start with a few observations in this regard. The first observation is related to a special restriction applying to bare consecutive clauses: they are aspectually bound. In the case of unergative predicates, they are bound to modify the process subevent. This is not a general property of fully fledged consecutive clauses, which can modify several types of entities. Consider for instance the following contrast:

- (65) a. *Tu padre habló que fue inoportuno
 your father talked that it-was inappropriate
 'Your father talked in such a way that it was inappropriate'
 b. Tu padre habló tanto que fue inoportuno
 your father talked so much that it-was inappropriate
 'Your father talked so much that it was inappropriate'

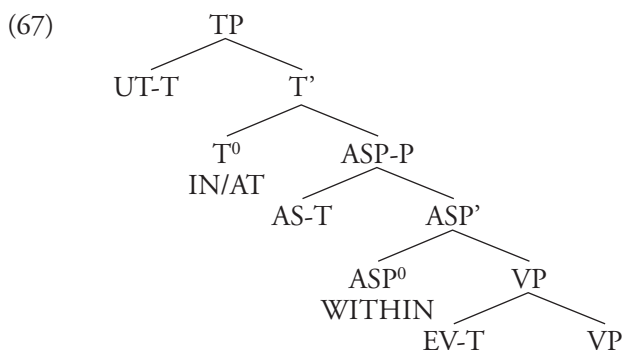
With an overt degree predicate, the consecutive clause can modify the proposition: it is the fact of talking so much which makes it inappropriate. Nothing like that is possible with bare CP-clauses. Clausal attributive predicates therefore seem to

be bound to a subset of the possible modifying possibilities allotted to consecutive clauses. The modifying ability of CP-clauses is strictly linked to an aspectual configuration.

It is very common for aspectual relations to be expressed by topological or spatial notions, and those are typically conveyed (though not only) by adpositions. Demirdache and Uribe-Etxebarria (2000, 2002, 2004) develop an analysis of aspectual and temporal relations in which tenses and aspects are spatiotemporal predicates (also Stowell 1996; Zagona 2003). Those predicates establish topological relations —of precedence, inclusion and subsequence— between two arguments that denote time intervals. The main ingredients of Demirdache and Uribe-Etxebarria's topological approach to temporal-aspectual systems are two: a set of given temporal intervals; and a set of relations between them. The temporal intervals are drawn from Klein's system (1994): the Utterance Time (UT-T), the Time of the Assertion (AS-T) (the portion of time about which an assertion is made) and the Event Time (EV-T). Tenses and aspects order those intervals by means of a limited set of relations. Tense orders the temporal intervals denoted by the UT-T and the AS-T, whereas aspect orders the temporal intervals denoted by AS-T and EV-T. As an illustration of how their system works, consider their analysis of the present progressive in English (66):

(66) John is reading *Invisible Man*

The temporal syntax of the present progressive in English consists of the following predicative structure:



The progressive is a spatiotemporal predicate with a meaning akin to *within*. It establishes an inclusion relation between its two arguments: it orders the Assertion Time (the temporal portion of the event time which constitutes the object of assertion) inside the Event Time (the total reading time). It thus focalises a subinterval of the Event Time, as shown in the schema below:

(68) EV-T
 ...[.....[.....].....]...
 AS-T

The progressive focalises a phase in the internal temporal structure of the eventuality. The interval so captured, does not include the borders of the eventuality (its

inception and end) and so directly yields the unbound interpretation of the progressive. Finally (67) describes a present eventuality because the utterance time is *in* (not *after* or *before*) the assertion time (the time interval *about* which an assertion is made).

The intervals which are ordered by means of aspectual and temporal heads are semantically structured as the figure and the ground of the relation (Talmy 1983). The specifier of the aspectual or temporal head is mapped as the figure of the topological relation, while the complement of that head is mapped as the ground. Hale (1986) argues that spatiotemporal relations can be uniformly defined in terms of an opposition: the coincidence between the figure and the ground is either central, or non central. A predicate which expresses central coincidence specifies that the situation, the path, the positioning of the figure (F) centrally coincides with the ground (G). A predicate of non-central coincidence specifies that the localisation, path or positioning of F does not centrally coincide with G. The later predicates divide in (at least) two different types: the [-central; +centripete] predicates place the figure before the ground, or indicates that the path F follows goes towards G. The [-central;+centrifuge] predicates indicate that the localisation of F is after G, or that the path followed by F departs or comes from G. Adpositions are the typologically privileged means to express those topological notions, so it is not surprising to find them once and again across languages in the aspectual/temporal realm (see a.o. Bybee, Pagliuca and Perkins 1994).

CP-modifiers are strictly aspect-bound. From an aspectual point of view, they could be represented as one of the terms in a binary relation: the one established between the main process event and the event denoted by the modifying clause. In a sentence like (69) the event represented by (not) shutting up takes up the same space occupied by the process event of talking:

- (69) Tu padre habló que no calló
 your father talked that neg he-shut up
 ‘Your father talked in such a way that he would not shut up’

The two eventualities thus seem to be related by a predicate with the properties of central coincidence. Demirdache and Uribe-Etxebarria (2005) note that prepositions of central coincidence, as the unmarked case in the system, tend to be phonologically unrealized. I conclude that the relevant preposition underlying the relation between the main event and the modifying-clause is an aspectual predicate, realized as a null, central coincidence, preposition. A sentence like (69) therefore, will be syntactically represented as (70), with AST-T representing the assertion time. The *que*-clause, headed by a silent preposition, would modify the assertion time of the talking event:

- (70) ..._[AspP]_[NP] AST-T _[PP] P_{central} _[CP] [Degree P] *que* _[SC] CP (DegreeP)]]] Asp⁰ _[VP] DO¹ [talk]]]

The structure should informally read as: “DO talk in the degree (p), p a proposition”. In (70) I follow Hale and Keyser’s traditional analysis (1993) of the lexical structure of unergative verbs like *talk*. In the case of attributive predicates of aspectual verbs as in (71a), the relevant structure would be (71b):

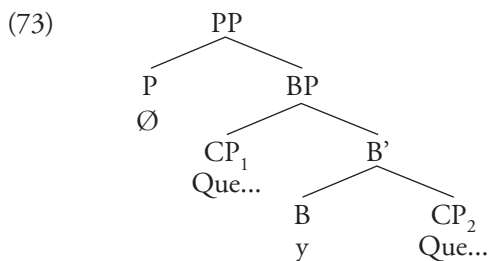
- (71) a. Tu padre está que no puede más
 your father is that neg he-can more
 ‘Your father is in such a state that he can’t do more’
 b. ..._{[VP} BE _{[PP} *tu padre* P_{central} _{[CP} [Degree P] *que* _{[SC} CP (DegreeP)]]]]]]

With the subject generated in the Spec of the aspectual projection (see Hale and Keyser 2005). BE is spelled out as *estar* ‘to be in a location’ when it selects [+central coincidence] preposition (for the aspectual properties of the *ser/estar* distinction in Spanish, see Schmitt 1996).

The structures in (70-71) immediately account for an important property of bare CP-modifiers: their invariable clause-final position, which follows from being a complement of the silent preposition, which is either the modifier of a low aspectual projection (70), or an attributive predicate (71). They also provide us with structural means to account for another intriguing syntactic property of consecutive clauses: they cannot be conjoined:

- (72) Tu padre habla que no calla (*y que aburre a los demás)
 your father talks that neg shuts-up and that he-bores to the rest
 ‘Your father talks in such a way that he will not shut up and that he bores the rest of the people’

If the merging of the null P with CP is automatically followed by conflation (copying of the phonetic matrix of the selected head, Hale and Keyser 2002), and conflation requires a head-head relation, then we can explain why degree-clauses here can not be conjoined. The coordination of the degree clause under a boolean projection breaks up the head-government configuration between the null preposition and *que* required for conflation, and the null preposition can not be phonologically supported:



4.4. The quotative dependents

The option of inserting a null aspectual preposition in Spanish also provides a solution for the way in which quotative dependents relate to the verbal structure. Full quotatives would be represented as in (74), with P a silent aspectual preposition with the value [+central coincidence] and BE an auxiliary:

- (74) a. Tu padre está que cuándo vamos
 your father is that when we-are-going
 ‘Your father is saying: “when are you coming?”’
 b. ..._{[VP} BE _{[PP} *tu padre* P_{central} _{[CP} que cuándo vamos]]]]

That this option is not purely speculative is shown by the existence of parallel quotative constructions which do present an overt central coincidence preposition:

- (75) a. Tu siempre con que es demasiado tarde para comer
 you always with that it-is too late for lunch
 ‘You always with this story that it is too late for lunch’
 b. Tu siempre con que cuándo vamos
 you always with that when we are going
 ‘You always with this question of when we are going’
 c. Tu siempre con que qué sano es hacer footing
 you always with that how healthy is to do jogging
 ‘You always with this exclamation that how healthy is to go for a jog’
 d. Tú siempre con que no vaya allí
 you always with that neg I-go-subj there
 ‘You always with this order that I should not go there’

Like our simple quotative constructions, *con* necessarily requires a saying as part of the interpretation of its clausal dependent. In order for the utterance of (75a) to be faithful to the facts, the subject to which the utterance makes reference must say (and not just believe or think), that it is too late for lunch. *Con* is impossible with manner modifiers not expressing the content of a speech event:

- (76) Juan hablaba (*con) que no callaba
 Juan spoke with that neg he-shut-up
 ‘Juan spoke in such a way that he would not shut up’

And it shows the same quantificational restrictions as simple quotatives:

- (77) *Tu nunca/rara vez/alguna vez con que cuando voy
 you never/rarely/sometime with that when I-am-going
 ‘You are always with this saying: “when are you going?”’

Unlike simple quotatives, though, this construction cannot represent a punctual speech event: it conveys the idea that the speech eventuality repeats itself very often. Consider the following situation, in which a punctual reading is enforced:

- (78) (Context: Jon has just called suggesting going out for a drink. A holds up the phone and asks his/her partner)
 A: Jon (*con) que si queremos salir a tomar algo. Qué le digo?
 Jon with that if we want to go out for a drink. What should I say
 ‘Jon is asking whether we feel like going out for a drink. What should I say?’

Note that in the absence of an overt aspectual auxiliary, *con que* dependents require modification by a quantifier like *siempre* ‘always’, expressing frequency or habituality. It thus seems that the overt preposition differs from the null one in its aspectual properties: it expresses habitual or frequentative aspect. Let me therefore conclude that the aspectual preposition has two possible realizations: a null one, expressing central coincidence (and operative also in manner-CP constructions), and an overt one, expressing central coincidence and habituality/frequency.

The reduced quotative would be represented as in (79), with an abstract intransitive verb that supports event modification (see section 2):

- (79) a. Tu padre que cuándo vienes
 b. ..._{[VP} Tu padre GO _{[PP} P_{central} _{[CP} que cuándo vamos]]]

(79) adopts Travis' idea that the relevant aspectual head is actually below the higher lexical verb (Travis' inner aspect head, 2006).

4.5. The saying event

We were not explicit about how a saying is expressed in the relevant configurations. This saying must be somehow represented in the syntactic structure of quotative constructions, since it is semantically required by them. As in the case of degree-clauses, I will take *que* to introduce a predication relation: one that relates a CP with, possibly, features of main clauses (semi-direct speech), and a nominal predicate. This nominal predicate is an indefinite description, whose only lexical feature is [linguistic communication], and which is interpreted as an existential quantification over utterances (Lahiri 2002). As in the case of degree-clauses, the result is a complex noun phrase construction, à la Stowell (1981):

- (80) ..._{[CP} que _{[SC} CP DP_{LC}]

The predicate raises to Spec of CP:

- (81) ..._{[CP} DP_{LC} que _{[SC} CP (DP_{LC})]

Yielding the complex noun phrase « a saying that CP » (see also Kayne 1994). This is semantically an event description, and as such can enter into aspectual relations via aspectual prepositions. The central coincidence prepositions P_{null} and *con* “with” relate the time of the saying event to a subpart of that time, yielding a reading akin to a progressive:

- (82) ...P_{central} _{[CP} DP_{LC} que _{[SC} CP (DP_{LC})]

This aspectual projection is selected by a copula:

- (83) ..._{[VP} BE _{[PP} Subject P_{central} _{[CP} DP_{LC} que _{[SC} CP (DP_{LC})]]]

The copula may be spelled out as it is (*estar*) or may merge to a higher frequentative aspect head, yielding the aspectual auxiliary verb *andar* “walk”.

Under this view, minimal quotatives are just low aspectual phrases, phrases containing neither Tense nor higher aspectual auxiliaries. Just aspectual PPs:

- (84) a. Tu padre con que cuándo vienes
 b. _{[PP} Tu padre P_{central} _{[CP} DP_{LC} que _{[SC} cuándo vienes (DP_{LC})]]]

5. A note on the anchoring of the quotative construction

An aspect of minimal quotative constructions that becomes manifest to anyone working on these cases is their dependency on discourse particles, vocatives or loca-

tional adverbs. Something like (13a), repeated below, becomes extremely odd without the presence of *aquí* “here”:

- (85) *(*Aquí*) todo dios que cuándo les van a subir el sueldo
 here everyone that when cl they-are-going to raise the salary
 ‘Everyone is saying: “When are they going to give us a raise?”’

In the same way, discourse particles like *oye* “hey” or vocatives like *Juan* below seem to be necessary in the absence on any overt event participant:

- (86) *(*Oye/Juan*), que el Barça ha ganado la Champions
 listen that the Barça has won the Champions League
 ‘Hey, there’s a saying that Barça has won the Champions League’

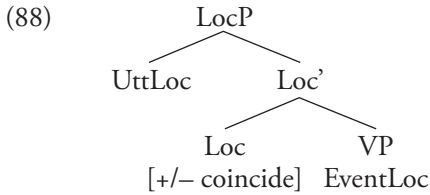
The conditions under which the different options are put into use are not yet clear, but they all seem to be related to the necessity of anchoring the report in the discourse. This anchoring requirement disappears with full fledged quotative constructions containing finite auxiliaries:

- (87) Todo dios anda que cuándo les van a subir el sueldo
 everyone walks that when cl they-are-going to raise the pay
 ‘Everyone is continuously saying: “when are they going to give us a raise?”’

Anchoring particles are therefore obligatory in cases when Tense is absent. Normally, Tense is the means by which an event is anchored to the utterance or some other salient point (the notion of *Anchoring Condition*, Enç 1987). But minimal quotatives may not have Tense (that is, they may contain no Tense morphology or show no temporal adverbial modifiers). How is anchoring effected in those cases? Ritter and Wiltschko (2005) have recently asked that question in the context of languages which seem not to have grammatical Tense. Their answer is that in those languages anchoring proceeds either spatially, via the syntactic category Location, or via speech act participants. The categories involved, as one can see, are very similar to the anchoring elements in the absence of Tense in Spanish quotatives: the vocative and the discourse particle *oye* (literally “listen”) are hearer-oriented elements, and belong in the structure of the utterance, rather than in the reported event¹⁰. On the other hand *aquí* “here” is a locative demonstrative.

Wiltschko and Ritter claim that in languages lacking grammatical Tense, the event is anchored in the utterance by expressing where it happened, instead of when it happened. Anchoring is driven by a category Location, which, following Demirdache and Uribe-Etxebarria (2000) they take to be a dyadic predicate expressing [+/- central coincidence]. This dyadic predicate combines the utterance location and the event location in a predication relation. A sentence can assert that the event location coincides with the utterance location (the event happens *here*) or that the event does not coincide with the utterance location (the event happens *there*):

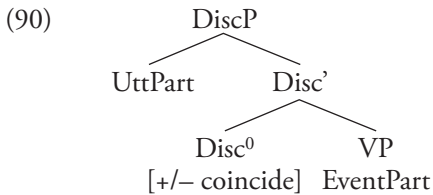
¹⁰ For instance, they are exempt from the truth functional evaluation of the sentence, unlike other addressee oriented elements such as second person pronouns.



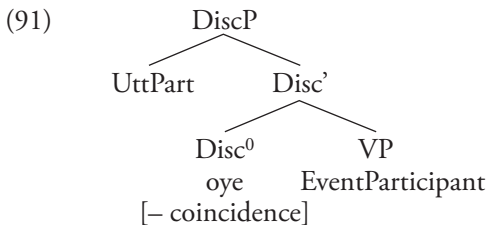
The presence of locative adverbials, like the obligatory *aquí* “here” in (85) signals the presence of a syntactic projection whose function is anchoring the reporting event in the utterance:



Anchoring can also be driven via speech act participants. Wiltschko and Ritter claim that this is the case in Blackfoot. In Blackfoot event anchoring proceeds via participants of the utterance and the event. In this anchoring system, it is asserted who participated in the event with respect to who participated in the utterance. This relation is achieved by means of a different predicate of (non-)coincidence, that they call Discourse. If Discourse is a predicate of coincidence, the event participant is asserted to coincide with the utterance participant (e.g. actor=speaker). If Discourse is a predicate of non-coincidence then the event participant is asserted to not coincide with the utterance participant:



The choice of the anchoring participant in Blackfoot is driven by person morphology and theme marker morphology. The combination of those two morphological parameters gives rise to a rich and complex anchoring system. Spanish does not possess a theme-marker system, and in the cases we are considering, person morphology is absent. The anchoring system therefore must be much simpler. Let me advance the hypothesis that in Spanish, the anchoring speech act participant is always the speaker. *Oye* is the head of the Discourse Phrase. As it does not represent the speaker (it is a hearer oriented particle) it expresses a relation of non-coincidence between the event participants and the speech act participants:



This hypothesis is supported by a peculiar restriction of quotative constructions in Spanish: they don't admit self-reports with first person event participants (that is, cases where the speaker/actor and the utterer are the same person).

- (92) *Juan/oye, yo que cuándo vamos
 Juan/hey I that when we-are-going
 'I say: I am not going'

(92) contrasts sharply with the acceptable complete quotative construction:

- (93) Yo (siempre) estoy que cuándo vamos
 I always am that when we-are-going
 'I am (always) saying: when are we going?'

And with those quotative constructions which show overt temporal modification:

- (94) Ya verás, yo *(mañana) que cuándo vamos
 you'll see, I tomorrow that when we-are-going
 'You'll see, tomorrow I will be saying: 'when we are going?''

In both (93) and (94) anchoring is effected by Tense, and the discourse head does not need to be projected.

6. The absence of Person and Tense morphology in reduced quotatives

Consider the following two types of quotative construction, which I called "full" (95a) and reduced (95b):

- (95) a. Tu padre que cuándo vienes
 Your father that when you-are-coming
 'Your father is saying: "when are you coming?"'
 b. Tu padre está que cuándo vienes
 your father is that when you-are-coming
 'Your father is saying continuously: "when are you coming?"'

(95a) supports Tense modification, but cannot support overt Tense or Person morphology. We may wonder why. At the same time, (95b) but not (95a) can be extended by features typical of the left periphery, such as topic, focus, or sentence adverbial:

- (96) a. TU PADRE está que cuándo viene, no tu madre
 Your father is that when he-is-coming, neg your mother
 'It is your father who is saying: "when is he coming?", not your mother'
 b. En cuanto al coche, tu padre está que cuándo lo va a vender
 as for the car your father is that when cl you-are-going to sell
 'As for the car, your father is saying: "when is he going to sell it?"'
 c. Desgraciadamente, tu padre anda que cuándo se irán
 unfortunately, your father walks that when refl they-will-leave
 'Unfortunately, your father keeps saying: "when are they going to leave?"'

None of that is possible in “reduced” quotative constructions (see section 3). One way of looking at those related restrictions (absence of Tense and Person morphology and absence of left-peripheral projections) is under the following generalization: if the construction has an auxiliary which can independently pick up Tense and Person morphology, we have a full clause; if we don’t have an auxiliary, the embedded quotative verbal structure does not seem able to support such morphology, and all the left periphery disappears. In those cases we are left with bare TPs.

Let us consider again the syntactic structure of reduced quotatives. They would go along the following lines:

(97) ...v_{GO} [_{PP/AspP} P_{Asp} [a saying *que* [[Force P] (a saying)]]]]

P and v are null in Spanish. By conflation (see section 4.3), the null P and the null v will get lexicalized by the complementizer *que*, a nominal category. Now: there is no morphological item in Spanish which would possess the following morphological structure:

(98) *_[INFL] INFL [*que*]

In other words, combining the Tense/Person heads with *que* yields an impossible morphological object in Spanish. I would like to relate this to the otherwise intriguing fact that the absence of such morphology goes hand in hand with the absence of left-peripheral elements. Note that T itself is semantically present, as it supports temporal modification. I would like to claim that the way Spanish circumvents the impossible morphological object in (98) is by not spelling out the whole clause. If morphology is checked independently in a post-syntactic module, as Distributed Morphology wants, then the ungrammaticality of (98) is strictly a morphological phenomenon. The correct configuration of morphological words is checked at Spell Out. If you don’t spell out, then existence of configurations such as (98) depends on strict syntactic motivations, and those, we saw, argue for the underlying presence of sequences of T and lexicalized v/P. Assuming a version of Phases which has C as a strong Phase (Chomsky 2001), the only way of not spelling out (98) is by not ever getting to C (more precisely, the extended C domain), hence the absence of left-peripheral elements in reduced quotatives.

References

- Alarcos, E., 1982, «Español *que*». In E. Alarcos, *Estudios de Gramática Funcional del Español*, 260-274. Madrid:Gredos.
- Álvarez, A., 1999, «Las construcciones consecutivas». In *Gramática Descriptiva de la Lengua Española*, volume 3, ed. by I. Bosque and V. Demonte, 3.739-3.804. Madrid: Real Academia Española.
- Amberber, M., 1996, «The transitivity of verbs of saying revisited». *The Proceedings of the Fourteenth West Coast Conference on Formal Linguistics*. J. Camacho, L. Choueiri and M. Watanabe (eds.), Stanford: CSLI Publications. 1-15.
- 1996, *Transitivity Alternations, Event-types and Light Verbs*. Ph.D. diss, McGill University. Canada: Quebec.

- 2002, *Verb Classes and Transitivity in Amharic*. Lincom Studies in Afro-Asiatic Languages 6. Munich: Lincom.
- Asher, N., 1993, *Reference to Abstract Objects in Discourse*. Kluwer Academic Publishers.
- Baker, M., 1996, «On the structural position of Themes and Goals». J. Roryck and L. Zaring (eds.), *Phrase Structure and the Lexicon*. Dordrecht: Kluwer Academic Publishers.7-39.
- Beck, S., 1996, «Quantified Structures as Barriers for LF Movement», *Natural Language Semantics* 4: 1-56.
- Boer, S. E. and W. G. Lycan, 1980, «A performadox in truth-conditional semantics», *Linguistics and Philosophy* 4-1:71-100.
- Borer, H., 2005, *The Normal Course of Events. Structuring Sense*, volume II. Oxford: Oxford U. P.
- Brucart, J.-M., 1992, «Sobre la estructura de SCOMP en Español». In *Sintaxi: Teoria I Perspectives*, ed. by A. Viana, 59-102. Barcelona: Pages Editors.
- Bybee, J., W. Pagliuca and R. Perkins, 1994, *The evolution of Grammar: Tense, Aspect and Modality in the languages of the world*. Chicago: The University of Chicago Press.
- Campos, H., 1992, «Enunciative elements in Gascon», *Linguistics* 30: 911-940.
- and M. Zampini, 1990, «Focalization strategies in Spanish», *Probus* 2-1: 47-64.
- Chafe, W. L. and J. Nichols (eds.), 1986, «Evidentiality: the Linguistic Coding of Epistemology», *Advances in Discourse Processes* 20. Norwood, NJ: Ablex.
- Chierchia, G., 1993, «Questions with quantifiers», *Natural Language Semantics* 1: 181-234.
- Chomsky, N., 1995, *The Minimalist Program*, MIT Press.
- Cinque, G., 1977, «The Movement Nature of Left Dislocation», *LI* 8: 397-411.
- 1999, *Adverbs and Functional Heads*. Oxford U. P.
- Collins, C., 1997, *Local Economy*. Massachusetts: MIT Press [Linguistic Inquiry Monograph 29].
- Coulmas, F., 1986, *Direct and Indirect Speech*. Berlin: Mouton de Gruyter.
- D'Arcy, A. F., 2005, «Like». *Syntax and Development*. University of Toronto Ph.D dissertation.
- Delbecque, N. and B. Lamiroy, 1999, «La subordinación sustantiva: Las subordinadas enunciativas en los complementos verbales». In *Gramática Descriptiva de la Lengua Española*, volume 2, ed. by I. Bosque and V. Demonte, 1.965-2.082. Madrid: Real Academia Española.
- Demirdache, H. and M. Uribe-Etxebarria, 2000, «The primitive of temporal relations». In *Step by step: Minimalist Essays in Honor of Howard Lasnik*, ed. by R. Martin, D. Michaels and J. Uriagereka, 157-175. Cambridge, MA: MIT Press.
- 2002, «La grammaire des predicats spatio-temporels: Temps, aspect et adverbos de temps». In *Temps et Aspect : De la morphologie à l'interpretation*, ed. by B. Laca, 125-175. Paris: Presses Universitaires de Vincennes.
- 2004, «The Syntax of Time Adverbs». In *The Syntax of Time*, ed. by J. Guéron and J. Lecarme, 143-180. Cambridge, MA: MIT Press.
- Demonte, V. and P. J. Masullo, 1999, «La predicación: Los complementos predicativos». In *Gramática Descriptiva de la Lengua Española*, volume 2, ed. by I. Bosque and V. Demonte, 2.461-2.524. Madrid: Real Academia Española.
- Dendale, P. and L. Tasmowski, 2001, «Introduction. Evidentiality and Related Notions», *Journal of Pragmatics* 33-3: 339-348.

- De Roeck, M., 1994, «A functional typology of speech reports». *Function and Expression in Functional Grammar*. E. Engberg-Pedersen, L. Falster Jakobsen and L. Schack Rasmussen (eds.). Berlin: Mouton de Gruyter. 331-351.
- Ernst, T., 2002, *The Syntax of Adverbs*. Cambridge: Cambridge U. P.
- Ettxepare, R., 1998, *On the grammatical representation of speech events*. Ph.D. diss, University of Maryland.
- and M. Uribe-Etxebarria, 2005, «In-situ wh-phrases in Spanish: Locality and Quantification», *Recherches Linguistiques de Vincennes* 33: 9-34.
- (to appear), «Quotative constructions in Iberian Spanish». In *The Pragmatics of Clause Combining*, ed. By R. Laurie and E.-L. Seepanen. John Benjamins.
- Frajzyngier, Z., 1996, «Grammaticalization of the Complex Sentence. A Case Study in Chadic». Amsterdam: John Benjamins [*Studies in Language Companion*, series 32].
- García, E., 1996, «¿Cómo que qué?», *Hispanic Linguistics* 8: 59-93.
- Givón, T., 1982, «Evidentiality and epistemic space», *Studies in Language* 6-1: 23-49.
- Güldemann, T., 2001, *Quotative Constructions in African Languages: a Synchronic and Diachronic Survey*, Habilitationsschrift. Leipzig University.
- Heine, B. and T. Kuteva, 2002, *World Lexicon of Grammaticalization*. Cambridge U. P.
- Hale, K., 1986, «Notes on world view and semantic categories: Some Warlpiri examples». *Features and Projections*. P. Muysken and H. van Riemsdijk (eds.), Dordrecht: Foris. 233-254.
- and S. J. Keyser, 1993, «On argument structure and the lexical expression of syntactic relations». *The View from Building 20. Essays in Linguistics in Honor of Sylvain Bromberger*, K. Hale and J. Keyser (eds.), Massachusetts: MIT Press. 53-110 [*Current Studies in Linguistics* 24]
- 2002, *Prolegomenon to a theory of argument structure*, Massachusetts: MIT Press. Linguistic Inquiry Monograph 39.
- 2005, «Aspect and the Syntax of Argument Structure». In *The Syntax of Aspect*, ed. by Nomi Esterschik-Shir and Tova Rapoport. Oxford U. P. 11-41 [Oxford Studies in Theoretical Linguistics].
- Harley, H., 1995, *Subjects, Events and Licensing*. Ph.D. dissertation, MIT, Massachusetts.
- Hernández Alonso, C., 1967, «El que español», *RFE* 50, 257-271.
- Honcoop, M., 1998, *Dynamic Excursions on Weak Islands*. The Hague: Holland Academic Graphics.
- Kayne, R., 1994, «The Antisymmetry of Syntax». Massachusetts: MIT Press [Linguistic Inquiry Monographs 25].
- Keenan, E. and L. Faltz, 1985, *Boolean Semantics for Natural Language*. Dordrecht: Reidel.
- Klamer, M., 2000, «How report verbs become quote markers and complementisers», *Lingua* 110, 69-98.
- Kratzer, A., 1996, «Severing the external argument from the verb». *Phrase Structure and the Lexicon*, J. Rooryck and L. Zaring (eds.), 109-138. Dordrecht: Kluwer Academic Publishers.
- Krifka, M., 2001, «Quantifying into question acts», *Natural Language Semantics* 9-1, 1-40.
- , 2003, «Quantifiers in questions», *Korean Journal of English Language and Linguistics* 3: 449-526.
- Lahiri, U., 1991, *On embedded interrogatives and predicates that embed them*, Ph.D. diss, MIT, Massachusetts.
- 2002, *Questions and Answers in Embedded Contexts*, Oxford U. P.

- Larson, R. K., 1988, «On the double object construction», *LI* 1-9: 335-391.
- Lenci, A. and P. M. Bertinetto, 2000, «Aspects, Adverbs and Events: Habituality vs. Perfectivity». *Speaking of Events*. J. Higginbotham, F. Pianesi and A. Varzi (eds.), 245-288. Oxford: Oxford U. P.
- Levin, B. and M. Rappaport, 1998, «Building Verb Meanings». *The Projection of Arguments*. M. Butt and W. Geuder (eds.), Stanford: CSLI Publications. 97-134.
- Levinson, S., 1983, *Pragmatics*. Cambridge: Cambridge U. P.
- Lord, C., 1993, *Historical Change in Serial Verb Constructions*. John Benjamins.
- Macaulay, R., 2001, «You're like "Why not?" the Quotative Expressions of Glasgow Adolescents», *Journal of Sociolinguistics* 5-1: 3-21.
- Mathieu, E., 2002, *The Syntax of Non-Canonical Quantification*. Ph.D Dissertation. University College, London.
- Munro, P., 1982, «On the transitivity of "say" verbs». *Studies in Transitivity*. P. Hopper and S. Thompson (eds.), *Syntax and Semantics* 15. New York: Academic Press. 301-318.
- Pafel, J., 1993, «Scope and Word Order». *Syntax. An international handbook of contemporary research*, volume 1. Jacobs, J. et alii (eds.), Berlin: Mouton. 867-879.
- Parsons, T., 1990, «Events in the semantics of English. A Stud in subatomic semantics». Massachusetts: MIT Press [Current Studies in Linguistics 19].
- Pesetsky, D., 2000, *Phrasal Movement and Its Kin*. Cambridge: MIT Press.
- Picallo, C., 2002, «Abstract Agreement and Clausal Arguments», *Syntax* 5: 116-147.
- Plann, S., 1982, «Indirect questions in Spanish», *LI* 13: 297-312.
- Porroche Ballesteros, M., 1995, «Algunos aspectos del uso de *que* en el español conversacional (*que* como introduccion de oraciones "independientes")». *Atti del XXI Congresso Internazionale di Linguistica e Filologia Romanza. Centro di studi filologici e linguistici siciliani. Università di Palermo 18-24 settembre 1995*, volume II, G. Rufino (ed.), 245-255.
- Pustejovsky, J., 1991, «The syntax of event structure», *Cognition* 41: 47-81.
- 1995, *The Generative Lexicon*. Massachusetts: MIT Press.
- Ritter, E. and S. Rosen, 1998, «Delimiting Events in Syntax». *The Projection of Arguments*. Miriam Butt and Wilhelm Geuder (eds.). Stanford: CSLI Publications. 135-164.
- and M. Wiltschko, 2005, «Anchoring events to Utterances without Tense», *Proceedings of the 24th West Coast Conference on Formal Linguistics*. J. Alderete, Chung-Hye-Han and A. Kochetov (eds.), Somerset, MA: Cascadilla Press.
- Rizzi, L., 1997, «The fine structure of the left periphery». *Elements of Grammar. Handbook of Generative Syntax*. L. Haegeman (ed.), Dordrecht: Kluwer Academic Publishers. 281-338.
- Rohlf, G., 1977, *Le Gascon. Études de philologie pyrénéenne*. Tübingen and Pau: Niemeyer Verlag-Éditions Marrimpouey Jeune.
- Rooryck, J., 2001a, «Evidentiality I», *Glott International* 5-4: 125-133.
- 2001b, «Evidentiality II». *Glott International* 5-5: 161-168.
- Ross, J. R., 1970, «On declarative sentences». *Readings in English Transformational Grammar*. R. A. Jacobs and P. S. Rosenbaum (eds.). Waltham: Ginn.
- Schmitt, C., 1996, *Aspect and the Syntax of Noun Phrases*. Ph.D. dissertation, University of Maryland.
- Shlonsky, U., 2000, «Remarks on the Complementizer layer of Standard Arabic». *Research in Afroasiatic Grammar. Papers from the Third Conference on Afroasiatic Languages. Sophia Antipolis. 1996*. J. Lecarme, J. Lowenstamm et U. Shlonsky (eds.), Amsterdam: John Benjamins.

- Smith, C., 1991, *The parameter of Aspect*. Kluwer Academic Publishers.
- Spitzer, L., 1942, «Notas sintáctico-estilísticas a propósito del español *que*», *Revista de Filología Hispánica* 4: 105-126.
- Stowell, T., 1982, *Origins of Phrase Structure*. Ph.D.dissertation, MIT, Massachusetts.
- 1996, «The Phrase Structure of Tense. Phrase Structure and the Lexicon». J. Rooryck and L. Zaring (eds.). Dordrecht:Kluwer Academic Publishers.
- Suñer, M., 1993, «About indirect questions», *Linguistics and Philosophy* 16: 45-77.
- Talmy, L., 1983, «How language structures space». *Spatial Orientation: Theory, research, and application*. H. L. Pick, Jr. and L. P. Acredolo (eds.), New York: Plenum Press. 225-282.
- Tenny, C., 1994, *Aspectual Roles and the Syntax-Semantics Interface*. Dordrecht: Kluwer Academic Publishers.
- , 2000, «Core events and adverbial modification». *Events as Grammatical Objects*. C. Tenny and J. Pustejovsky (eds.), Stanford: CSLI Publications. 285-334.
- Torrego, E., 1984, «On inversion in Spanish and some of its effects», *LI* 15: 75-102.
- Travis, L., 2000, «Event structure in syntax». *Events as grammatical objects*. C. Tenny and J. Pustejovsky (eds.). Stanford: CSLI Publications. 145-185.
- Uriagereka, J., 1995, «An F position in Romance». *Discourse Configurational Languages*. K. Kiss (ed.), Oxford: Oxford U. P.
- Zagona, K., 1995, «Temporal Argument Structure: Configurational Elements of Construal». *Temporal Reference, Aspect and Actionality, vol 1: Semantic and Syntactic Perspectives*. P. M. Bertinetto, V. Bianchi, J. Higginbotham and M. Squartini (eds.), Turin: Rosenberg and Sellier.
- Zagona, K., 2003, «Tense and Anaphora. Is There a Tense-specific Theory of Coreference?». *Anaphora: A Reference Guide*. A. Barss (ed.), Blackwell: 140-171.
- Zwicky, A., 1971, «In a manner of speaking», *LI* 2: 223-232.

COMPLEMENTS AND ADJUNCTS IN MACHINE TRANSLATION

Kata Gàbor, Enikő Héja

Eötvös Loránd University and Hungarian Academy of Sciences

Abstract

A significant number of natural language processing applications cannot work without syntactic parsing. The automatic syntactic analysis of natural language texts in turn requires an efficient method for differentiating between elements that belong to the predicate's argument structure and those that are attached to it as adjuncts. The focus of our paper is a specific method we are working on for differentiating between verbal complements and adjuncts, which we intend to use for the elaboration of a Hungarian verbal argument structure database, particularly suited for machine translation purposes.

1. Introduction

Both linguistic theories and rule-based natural language processing applications rely on a strict differentiation between verbal complements (elements that figure in the subcategorization frame of the verb's lexical entry) and adjuncts (elements that are optionally added to the verb (phrase) by syntactic rules). Although adjuncts are optional and hence their appearance is not predictable, the possibility to extend a verbal structure by an adjunct is predictable. In opposition, the behavior of complements is not predictable by general syntactic rules of a given language, this is why they are widely conceived as lexical properties of verbs and they are treated in the lexicon. Consequently, Natural Language Processing (NLP) applications that involve syntactic parsing of texts need to use a lexical database of verbal argument structures which describe all the relevant properties of every single verb's arguments. However, for the database to be coherent and homogeneous, coders need to be given exact and explicit instructions about what a complement is. This boils down to our basic question: what could be the method for making the difference between complements and adjuncts?

We examined two linguistic theories: Government and Binding theory (GB; on the basis of Radford 1988) and Lexical Functional Grammar (LFG; on the basis of Komlósy 2001) with respect to how they describe verbal argument structure and the way arguments are represented in the surface structure of natural languages. The most significant difference between these theories is that GB proposes a configurational model of natural languages, i.e. it encodes constituents' grammatical functions

by dominance and precedence relations in the tree structure, while in LFG grammatical functions are coded in a separate level of representation which does not prescribe their possible surface representation. The importance of this difference is that in several languages complementness and syntactic functions are reflected not as much in surface constituent order as in morphological properties. In Hungarian configurationality is used to express discourse functions instead of syntactic functions, thus we would predict that an LFG analysis would describe better complementness criteria in Hungarian. On the other hand, both theories agree that complementness is a relational notion: a given element can only be the complement of a governing element, but not in itself.

2. The role of complements in machine translation systems

Among current theoretical and methodological approaches to machine translation (MT) two main branches can be distinguished (Jurafsky and Martin 2000): *rule-based* and *statistical/corpus-based* systems. Rule-based systems use linguistic knowledge: they contain one or more modules which analyze source language text units at several linguistic levels, and rules map the output of the source language analysis to the target language or to an intermediate representation.

The main advantage of rule-based MT systems as opposed to statistical ones is that they are more easily maintainable: to find the source of an incorrect translation is relatively trivial in a well-designed rule-based system, while it can be very complicated in a statistical one.

Rule-based MT systems can be subdivided into three types:

- 1) *direct transfer*
- 2) *transfer-based translation*
- 3) *interlingual translation*

As its name implies, direct transfer is a simple method based on the supposed similarity between closely related languages: it makes wide use of bilingual dictionaries, but does not direct much attention to structural differences. The role of grammatical rules for source language analysis and translation is marginal: they mainly serve as disambiguation rules.

Transfer-based systems analyze the source language text at both morphological and syntactic levels by monolingual rules and databases, and use the so-called *transfer rules* to map the output of the analysis into the target language. The final stage of the translation is the set of monolingual target language grammar rules which correct the output of the transfer phase. The key module of the process is the bilingual *transfer module* which is composed of the bilingual dictionary and the transfer rules. These rules carry out the task of mapping grammatical characteristics of the source language into the target language. Hence, this module is totally specific to language pairs as it only deals with phenomena that differ across the given language pair.

As opposed to transfer-based translation, *interlingual* systems project the source language text into an intermediate representation which is a language-independent structure intended for outlining the information contained in the sentence as well as its logical structure. Target language equivalents of the sentence are then calcu-

lated from the intermediate representation. While transfer-based systems attain the target language translation of a sentence by means of transforming the elements and the structure of the original source language text, interlingual systems aim at extracting the meaning of the source text and produce a target language text with the same meaning.

Both transfer-based and interlingual MT systems rely strongly on the syntactic parsing of the source language, all the more because it plays an important role in disambiguation. Most systems lay emphasis on setting apart lexical information and general sentence formation rules. The reason for it is the assumption that while the translation of lexical information is unpredictable, regular phenomena can be translated by rules to another natural language (in transfer-based systems) or to an intermediate representation (in interlingual systems). This distinction applies to verbal subcategorization and adjunction. On the one hand, syntactic behavior of the elements which fulfill complement or adjunct functions can be predicted within a given language and translated by rules: e. g. we can state that in Hungarian the top-level NP constituent in nominative case will be the subject of the clause, and construct a rule which translates it into English by moving this NP before the verb. On the other hand, whether a given Hungarian verb *can have* a subject and whether it will keep this function throughout translation is a piece of unpredictable, though important information which has to be coded.

3. Tests for complementness

We made a comparison between a configurational and a lexicalist linguistic theory (Government and Binding theory and Lexical Functional Grammar, respectively) with respect to how they describe verbal argument structure and how they represent arguments in the surface structure of natural languages.

3.1. Complements and subcategorization in GB

Government and Binding theory defines complements as constituents which compulsorily appear in the close local context of the verb. Their syntactic behavior cannot be described by general phrase-structure rules as their appearance is not predictable. The reason for this is that complementness is conceived as a relation: constituents which fulfill a complement function in a sentence with respect to its predicate cannot have this same function in other sentences. Predicates' ability to take complements is their idiosyncratic lexical property. Consequently, lexical entries of verbs have to contain as much information as necessary for the syntactic rules to generate surface forms of complements. Hence, lexical entries of verbs comprise syntactic description (i.e. the category) of their complements. Moreover, since syntactic complements are surface representations of semantic arguments, it is worth coding the thematic roles of semantic arguments in the lexicon as it allows certain generalizations over the surface representation of semantic arguments: a part of the complement structure can be derived from thematic roles.

3.2. Complement tests in GB

According to X-bar theory, if we want to test whether a given constituent is a complement or an adjunct we have to examine its structural position. While complements are located within the syntactic tree in a sister node of the X (verbal) head and together they form an X' projection, adjuncts are sisters of the X' projection and form a new X' with it. The position that adjuncts and complements occupy in the syntactic tree is universal among languages, but their surface order in relation to the head is language-specific. Unfortunately, this implies that we can only rely on language-specific tests for verifying the different structural position of given constituents. Radford (1988) mentions four tests for English:

- a) *Passivization*: NPs raised from a complement PP can be passivised while NPs from an adjunct PP cannot:
 - [This job] needs to be worked at by an expert.
 - *[This office] is worked at by a lot of people.
- b) *Pronominalization*: the *do so* structure, which replaces the category V', can include adjuncts that are attached to a V' to form a new V' projection with it (i), but adjuncts can also be omitted from it (ii), while complements are compulsorily included (iii), they cannot be omitted as in (iv).
 - i) John will [buy the book on Tuesday] and Paul will *do so* as well.
 - ii) John will [buy the book] on Tuesday and Paul will *do so* on Thursday.
 - iii) John will [put the book on the table] and Paul will *do so* as well.
 - iv) *John will [put the book] on the table and Paul will *do so* on the chair.
- c) *Surface order*: Complements are closer to the verb than adjuncts because they connect to the verb in the syntactic tree earlier than adjuncts, and crossing branches are forbidden.
- d) *Ellipsis*: Any phrasal category can be subject to ellipsis. Constituents of the category of V' can be ellipsed if they consist of the verbal head with its complements and adjuncts (i), the head with its complements but without adjuncts (ii), but the head with one of its complements and without the other one does not form a constituent, hence it cannot be ellipsed (iii):
 - i) — Who might be going to the cinema on Tuesday?
— *John might be ...*
 - ii) — Who might be going to the cinema when?
— *John might be ... on Tuesday.*
 - iii) — Who will put the book where?
 - iv) **John will ... on the table.*

3.3. Complements and subcategorization in LFG

In the LFG model different structural levels of the sentence carry the same functional information but are represented at different levels. However, information about grammatical function is present at every level of representation. Accordingly, grammatical information is represented at three levels:

1. a-structure stores lexical information, i. e. argument structure;
2. c-structure stores surface constituent structure;
3. f-structure represents the language-independent functional structure which can be extracted from the two language-specific representations.

Surface structure is not an independent level of representation: it is generated from c-structure by inserting lexical elements.

Argument structure and other lexical information are stored in a-structure. Arguments are conceived as ‘unfilled’ slots in the meaning of the verb. Accordingly, the bare a-structure comprises semantic arguments of the verb with their thematic roles. Surface representation of semantic arguments depends on the grammatical functions associated with them. The first step in the process of mapping argument structure to surface complements is the annotation of bare lexical structure. Annotation assigns a function to semantic arguments. In LFG verbal subcategorization frames do not contain categorial information about complements, they only refer to their grammatical function. Correspondences between semantic argument positions and grammatical functions are coded in the annotated lexical structure —argument roles cannot be bound to universal structural positions (as in GB, where subjects and direct objects are assigned their function on the basis of their structural positions).

The set of functions verbs can prescribe for their complements is restricted. There are complement and adjunct functions. Among complement functions, the most interesting distinction is between *thematically bound oblique complements* and *labeled complements*. Thematically bound oblique complements are those complements whose thematic role is determined by the verb, but whose syntactic functor is not. One typical example is constituted by locative complements. On the other hand, in the case of labeled complements, not only the thematic role but also the exact form of the syntactic functor (e.g. its case suffix or preposition) is prescribed by the verb.

Surface representation of constituents with complement functions is generated by language-specific context free rules of the c-structure. Surface representation of grammatical functions may be coded either configurationally or by other (typically morphological) means, and this divergence may be present within one language. In the case of configurationally coded functions, the LFG variant of X-bar theory conditions the construction of the tree structure, while in the case of non-configurational coding, complement functions are associated not to the c-structure but to case-marking and agreement properties. The way functions are associated to case-bearing constituents is by functional annotation schemes realized as implications: “*a constituent X may be associated with a function Y if it bears case Z.*” On the other hand, agreement is handled by *head marking*: the constituent’s relevant AGR features are checked, and the constituent is associated with the given function if the value of the features equals those prescribed by the functional structure of the constituent which dominates it.

3.4. Complement tests in LFG

Komlósy (1992), in his LFG analysis on Hungarian verbal argument structure, defines complements as elements whose syntactic and semantic properties are subcategorized by the governing verb. For analyzing given structures, he suggests using the following three tests:

- e) if a constituent is obligatory in any level of sentence structure, it is a complement;
- f) if a constituent's appearance in the structure allows to expand it further by an optional adjunct (which could not be present in the original structure), then this constituent is an (optional) complement;
- g) if a word X has an expansion Y, and there is a word Z which can systematically replace X+Y, and can replace X when X is not expanded by Y, but cannot replace X when Y is present, then Y is an optional complement of X.

3.5. Hungarian syntax

Hungarian is a highly inflective language with 18 cases and a (roughly) free word order: this means that almost any ordering of the verb and its expansions is acceptable, although they yield slightly different interpretations. As described in É. Kiss (2002), in the neutral sentence verbal complements and adjuncts follow the predicate within the VP. However, in most sentences, at least one verbal complement precedes the verb —this is the *topicalized* constituent, which is raised to the first position in the sentence. Another syntactic movement that changes the neutral word order is *focusing*: the focus position is the one immediately preceding the finite verb. Any type of complements or adjuncts can be topicalized and focused, hence moved outside the VP. Furthermore, *verb modifiers*, i. e. verbal prefixes, adverbs or bare NP complements also precede the verb they modify. Thus, most verbal complements and adjuncts are free to appear before the predicate. When parsing Hungarian texts, we face the difficulty of being unable to determine dependency relations and grammatical functions on the basis of constituent order. On the other hand, Hungarian morphology is very rich, thus we have to rely on constituents' morphological features, in particular on case marking. In compliance with these features of Hungarian, we find that most of the above-mentioned GB and LFG-related tests either do not apply or are insufficient for deciding on complementness.

GB tests

Tests a) and b) do not apply to Hungarian due to the lack of passivization and pronominalization. As to condition c), it is not always met in the surface order of Hungarian sentences:

A gyerekek nyírják a kertben a füvet.
 The children[NOM] cut[PL.3] the garden[INE] the grass[ACC].
 The children are cutting the grass in the garden.

Bea megtalálta tegnap a kutyádat.
 Bea found yesterday the dog[POSS.S2+ACC]
Bea found your dog yesterday.

In the sentences above, adjuncts (a locative NP in inessive case: *a kertben*-‘in the garden’ and a temporal adverb: *tegnap*, ‘yesterday’) precede the obligatory complements (direct objects in accusative case: *a füvet* —‘the grass’ and *a kutyádat*— ‘your dog’) and hence they are wedged in between the verb and its complements.

Assuming that *tesz*, the Hungarian counterpart of ‘put’, similarly to the English verb has three argument places, we can prove that condition d) concerning ellipsis possibilities within the VP does not hold for Hungarian:

<i>Ki megy hová kedden?</i>	— János ... moziba ...
Who goes where on Tuesday?	— John ... to the movie ...
<i>Ki tette a könyvet hová?</i>	— János ... az asztalra.
Who put the book where?	— John ... on the table.

LFG tests

There is a counter-argument for e) (which only holds for obligatory complements): any of the verbal complements can be omitted in Hungarian. As for f) and g) tests, they can only be applied to a limited number of verbs: the expansion of the structure by a complement does not always entail the possible appearance of an adjunct (as in f), and we cannot be sure to find a synonym (with different valence) for each verb we are dealing with (as in g). Thus these criteria do not seem to suffice for our purposes.

4. Compositionality as a criterion

Lexical entries of verbs thus contain those elements that appear in the local context of the verb and which cannot be derived by general phrase structure rules. In compliance with GB, we would like to construct lexical entries whose subcategorization frames specify the category and (in the case of postpositional complements) the lemma of the complements. On the other hand, instead of relying on the local context, we loosen this constraint and look for complements and adjuncts in the whole extent of the clause which contains the finite verb. The reason for it is the phenomenon known as “scrambling”: certain non-configurational languages with a rich morphology show a much bigger diversity in the surface order of sentence constituents than, for instance, English. Complements and adjuncts are free to mingle, they might even leave their clause after having received a case from the verb.

This implies that surface order cannot be used for separating complement and adjunct functions. On the other hand, the markedly rich morphological system can serve as a basis for our investigations. Thus, instead of using configurational information, we intend to use morphology, especially case feature as a marker of the syntactic role. The basic assumption is that not every occurrence of an NP with a case suffix is lexically subcategorized by a verb: some of them are added to the sentence by productive rules. Since such rules prescribe as a syntactic requirement the appearance of a certain case suffix on the NP and associate a syntactic role to it, we conclude that these suffixes are elements which enable NPs to fulfill certain roles. In other words, default meanings can be associated with case suffixes. Another important assumption we relied on is that once we manage to state the function of the NP with a particular case, its translation can be generated from the translation of the NP by the application of translation rules. Hence our definition of complementness and adjunctness will be based on their degree of compositionality.

With these presuppositions, our work starts by enumerating possible syntactic and semantic functions of case suffixes. This means that we try not to see complements in their relation to the predicates, and we conceive the predicate-argument functions as one possible function that case suffixes may bear. We find that there are two grammatical cases (nominative and accusative) that cannot have a default meaning and can only occur with verbal complements. These cases have to be included in verbal valence structures. As for the other cases, we try to define all their syntactic and semantic properties that can be described—and translated in a machine translation system—by general rules. Such rules specify one or more translations for the given case suffix, and may refer to semantic or syntactic features of the constituent they appear in, but general rules may not refer to the predicate. For example, the case suffix *-ban* (“inessive case”) indicates the exact date if it appears on a constituent expressing time: it forms a regular adjunct of time. Otherwise, it expresses location, and also forms a regular adjunct. These two rules will work as *default rules* for the case suffix *-ban*, assuming that in absence of lexical rules, the case is associated to one of these functions independently of the context:

[NP.case = ins, semantics=time]	→	[NP.role = time.adjunct]
[NP.case = ins]	→	[NP.role = loc.adjunct]

The default rule is conceived as a definition of the relation between the case-bearing element and the verb. In the example above, NPs get labeled as time or locative adjuncts which specify the kind of information they add to the predicate’s meaning.

While defining default rules, it comes into light that some [V + NP.case] structures are midway between rule-based constructions and total lexicalism. This means that their function can be stated, but their appearance depends on the semantic class of the predicate. For example, the ablative case *-tól* may have two default meanings: with movement verbs it marks the starting point of a movement; with verbs that express a change in someone’s state, it expresses the cause of the change. Thus, we can associate a function to the case suffix with rules that refer to the semantic class of the verb it occurs with. This kind of rule cannot be considered as default because it refers to the predicate. However, we find that even the appearance of regular adjuncts like time adverbs are constrained by the semantics of the verb they modify, but still we would not like to consider them as being part of the verb’s valence. These semi-productive rules represent a new category in-between complements and adjuncts: accordingly, when performing syntactic parsing, their application follows verbal valence matching but precedes default rules.

Because semi-default rules refer to verb classes when they apply, we had to tackle the task of creating predicate classes on reasonable grounds. The main characteristics of the classes are described by metapredicates. At this point we assume that systematic *syntactic* and *morphological* alternations are able to serve as good hints while defining our metapredicates. This presupposition is in accordance with what we stated before, namely, the NPs’ syntactic function in relation to the predicate of the clause depends on the existence of the case suffix’s meaning. This meaning also determines the given NP’s semantic relation to the verb.

First, let us examine the criterion of morphological changes of the predicate. Productive derivational processes change the meaning of the verb in a systematic way,

and sometimes also the argument structure of it. When the argument structure does not change we can conclude that the meaning denoted by the derivational suffix is of no importance regarding the NPs' syntactic or semantic role in the sentence, or that the given NP might be a complement the case suffix of which bears no meaning at all. As for the first possibility, the interesting thing here is that in most cases we cannot make such generalizations over the totality of verbs. This means that usually even if a derivational suffix does not have an effect on the argument structure of the majority of predicates, there might be also verbs whose argument structure does undergo certain changes. Indeed, this difference strongly implies the metapredicates we should use when describing a verb class, since we might expect that when the NP with its case suffix stays as it was, the meaning of the case suffix is compatible with that of the derivational suffix, while regarding the other case the meanings are controversial that is why the argument structure changes along with the meaning of the predicate itself.

For instance, *-gAt* is a derivational suffix which expresses two different *aktionsarts*, diminutive and iterative —depending on some semantic features of the base verb (Kiefer, Ladányi 2000). Usually *-gAt* does not change the argument structure of the base verb. Just as in the following verb pairs:

olvas - olvasgat 'keep reading'; *lövöldöz vkre - lövöldözget vkre* 'keep shooting at sg'

In fact, *-gAt* may also attach to verbs whose argument structure do change:

iszik vmre 'drink to sg' vs. **Iszogat vmre* 'keep drinking to sg';

halaszt vmit vmire 'postpone sg to sg' vs. **halogat/halasztgat vmit vmire* 'keep postponing sg to sg';

From the examples above it follows that there is a verb class which has a meaning component —expressed by the relevant NP's case suffix— that is not compatible with the meaning of *-gAt*.

Another basis we use to distinguish among predicate classes is the systematic change in the argument structure. It could be stated that from a linguistic point of view this criterion is the same as the morphological one. To supply proof for this hypothesis we only need to stipulate the existence of some zero-morphemes which are responsible for the alternation of argument structures. On the one hand this method is supposed to guarantee the reliability of the metapredicates. On the other hand we expect that these alternations have no effect on complements but on adjuncts, i.e. complements also might appear in the structure after the derivation with the same suffix. This follows from our presupposition, namely that in the case of complements the case suffix appearing on the head of the NP adds nothing to the whole structure's meaning. While derivational suffixes are considered as forms that yield always the same meaning —and obviously that is the way we want to look at them— we can say that derivational suffixes are functions that take only the verbs' meaning as input. As opposed to adjuncts, complements' suffixes play no role from the point of view of the verb's meaning, hence we would expect that any change of the verbs' meaning would leave such suffixes untouched. Regarding adjuncts we expect that —because the case suffix meaning is composed with the predicate meaning— they can undergo alteration when the appropriate derivational suffix is attached to the base verb.

Now we present how we use the conditions above in the case of the Hungarian suffix *-tól* (instrumental case). What kind of restrictions can be formulated regarding the three different argument structures below?

[11.a] *János felébresztette Mari - t a zaj - jal.*
 John awoke Mary - ACC the noise - INS
 'John awoke Mary with the noise.'

[11.b] *A zaj felébresztette Mari - t.*
 The noise awoke Mary - ACC
 'The noise awoke Mary.'

[11.c.] *Mari felébredt a zaj - tól.*
 Mary woke up the noise - ABL
 'Mary was awoken by the noise.'

We supposed that the semantic representation of the verbs belonging to this class is as follows:

CAUSE (John, E), where $E < \text{noise}, \text{CHANGE}(S, S') >$ and CAUSE(cause, S')

which means that John brought (CAUSE) a situation (E) into existence, and E is a two-argument predicate, such that there is an x ('noise'), which causes (CAUSE) a change in Mary's *mental* state, namely a change from S into S'. The next question is, how could we verify syntactically these three semantic components (i.e. CAUSE, MENTAL, CHANGE)? We suppose that a verb belongs to this class if and only if it can undergo systematically the syntactic alternations represented in [11.a.], [11.b.] and [11.c.].

As [11.a.] and [11.b.] show, the predicates belonging to this verb class have to have at least one interpretation where the subject is non-agentive. Otherwise [11.b.] should be ungrammatical, since the denotata of such subjects cannot carry out an action voluntarily. This requirement is responsible for the fact that most verbs in this class—not all, though—are mental verbs. (Note that all mental verbs with this argument structure have a non-agentive interpretation.)

[11.c.] illustrates the necessity of the metapredicates CAUSE and CHANGE. According to Komlósy (2000) one default meaning of the ablative case *-tól* is the CAUSE of something. Though in cases which are similar to the example above, i.e. where all the three argument structures are well-formed, the change has to be a transition from a state (S) *into* another state (S'). This transition is referred to by the metapredicate CHANGE. There are two arguments to support this thesis. The first relies on the English translation¹; the elements of this verb class are inclined to be translated into English by perfective verb forms. The structure in sentence [11c] cannot even be put in an imperfective form with the same argument structure. This fact is in accordance with our expectation that sentences with the perfective forms of these structures involve the complete transition between two states, while imperfective forms express the process of transition, but do not imply the end of this process. The other argument takes as its starting point

¹ We have to use English translation, since Hungarian lacks perfective-imperfective distinction expressed by tense.

the observation that there is a verb class with verbs such that the argument with instrumental suffix represents the CAUSE as in the instances above, but there is no transition between definite states which means that CHANGE predicate cannot apply:

[12a] *Az igazgató János - t terhelte a feladat - tal.*
 The director John - ACC burden the task - INS
 ‘The director burdened John with the task’

[12b] *A feladat János - t terhelte.*
 The task John - ACC burdened
 ‘The task burdened John.’

[12c] *János terhelve van.* [12d] **János terhelve van a feladat - tól.*
 John burdened is John burdened is the task - ABL
 ‘John is burdened.’ ‘John is burdened by the task.’

This semantic intuition is caught by the explicit criterion of the syntactical ill-formedness of the sentence [12d]. As the counterexample demonstrates the metapredicate CHANGE is distinctive, that is why we need it independently of CAUSE.

6. Conclusion

Our work aims at creating a well-defined and efficient method for NLP applications to distinguish between verbal complements and adjuncts. The usability of such an algorithm depends basically on two parameters:

- it has to be explicit enough so that different people working on parallel on argument structure descriptions produce coherent, homogeneous work,
- every piece of relevant information that is not predictable by general rules has to be classified as lexical.

We built up our method upon these criteria, assuming that *case suffixes*, the syntactic markers of grammatical functions are not only markers but that there are verb + adjunct structures in which they take part with their own morphosyntactic properties and meanings. These structures are compositional: the verb, the NP and its case suffix form a syntactic unit and the meaning of this unit is calculated adding up the meaning of the verb, the meaning of the NP and that of the case suffix. On the other edge of the scale we find *complements*: they figure in non-compositional structures where the case suffix of the complement NP does not contribute any predictable meaning on its own. The semantic role of the complement NP, i.e. the relation between its denotatum and the action/fact referred to by the verb, depends solely on the verb’s lexical properties. Midway between these categories, we found a set of structures in which the case suffix behaves the same way as in adjunction but which are restricted to semantically characterizable classes of predicates. As we would like to reduce the amount of data stored in the lexicon, we decided to capture semi-adjunct structures by the so-called *non-default rules* which only apply to given sets of predicates. These predicates are described by semantic metapredicates.

In accordance with our expectations, the number of ‘real’ complements reduced considerably. As a secondary result of our work, a cluster of syntactically relevant se-

mantic features is shaping up from the metapredicates that define semantic classes (e.g. cause, change). We expect metapredicates to be language-independent. If this assumption proves to be right, not only default rules but also non-default rules can behave as translation rules for case suffixes: the only modification needed in the rules is the replacement of linguistic labels of NPs in the output of the rule by the target language syntactic marker of the role referred to by the label, e.g.:

<i>rule type</i>	<i>source language (HUN)</i>			<i>target language (EN)</i>		
non-default	V + change_state	NP	case = ABL	V	NP	prep = 'because of'
default	V	NP	case = INS	V	NP	prep = 'with'

The figure illustrates how a non-default and a default rule may be captured in the bilingual module of a Hungarian-to-English MT system. The non-default rule takes as source language input a verb which denotes a change in someone's state, modified by an NP in ablative case, and translates the case suffix by the preposition *by*. The default rule does not place any restriction on the verb, and states that any NP in instrumental case which has not been matched by earlier rules has to be translated as an instrument, by the preposition *with*.

The most important future task is to find a way to verify the language-independent character of metapredicates. Meanwhile, the precise elaboration of the lexical argument structure database and in parallel, its use in a rule-based machine translation system (Prószycki and Tihanyi 2002) are being carried out. Considering the very strict claim MT sets up for separating language-independent and language-specific information, as well as the practical requirement to cover by rules as many phenomena as possible, we believe that MT as an application is also a relevant test to verify the foundations of our method.

References

- Kiefer, F. and Ladányi, M., 2000, «Morfoszintaktikailag semleges képzések». In: *Strukturális magyar nyelvtan* 3. ed. by Kiefer, F., pp. 165-214. Budapest. Akadémiai Kiadó («Morpho-syntactically neutral derivational processes». In: *Structural Grammar of Hungarian* 3.)
- Kiss, K., 2002, «The Syntax of Hungarian». Cambridge. Cambridge U. P.
- Jurafsky, D. and M., J., 2000, *Speech and Language Processing*. Prentice Hall.
- Komlósy, A., 1992, «Régensek és vonzatok». In: *Strukturális magyar nyelvtan* 1. ed. by Kiefer, F., pp. 299-529. Budapest. Akadémiai Kiadó («Predicates and Complements». In: *Structural Grammar of Hungarian* 1.)
- 2000, «A műveltetés». In: *Strukturális magyar nyelvtan* 3. ed. by Kiefer, F., pp. 215-291. Budapest. Akadémiai Kiadó («Factitive verbs». In: *Structural Grammar of Hungarian* 3.)
- 2001, *A lexikai-funkcionális grammatika mondattanának alapfogalmai*. Budapest. Tinta kiadó («Syntax in Lexical-Functional Grammar»)
- Prószycki, G. and Tihanyi, L., 2002, «Metamorpho: A Pattern-Based Machine Translation Project». In: *Proceedings of the 24th Translating and the Computer Conference*. pp. 19-24. London.
- Radford, A., 1988, *Transformational Grammar*. Cambridge. Cambridge U. P.
- Silberstein, M., 1993, *Dictionnaires électroniques et analyse automatique de textes: Le système Intex*. Paris. Masson.

PHASE THEORY, CASE AND RELATIVE CLAUSES

Ángel J. Gallego

Universitat Autònoma de Barcelona

0. Introduction¹

The goal of this paper is to put forward an analysis of relative clauses which builds on Pesetsky & Torrego's (2001) proposal concerning the C-T connection and the nature of Case. In so doing, a unitary answer to two long-standing puzzles of the relative clause realm will be provided: the absence of both *that*-deletion and overt relative pronouns (unless introduced by a preposition) in Romance languages, which are shown in (1).

- (1) a. El hombre *(que) vi. (Spanish)
The man that see-PST.1SG
'The man (that) I saw'
- b. El hombre *(con) quien habló. (Spanish)
The man (with) who talk-PST.3SG
'The man who talked' / 'The man to whom (s)he talked'

As I argue below, the solution to the data in (1) will further prove useful in trying to explain a more general paradigm of asymmetries between English and Null Subject Languages which seems to point to Case Theory as the *Locus* of parametric variation; in particular, evidence will be provided showing that languages may differ with respect to the derivational stage at which subjects get their Case checked, with non-trivial consequences for additional operations taking place in the CP phase.

The present proposal differs from previous ones (cf. Arregui 1998, Bianchi 1999, Brucart 1992, Law 2002, Ojea 1992, and Toribio 1992, *inter alia*) in dispensing with cartographic, government and Optimality Theory based accounts, underscoring the important role of Case and its bearing on computational processes. The paper is divided as follows: in section 1, I lay out the system and technical operations I assume throughout; section 2 focuses on the recent revival of Vergnaud's (1974) 'raising analysis' of relative clauses by Kayne (1994), and the subsequent refinements introduced by Bianchi (1999); in section 3, I put forward a minimalist analysis for

¹ I would like to thank Ricardo Etxepare, Susana Huidobro, and Aritz Irurtzun for helpful discussion. Special thanks go to Valentina Bianchi and Esther Torrego for their generous and insightful comments. A previous version of this paper was presented at the 2004 *Going Romance* (9 december, 2004, Universiteit Leiden). Usual disclaimers apply.

relative clauses that highlights the role of Case and the syntactic dependency between C and T. Section 4 summarizes the main conclusions.

1. T-to-C Movement

Much research and comparative work stemming from Den Besten (1983) has shown that a key syntactic relation exists between the functional categories C and T in natural languages. Such dependency is sometimes abstract, although it has mainly been explored in terms of familiar phenomena: verb movement to C (in V2 languages) and *that*-trace effects.² The conclusion drawn from that evidence seems to be that a T element has to move to C, a fact which Pesetsky & Torrego (2001) encode as follows:

(2) *Motivation for T-to-C Movement*

C bears an uninterpretable T feature (henceforth [uT]) with the EPP property.
[from Pesetsky & Torrego 2001: 360]

By the ‘EPP property’ Pesetsky & Torrego (2001) understand a trait of a feature, not a feature itself; put differently, if a feature F is endowed with the EPP property, it will trigger overt movement (what Chomsky 2004 dubs *internal-Merge*).³

In the context of the present discussion, it is important to step back a little bit and introduce the basics of an operation crucially related to movement:⁴ *Agree*. Minimalism makes a central distinction between interpretable and uninterpretable features. As Pesetsky & Torrego (2004b) point out, this cut does not capitalize on features *per se*, but rather on whether a given feature makes a semantic contribution in the lexical item in which it appears. In this vein, Chomsky’s (2000, 2001, 2004) attention is placed in the Case/Agreement systems, taking these notions to be the two sides of the same coin: ϕ -features (i.e., nominal inflectional features like gender, number and person) are interpretable in nouns, not in verbs; therefore —Chomsky argues—, uninterpretable ϕ -features placed in verbal morphology enter syntax without a value, which makes them act as a Probe seeking for a Goal, (typically) a DP down in the tree endowed with interpretable ϕ -features: the Goal’s ϕ -features value those of the Probe, and, as a result, it receives structural Case. Chomsky (2000) calls this operation *Agree*.⁵ Note that, as stated, all *Agree* cares about is valuation, not movement, but it is an empirical fact that valuation is followed by *internal-Merge* of the Goal under certain circumstances, creating a SPEC: this is precisely the role of the EPP property.

² Cf. Koster (2003), Pesetsky & Torrego (2001), Rizzi (1990), and references therein.

³ A reviewer asks what the difference between the EPP being a feature proper or a trait of a feature is. Technically, the difference is important: only *bona fide* features (e.g., ϕ -features) can *Match* other features by means of *Agree*. On the other hand, the EPP property, as understood in Pesetsky & Torrego (2001), cannot *Match* anything: it is simply a mechanism parasitic on *Agree*.

⁴ I put aside the modifications in Chomsky (2005), where overt movement does not always invoke *Agree*.

⁵ This process of long-distance checking dispenses with Chomsky’s (1995) *Attract*, which was viewed as head-movement. Cf. Boeckx (2003a, 2003b, 2004) for discussion.

With this theoretical background in mind, I assume, following Pesetsky & Torrego (2001), that whenever *internal-Merge* occurs, the relevant Probe has the EPP property (making it ‘strong’, a notion supposed to capture the overt/covert nature of operations in previous models).⁶ Let us consider the examples in (3) in order to see the role of the EPP property. Adopting the view that the traditional EPP (i.e., the need for SPEC-T to be filled in) is related to T’s ϕ -features, a language like Catalan has the two options depicted in (3), depending on whether the EPP property is active or not:

- (3) a. $[_{TP} [_T \text{Canta}_i [_T T_{[\neq\phi]}]] [_{\nu^*P} \text{en Joan}_{[i\phi]} t_i]]$ (Catalan)
 sing-PRS.3SG the Joan
 ‘Joan sings’
- b. $[_{TP} [_{DP} \text{En Joan}_j [_i\phi]] [_T \text{canta}_i [_T T_{[\neq\phi, \text{EPP}]}]] [_{\nu^*P} t_j t_i]]$ (Catalan)
 The Joan sing-PRS.3SG
 ‘Joan sings’

In (3) T’s ϕ -Probe scans its complement domain looking for a Goal: the subject DP *En Joan*. The main difference between (3a) and (3b) has to do with *internal-Merge*: in (3a) T’s ϕ -features are not endowed with the EPP property (hence valuation alone suffices), whereas in (3b), they are, triggering *internal-Merge* of the subject DP.^{7 8}

Let us now return to Pesetsky & Torrego’s (2001) proposal. To start with, consider the next paradigm, originally noted by Koopman (1983):

- (4) *T-to-C Asymmetry in Matrix Interrogative Clauses*
- What did Mary buy?
 - *What Mary bought?
 - *Who did buy the book? [*unless *did* is focused]
 - Who bought the book?

[from Pesetsky & Torrego 2001: 357]

Descriptively speaking, what is going on in (4) is very clear: *do*-insertion is blocked whenever a subject DP undergoes *wh*-movement to SPEC-C. Contrary to Koopman’s (1983) approach, which relied on government (a device no longer available within the current framework), Pesetsky & Torrego (2001) account for the examples in (4) by claiming that what we call ‘Case’ is an uninterpretable Tense feature on D heads. Let me elaborate. For Chomsky (2000, 2001, 2004, 2005), Case features have no matching counterpart whatsoever, they are purely formal uninterpretable features: when the ϕ -features of T and ν^* are valued, the nominals they agree

⁶ Cf. Chomsky (1993, 1995), Nissenbaum (2000), and Pesetsky (2000).

⁷ Note that this analysis does not invoke an expletive *pro* in SPEC-T, disregarding the universality of the EPP. Since this issue is orthogonal to the focus of this paper, I will put it aside. Cf. Boeckx (2003b).

⁸ A reviewer is concerned about the preverbal vs. postverbal position of the subject DP and its bearing on information structure. In the analysis I am assuming, those interpretive effects follow from T’s ϕ -features having (or not) the EPP property: if the subject is preverbal, it receives a topic interpretation; if it is postverbal, a non-contrastive focus interpretation arises (cf. Belletti 2004). Accordingly, I take effects on information structure (what Chomsky 2004 calls *edge-semantics*) to follow from *internal-Merge*.

with get Case, period. The asymmetry is blatant, as Pesetsky & Torrego (2004b: 10) correctly note:

The [Minimalist Inquiries]/[Derivation by Phase] framework does not view structural case as the uninterpretable counterpart of an otherwise interpretable feature. Instead, it is a *sui generis* feature with a special relation to the ϕ -features: it gets valued only as a by-product of ϕ -feature agreement. Thus, when the unvalued ϕ -features of finite T probe, on this approach, and find a suitable goal—for example, a DP with a full set of ϕ -features—the unvalued case feature of that DP gets valued as a kind of ‘bonus’.

An alternative view on Case like Pesetsky & Torrego’s (2001) is interesting inasmuch as it holds that all grammatical features have some potential semantic value. This is conceptually preferable and, furthermore, restores the asymmetry of Chomsky’s view about structural Case: both T (formerly, Case) and ϕ -features have matching counterparts. The bottom line of this view can be stated as in (5):⁹

- (5) *The Nature of Case*
Case is [u T] on D

Now, have a look at (4) again. What must be answered is why the subject’s *wh*-movement does not trigger *do*-insertion, which is itself an instance of T-to-C movement within this system. According to Pesetsky & Torrego’s (2001), *do*-insertion is barred because the nominative Case feature (that is, [u T]) of the subject DP can delete C’s [u T], rendering *do*-insertion as redundant. Graphically:

- (6) a. [_{CP} Who_i_[u T] [u Wh]] C_[u T, EPP] [u Wh, EPP] [_{TP} t_i bought the book]]
b. * [_{CP} Who_i_[u T] [u Wh]] did_j C_[u T, EPP] [u Wh, EPP] [_{TP} t_i T_j buy the book]]

Under (6) lies a core property of the computational system: economy. As the reader may easily see, if one operation suffices to check two uninterpretable features, no extra operations are needed. In (6a), the T feature of the subject DP is closer to C than T itself (taking strict c-command to signal closeness, cf. (8) below),¹⁰ and, in addition, it can also be used to check the [u Wh] feature:¹¹ by a principle of computational economy like (7), moving the subject DP should be enough to satisfy C’s requirements, and it is indeed, as (4) shows. On the other hand, when object DPs move, T is always closer to C, so pure T-to-C movement (i.e., *do*-insertion) must occur.¹²

⁹ Cf. Svenonius (2002) for a similar view on Case.

¹⁰ That is, what matters for being a closer Goal is strict c-command (putting aside equidistance-based definitions; cf. Chomsky 2001). This can be spelled-out as in (i), from Pesetsky & Torrego (2001: 362):

- (i) *Closeness*
Y is closer to K than X if K c-commands Y and Y c-commands X.

¹¹ I assume that matrix interrogative C bears an uninterpretable [Wh] feature endowed with the EPP property. Things are different in Chomsky (2005), for all A'-Movements are triggered by EPP/edge-Probes. Since nothing I have to say here crucially hinges on this notational alternative, I will ignore it.

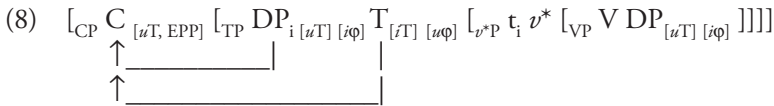
¹² At first glance, there is a non-trivial drawback to this proposal: how can it be the case that C’s [u T] be valued by the subject’s [u T], since both features are unvalued? First of all, it must be noted that this possibility is severely restricted, for an unvalued feature can be used to value another unvalued feature only within the phase it has been marked for deletion, as Pesetsky & Torrego (2004a) argue. Second, in

(7) *Economy Condition*

A head H triggers the minimum number of operations necessary to satisfy the properties (including EPP) of its uninterpretable features.

[from Pesetsky & Torrego 2001: 359]

As (8) shows, subject DPs are indeed closer to C than T, under strict c-command (object DPs are obviously too buried in the structure, as noted):



Are there any other cases of T-to-C movement? In Pesetsky & Torrego (2001), *that*, the morpheme assumed to fill in the C position, is analyzed as a clitic head doubling T which deletes C's [uT]. By parity of reasoning, the same should hold for the so-called 'prepositional complementizers' (cf. Bresnan 1972, Kayne 2000 and Pesetsky & Torrego 2001, 2004a). Interestingly enough, this take on complementizers derives *that/for*-trace effects straightforwardly:

- (9) a. Who_i did John say $[_{CP} t_i C_{[uT, EPP]} [_{TP} t_i \text{ called Mary}]]$?
- b. *Who_i did John say $[_{CP} t_i \text{ that}_j C_{[uT, EPP]} [_{TP} t_i T_j \text{ called Mary}]]$?
- (10) a. Who_i would John like $[_{CP} t_i C_{[uT, EPP]} [_{TP} t_i \text{ to buy the book}]]$?
- b. *Who_i would John like $[_{CP} t_i \text{ for}_j C_{[uT, EPP]} [_{TP} t_i \text{ to}_j \text{ buy the book}]]$?

If *that* deletes C's [uT] and deletion of uninterpretable features is required for convergence at the interfaces, one might now wonder what to do with *that*-deletion (cf. (11) below): how is C's [uT] deleted in those cases? Pesetsky & Torrego (2001) argue that both TP and the DP in SPEC-T¹³ are equally able to delete C's [uT], since, c-command-wise, both are equally close to C (that is, they are 'equidistant').^{14 15}

- (11) a. John thinks $[_{CP} \text{ that}_j C_{[uT, EPP]} [_{TP} \text{ Mary } T_j \text{ is gorgeous}]]$
- b. John thinks $[_{CP} \text{ Mary}_i \text{ for}_j C_{[uT, EPP]} [_{TP} t_i \text{ is gorgeous}]]$

For the purposes of the present section, we can stop at this point. I have presented the main aspects of Pesetsky & Torrego's (2001) analysis of Case features (henceforth, [uT] features) and the C-T interaction. As we have seen, their proposal accounts for some well-known phenomena in a unitary fashion, with the additional advantage of giving Case a more coherent treatment within a Probe-Goal system.

Pesetsky & Torrego (2004b), a possible way out is sketched: all instances of T features form a sort of abstract syntactic dependency (technically, *Agreement* is regarded as *Feature Sharing*; cf. Frampton & Gutmann 2000) so that an unvalued link is not 'alone' when valuing another unvalued feature appearing up-stairs in the tree: the chain works 'together', as a whole, in valuation. Another possible implementation of this technical solution is Hiraiwa's (2001) *Multiple Agree*. Cf. section 3 for more relevant discussion.

¹³ Although I say TP here, it is actually the T head that can move to C, being spelled-out as *that*. Cf. Pesetsky & Torrego (2001) for details about 'equidistance' between TP and SPEC-T.

¹⁴ Another possibility would be for C to delete its [uT] feature by mere *Agree*.

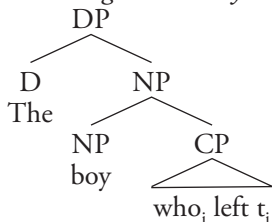
¹⁵ Cf. Chomsky (1993, 1995, 2000) and Hiraiwa (2001) on 'equidistance'. If this notion is eliminated, as in Chomsky (2001), the possibility to use subject DPs' [uT] to value C's [uT] could still take place in a *Multiple Agree/Feature Sharing* fashion, but I put this aside here.

2. The Raising Analysis of Relative Clauses

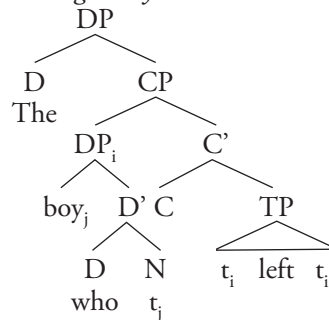
In this section I introduce some evidence in favor of the ‘raising analysis’ of relative clauses (originally proposed by Brame 1968 and developed later on by Schachter 1973, Carlson 1977, and specially Vergnaud 1974), focusing on Kayne’s (1994) and Bianchi’s (1999, 2000) particular implementations.

Taking the base position of their head as a classifying criterion, it can be said that relative clauses have received two main approaches in the literature:¹⁶ the ‘matching’ and the ‘raising’ analyses. In the latter, the nominal head is generated inside the relative clause prior to its movement to SPEC-C; in the former, it is generated outside, and the relative clause is an adjunct. Consider these differences in (12):

(12a) *Matching / Wh-Analysis*¹⁷



(12b) *Raising Analysis*



Due to the technical limitations imposed by the Antisymmetry framework, Kayne (1994) adopts the raising account: since right adjunction is not an option under Kayne’s (1994) LCA, the relative CP and the D head must directly undergo *external-Merge*, as depicted in (12b). There is robust evidence in the literature supporting this analysis (cf. Bhatt 2002, Bianchi 1999, 2000, Brame 1968, Kayne 1994, Sauerland 2000, Schachter 1973, *inter alia*). Consider some examples from binding (13a), definiteness effects (13b), and idiom interpretation (13c):¹⁸

¹⁶ For a historical review cf. Bianchi (2002a, 2002b). For evidence suggesting that both analysis (i.e., *external* and *internal-head-Merge*) actually exist, cf. Sauerland (2000) and Szczegielniak (2004).

¹⁷ The main difference between the *Matching* and the *Wh-* analyses is that the former involves two NPs (one of which gets deleted and replaced by a relative pronoun), whereas the latter involves just one. Importantly, both analyses share the idea that the relative clause is an adjunct to the NP: a constituent creating a two-segmented category (cf. Chomsky 1986), without altering the nature (i.e., the *label* or *type*) of the element it adjoins to.

¹⁸ Citko (2001) points out some problems for the ‘raising analysis’, the most important one having to do with anti-reconstruction effects (cf. Chomsky 1993, 2004, and Lebeaux 1991). There is some controversy on these data (cf. Bianchi 1999: 109-115), but the contrasts seem rather clear: in (i), the R-expression *John* can take *he* as its antecedent. As (ii) shows, the same pattern holds in Spanish:

- (i) [Which picture of Bill_z [that John_j liked]]_i did he_{j/z} buy t_i ?
 (ii) ¿[Qué libro [que María_i recibió ayer]]_i crees que pro_j leerá t_i antes? (Spanish)
 What book that María get-PST.3SG yesterday think-PRS.2SG that read-FUT.3SG before
 ‘Which book that María got yesterday do you think she will read first?’

Under Chomsky’s (2004) analysis of adjuncts, (i) and (ii) do not pose any problems for a raising account, since reconstruction only applies at the point where *Transfer* takes place (that is, reconstruction

- (13) a. Mary bought the [picture of himself]_i [_{CP} that John_j saw t_i]
 b. The men_i [_{CP} that there were t_i in the garden] were all diplomats.
 c. The headway_i [_{CP} that John made t_i] proved insufficient.

In a nutshell, the data in (13) support an analysis in which the head is not external to the relative clause: instead, it must be generated in a clause internal position and then undergo *internal-Merge* with C. Consider the binding datum in (13a) in more detail, for instance: under fairly standard assumptions about Condition (A) (cf. Chomsky 1993), the anaphor *himself* must be c-commanded by its antecedent (*John*, in the case at hand) at SEM; crucially, for that scenario to emerge, *himself* must be reconstructed into a clause internal position, an operation consistent with the ‘raising analysis’. The same logic applies in the other cases.

Going back to Kayne’s (1994) proposal, it is important to highlight two of its aspects: it treats relative pronouns (e.g., *who*, *which*, etc.) as determiners of the relative head and it assumes that the derivation of relative clauses unfolds in two basic steps: 1) *wh*-movement of the relative DP to SPEC-C and 2) movement of the head to SPEC-D, stranding the relative D. Bianchi (1999) adopts the basics of Kayne’s (1994) analysis, introducing some qualifications to which I return; before going into that, though, let me dwell on the D stranding operation for a moment: what I want to underscore here is the fact that such a process is optional, in the sense that relative clauses do not always contain a relative D, as is clear from the relativization patterns noted in Bianchi (1999):

- (14) a. The book [_{CP} *that* I read] *that*-relative
 b. The book [_{CP} *which* I read] *wh*-relative
 c. The book [_{CP} \emptyset I read] *zero*-relative

The examples in (14) differ in the formal element introducing the relative clause: the complementizer *that*, the relative D *which*, and a null head. As I said, regardless of their theoretical affinities, Bianchi’s (1999) analysis departs from Kayne’s (1994) in non-trivial respects. I will consider two aspects here, those related to the examples I started this paper with (cf. (1)). The first one has to do with the analysis of *zero*-relatives (or, alternatively, the *that*-deletion option, cf. (14a,c)), while the second one affects an asymmetry concerning what I will call ‘oblique relatives’, that is, *wh*-relatives that display a preposition (e.g., *The man to whom I talked*).

Let us consider the analysis of *zero*-relatives before tackling the asymmetries of (1). Contrary to Kayne (1994), who argues for NP raising to SPEC-C when there is no relative D, Bianchi (1999), building on Longobardi (1994), postulates a null relative operator heading the constituent, a turn that correctly qualifies the operation as a sub-case of A’-Movement:

- (15) a. [_{DP} The [_{CP} [_{NP} book]_i [_{CP} (that) [_{TP} I read t_i]]]] Kayne (1994)
 b. [_{DP} The [_{CP} [_{DP} D_{REL} book]_i [_{CP} (that) [_{TP} I read t_i]]]] Bianchi (1999)

only affects spelled-out copies, which are ‘simplified’ —reintroduced into the primary plane— by the time *Transfer* sends chunks of structure to PHON and SEM; cf. Chomsky 2004 for details).

By the end of the derivation, the internal null relative determiner D_{REL} incorporates into the external one by a government-based morphological process applying at PHON (when the relative D is overt, such incorporation does not obtain). Importantly, if a preposition intervenes between the external D and the internal one (i.e., D_{REL}), the derivation crashes, for incorporation fails, as (16b,c) show:

- (16) a. The man to whom I talked.
 b. *The man to that I talked. c. *The man to I talked.

At the outset of this paper I pointed out that there are two remarkable differences between English and Romance relative clauses. The first one concerns *zero*-relatives: these are impossible in Romance, but not in English. Consider the case of Catalan:

- (17) El llibre *(que) vaig comprar. (Catalan)
 The book (that) AUX-1SG buy-INF
 ‘The book (that) I bought’

At the same time, only English allows *wh*-relatives —Romance must introduce them by using a preposition. This is the second asymmetry:

- (18) The book which John read. (English)
- (19) a. *El libro el cual Juan leyó. (Spanish)
 The book the which Juan read-PST.3SG
 ‘The book which Juan read’
 b. *L’uomo il quale veniva. (Italian)
 The-man the which come-PST.3SG
 ‘The man who came’
 c. *L’homme lequél venait. (French)
 The-man the-which come-PST.3SG
 ‘The man who came’

That is, overt relative Ds must be introduced by a preposition in Romance for the derivation to converge. (20) confirms this.

- (20) a. El libro con el cual Juan estudió. (Spanish)
 The book with the which Juan study-PST.3SG
 ‘The book with which Juan studied’
 b. L’home amb el qual va venir. (Catalan)
 The man with the which AUX-3SG come-INF
 ‘The man with whom (s)he came’

Adopting Rizzi’s (1997) ‘CP-Split Hypothesis’, Bianchi (1999) postulates the next parameter in order to provide an explanation for these facts:

- (21) *Topic Parameter*
 ± Topic optionally supports the features [+declarative] and [+relative]
 [from Bianchi 1999: 186]

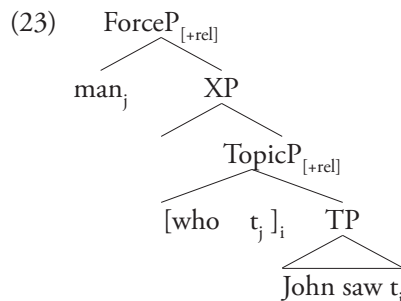
According to (21), Rizzi's (1997) Topic° can be endowed not only with a [+topic] feature (the default scenario), but also with [+declarative] and [+relative] ones. Crucially, such a repertoire is possible only in English, not in Romance. Furthermore, since it is null, this functional head is supposed to play a key role in *that*-deletion by Bianchi (1999): if Topic° bears [+declarative], an embedded declarative clause obtains (e.g., *Mary said John had left*); if it bears [+relative], then a *zero*-relative does (e.g., *The book John read*). However, notice that Topic° alone is not enough to derive *wh*-relatives: an extra head is needed, one which is supposed to carry [+relative] features by default —Rizzi's (1997) Force° . Things being so, English (a language for which (21) is marked positively), but not Romance, has two different heads being able to bear a [+relative] feature: Force° and Topic° . This is the key of Bianchi's (1999) analysis.

The facts in (18)-(19)-(20) have been addressed in the recent literature by many authors —some of them within the Government-and-Binding framework— (cf. Arregui 1998, Brucart 1992, Gutiérrez-Rexach & Mallén 2003, Law 2000, Ojea 1992, and Toribio 1992, *inter alia*). It is quite telling that Kayne (1994: 90) himself acknowledges that he does not understand what is going on; he just notes that Romance languages seem to lack 'enough room' in the CP-field for the head to strand the relative D:

The contrast between English, on the one hand, and French and Italian, on the other, can be stated by allowing English to use the specifier position of the *wh*-determiner itself as a landing site [...] (*At present, I have no account of why French and Italian differ from English in this respect.*) <Emphasis added: AJG>

Under Bianchi's (1999) account, therefore, *wh*-relatives have no problem in English, for this language has a positive setting of (21). In particular, Bianchi's (1999) analysis of (22) is as in (23) (irrelevant details omitted):¹⁹

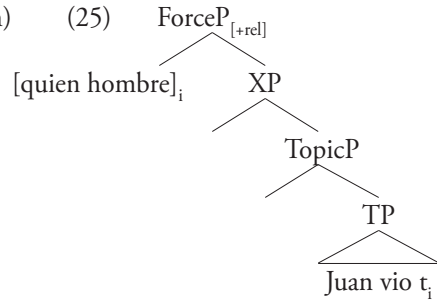
(22) The man who John saw.



Given that Romance languages only have one of the two required landing sites for *wh*-relatives (again, due to (21)), the derivation of (24) is doomed.

¹⁹ Note that the derivation of *wh*-relatives by Bianchi (1999) is different from Kayne's (1994) in that the relative head does not land in the SPEC of the relative D, but rather in the SPEC of Rizzi's (1997) Force° .

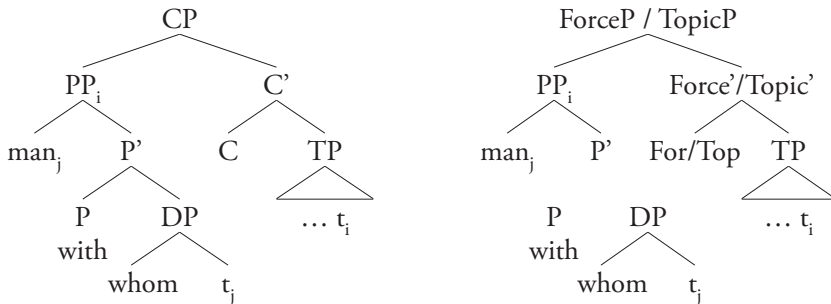
- (24) *El hombre quien Juan vio. (Spanish)
 The man who Juan see-PST.3SG
 ‘The man who Juan saw’



In (25), the relative DP *quien hombre* (Eng. *who man*) reaches the first available SPEC (namely, SPEC-Force), but then the relative head (i.e., *man*) cannot move any further, for there is no available SPEC with the [+relative] feature upwards in the tree.

As for ‘oblique relatives’, both groups of languages are able to license SPEC-P as a landing site for the relative head. The only remarkable difference concerns the landing SPEC of the moved PP: Kayne (1994) uses a standard CP structure, whereas Bianchi (1999) takes both TopP and ForceP to be potential landing sites.

- (26) *Oblique Relatives* [from Kayne 1994] (27) *Oblique Relatives* [from Bianchi 1999]



In this section I have summarized the basic properties of Bianchi’s (1999) and Kayne’s (1994) ‘raising’ proposals. In principle, both analyses (specially Bianchi’s 1999) seem to account for the main data, but they fail to provide a principled explanation of the asymmetries teasing English and Romance languages apart. In this respect, notice that one important drawback to Bianchi’s (1999) analysis is that it must stipulate the *Topic Parameter*, which, despite building on Rizzi’s (1997) ‘CP-Split-Hypothesis’, seems to contradict it, for it goes against one of the central claims by Rizzi (1997): each projection checks a unique feature, satisfying a dedicated *Criterion*. Given these problems, I will explore an alternative analysis in the next section.

3. The Proposal: a T-to-C Movement Account

Having seen the most recent analyses of relative clauses, now I turn to a proposal that assumes Pesetsky & Torrego’s (2001) findings regarding T-to-C movement and

Case Theory. Importantly, I also assume (28) as a principle of cyclic derivational dynamics:

(28) *Timing of Deletion of Uninterpretable Features*

An uninterpretable feature [uF] marked for deletion (i.e., [$\bar{u}F$]) within a completed phase P, is deleted the moment a new head H is merged to P.

[from Pesetsky & Torrego 2004a: 516]

In plain English, (28) can be paraphrased as follows: uninterpretable features can enter in checking processes within the phase they have been marked for deletion, but not beyond —when a new phase starts, all the features of the previous one become inert/useless for computational purposes.

What features does C have in relative clauses? I propose that, apart from [uT], C be endowed with an additional uninterpretable relative feature [$uRel$], whose nature is similar to a typical [Wh] feature.²⁰ This feature works as expected: as a Probe looking for a Goal in its c-command domain. Let us see how the three types of relative clauses in (14) would be analyzed under this proposal. Consider *wh*-relatives first.

(29) a. The man who loves Mary.

b. [_{DP} The [_{CP} C [_{uT , EPP}] [$uRel$, EPP] [_{TP} [_{DP} who man] _{i} [$uRel$] [$\bar{u}T$] loves Mary]]]

c. [_{DP} The [_{CP} [_{DP} who man] _{i} [$uRel$] [$\bar{u}T$] C [_{uT , EPP}] [$uRel$, EPP] [_{TP} t _{i} loves Mary]]]

How are C's features deleted in (29)? I argue that both [$uRel$] and [uT] are deleted by moving the relative subject DP: just like in matrix interrogative questions, and following Pesetsky & Torrego (2001), I assume that, in English, the [uT] of a subject DP can be used to delete C's [uT]. But we are not done yet; once we have arrived this far, what triggers the next movement? (i.e., what makes the N *man* in (29) strand the relative D *who*).²¹ For Kayne (1994) and Bianchi (1999) the answer is clear: the head must be in a configuration where it can receive Case, either by government or by another checking mechanism.²² Either way, we need some motivation for the head to move. Being extremely naïve about it, there are three candidates that come to mind:

(30) a. The external D.

b. An extra head between the CP and the external D.

c. C itself.

We can dismiss the third option right from the beginning: it would require not only to posit a new feature on C, but also to suppose that *Agree* can engage an ex-

²⁰ The proposal assumes that the [Rel] feature is interpretable in relative pronouns (as seems plausible), but not in C. This is consistent with the way of identifying relative clauses: by locating a relative operator. In other words: clauses are not relative or interrogatives *per se*, but rather because they contain an element which bears the [Rel] or [Wh] dimension as a defining characteristic.

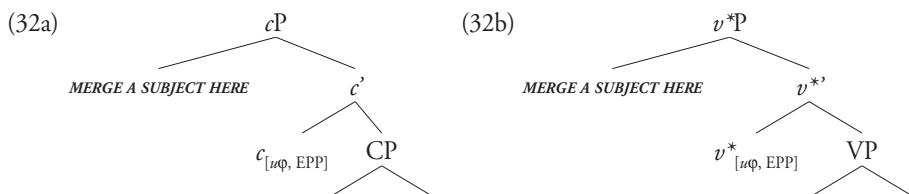
²¹ Notice that this D stranding process is very bizarre. If correct, this may indicate that relative DPs are not phases, at least not in Chomsky's (2000, 2001) terms (recall that phase heads cannot be stranded).

²² Actually, in Bianchi (1999, 2000), the entire process is even more obscure, since the head also moves in order to check a strong categorial feature that the external D is endowed with.

ceptional probing procedure (Probes can only scan their *c*-command domain, which does not include SPECs). Since the first option is essentially Kayne's (1994) and Bianchi's (1999) (or a slightly modified version of it, whereby the head moves to check its Case—an analysis incompatible with the Case Theory I am assuming here) let us explore the second one: an extra head. I will dub this head “*c*” in order to capture the fact that it is reasonably analogous (though not identical) to *v** within the VP-system, in the sense that it introduces a ‘subject of predication’.²³

- (31) [_{cP} man_j [_c *c*_{[uφ, EPP]] [_{CP} [who *t*_j]_i [_{iRel}] [_{uT}] [_C C_{[uT, EPP] [uRel, EPP]] [_{TP} *t*_i left *t*_i]]]]]}}

The final picture would be as in (32), which focuses on the EPP property I associate to the φ-features of both *v** and *c*:



So far, nothing has been said about the fact that relative clauses do not show *that*-trace effects when subjects are relativized. I will assess this matter right now, since it is related to Bianchi's (1999) *that/zero*-relatives. The relevant structures are the ones in (14a,c), repeated here as (33a,b) for convenience:

- (33) a. The book [_{CP} *that* I read] *that*-relative
 b. The book [_{CP} ∅ I read] *zero*-relative

In the system I am assuming, *that*-trace effects follow from *that* being a T head, as Pesetsky & Torrego (2001) hold. As for *that*-deletion, it involves the merger of the subject DP with C. This was previously shown in (11), repeated here as (34):

- (34) a. John thinks [_{CP} *that*_j [_C C<sub>[uT, EPP]] [_{TP} Mary *T*_j is gorgeous]]
 b. John thinks [_{CP} Mary_i [_C C_{[uT, EPP]] [_{TP} *t*_i is gorgeous]]}</sub>

All other things being equal, then, one would expect that relativization of subjects produce the same results that moving subjects do elsewhere (e.g., *that*-trace effects and the possibility of dropping complementizers), but things are not equal: no *that*-trace effects obtain and complementizers cannot be dropped²⁴.

- (35) The boy *(*that*) called Mary.

²³ Like *v**, *c* has the property of creating SPECs that go beyond s-selection. Unlike *v**, however, *c* does not seem to display different semantic flavors nor assign Case. Beyond that, notice that nothing really hinges on the label: I use *c*, but it could perfectly turn out to be that the most appropriate one is Bowers's (2001) Pred^o. In fact, if this proposal is on track, *c* and *v** may be simply phasal counterparts of Pred^o.

²⁴ In Bianchi's (1999) system, the anti-*that*-trace effects are explained through a much more complex set of assumptions that rely on a cartographic approach and the government mechanism, unavailable in the current framework. Cf. Bianchi (1999: 231-237) to see the details.

Note that the issue only arises with *that/zero*-relatives, which are analyzed as involving a null relative D by Bianchi (1999), as indicated in (36):

- (36) a. The boy that called Mary.
 b. $[_{DP} \text{The} [_{CP} [_{DP} D_{REL} \text{boy}]_j [_{CP} \text{that} [_{TP} t_j \text{called Mary}]]]]]$
 c. $[_{DP} \text{The}+D_{RELI} [_{CP} [_{DP} t_i \text{boy}]_j [_{CP} \text{that} [_{TP} t_j \text{called Mary}]]]]]$ (at PHON)

Recall that, in Bianchi (1999), *that* corresponds to Rizzi's (1997) Force^o, but we must follow a different route, given what I have been assuming all along (i.e., *that* is a T head). Here I would like to argue that there is a way of accounting for the impossibility of dropping the complementizer in (35) and the lack of *that*-trace effects in a unitary fashion. First, I hold that the operation in (36b) is not possible, since a covert operator cannot pied-pipe lexical material, as argued by Chomsky (2001):²⁵

- (37) [An] EC [Empty Category] disallows pied-piping
 [from Chomsky 2001: 28]

The good news of (37) is that it also accounts for the data in (38): (38b) and (38c) are out because the null relative D cannot pied-pipe the preposition *in*.

- (38) a. The school in which I studied.
 b. *The school in I studied. c. *The school in that I studied.

Things being so, suppose that relative DPs, when headed by a null D, never reach SPEC-C, obligatorily remaining in their first-Merge position. At this point, two questions emerge: 1) how does the head appear before *that*? and 2) how are C's [*uRel*] and [*uT*] deleted? I would like to suggest that the relative head moves to SPEC-*c* in order to delete *c*'s ϕ -Probe; as for the second question, I claim that C's uninterpretable features are deleted as follows: [*uT*] by moving a T head (i.e., *that*) and [*uRel*] by mere *Agree* between [*uRel*] and the null relative D. If the derivation unfolds as just indicated, the lack of *that*-trace effects receive a straightforward answer. Moreover, note that we also derive why *that* must be present: because there is no other way to check C's [*uT*] (the subject DP is too far away this time). The whole process is indicated in (39):

- (39) $[_{CP} \text{boy}_k [_c c_{[\phi, EPP]} [_{CP} \text{that}_i C_{[\text{uT}, EPP]} [\text{uRel}]] [_{TP} T_i [_{DP} D_{REL} t_k]_{[iRel]} \text{saw Mary}]]]]]$

What about cases in which object DPs are relativized? As before, different options are available:

- (40) a. The car $[_{CP} \text{which}]$ John sold] *wh*-relative
 b. The car $[_{CP} \text{that}]$ John sold] *that*-relative
 c. The car $[_{CP} \emptyset]$ John sold] *zero*-relative

²⁵ A reviewer wonders what happens with bare nouns if (37) is correct: how can they be pied-piped? The logic of the proposal forces us to assume that regardless of whether bare nouns are just NPs (cf. Chomsky 2000) or else they contain a DP layer (with possible N-to-D movement), it must be N that gets pied-piped.

The derivations would be roughly as in (41), which already incorporates the additional *cP* layer:²⁶

- (41) a. $[_{DP} \text{The } [_{cP} \text{car}_j \text{ } c_{[u\phi, EPP]}] [_{CP} [\text{which } t_i]_{[iRel]} \text{John}_{z[uT]} C_{[uRel, EPP]} [_{uT, EPP}] [_{TP} t_z \text{ sold } t_i]]]]]$
 b. $[_{DP} \text{The } [_{cP} \text{car}_j \text{ } c_{[u\phi, EPP]}] [_{CP} \text{that}_i C_{[uRel]} [_{uT, EPP}] [_{TP} \text{John } T_i \text{ sold } [_{DP} D_{REL} t_i]_{[iRel]}]]]]]]]$
 c. $[_{DP} \text{The } [_{cP} \text{car}_j \text{ } c_{[u\phi, EPP]}] [_{CP} \text{John}_i C_{[uRel]} [_{uT, EPP}] [_{TP} t_i \text{ sold } [_{DP} D_{REL} t_i]_{[iRel]}]]]]]]]$

Note that, when in SPEC-C, the relative object DP of (41a) can only check C's [*uRel*]; other strategies must be used to delete C's [*uT*], for the [*uT*] feature of object DPs is never alive long enough to do that job.²⁷ As usual, the candidates to delete C's [*uT*] are T itself and the subject DP. However, for reasons that are not clear to me, only the latter possibility yields a correct outcome.²⁸

- (42) a. *The car which that John sold. b. The car which John sold.

Finally, witness how other constituents show different relativization strategies as well. In (43) and (44), we have 'oblique relatives', with and without pied-piping:

- (43) a. The person whom John lives with. *wh*-relative
 b. The person that John lives with. *that*-relative
 c. The person John lives with. *zero*-relative
- (44) a. The person with whom John lives. *wh*-relative
 b. *The person with that John lives. *that*-relative
 c. *The person with John lives. *zero*-relative

(44b) and (44c) are directly ruled out under (37), but the remaining patterns are all possible. Given that I assess 'oblique relatives' in the remainder of this section, I do not delve into the derivational details of (44a) and (43a,b,c).

Let us then go back to the mysterious paradigm in (18)-(19). To begin with, recall Bianchi's (1999) explanation of the problem: English has the two landing sites needed to derive *wh*-relatives —namely, SPEC-Force and SPEC-Topic. Given that her analysis cannot be recast in our terms, an alternative explanation must be found. An empirical fact worth considering in connection with such an asymmetry is preposition stranding: Romance lacks it. However, promising as it may seem at first sight, this cannot be the solution: languages like Bulgarian and Russian, which also lack preposition stranding, display regular *wh*-relatives.

- (45) a. Ira govoriła s mal'čikom, kotoryj govorit po-ispanski. (Russian)
 Ira speak-PST.3SG with boy who speak-PRS.3SG Spanish
 'Ira spoke to the boy who speaks Spanish'

²⁶ According to (37), in (41b) and (41c) I am assuming that the relative DP, being headed by a null D, must stay in its first-Merge position (i.e., its theta-position). If this is so, C's [*uRel*] must be deleted by *Agree*, but then a problem emerges: [*uRel*] has to scan within *v**P's domain, overriding Chomsky's (2000, 2001) *Phase Impenetrability Condition*. A possible way out to this drawback is to suppose covert *internal-Merge* of the relative DP to an outer-SPEC-*v** (cf. Pesetsky 2000 and Nissenbaum 2000): since, strictly speaking, pied-piping is not invoked, the process does not violate (37).

²⁷ Recall that under Pesetsky & Torrego's (2001) proposal, the [*uT*] feature of object DPs is always deleted at the *v**P phase level, so it is by definition impossible for it to delete C's [*uT*].

²⁸ There is still another possibility: C's [*uT*] is deleted by *Agree* alone.

- b. Edin chovek koito govori s Bill. (Bulgarian)
 A person who talk-PST.3SG to Bill
 'A person who talked to Bill'

Crucially for my purposes here, the restriction on *wh*-relatives of Romance languages is also found in one specific environment of English, as noted by Bhatt (1999), Cinque (1982), Huddelston et al. (2002), and Pesetsky (1998): infinitival clauses. As these authors point out, the phenomenon has not received any satisfactory account.²⁹ Consider, in this sense, Bhatt's (1999) surprise when noticing the asymmetry:

With finite relative clauses and contra reduced relatives, object infinitivals permit relative pronouns cf. 14a.

- (14) a. A Knife [[with which]_i C^o [PRO to cut the bread t_i]]
 b. *A knife [[which]_i C^o [PRO to cut the bread with t_i]]
 (compare with *A Knife which John cut the bread with*)
 c. *The book [[which]_i C^o [PRO to read t_i]]
 d. A Knife [Op_i C^o [PRO to cut the bread with t_i]]

However, unlike finite relative clauses, overt material can be present in the [Spec, CP] of an infinitival only if it is part of a pied-piped PP. *It is not well understood why this difference exists between finite relative clauses and object infinitival relatives.* <Emphasis added: AJG>. [from Bhatt 1999: 13]

Huddelston et al. (2002) make the same point:

This construction is limited to somewhat formal style. It is found only with integrated relatives, and is subject to the following severe structural restrictions:

[2]

- i. *The relative phrase must consist of preposition + NP.*
 ii. There can be no expressed subject.

The first restriction excludes examples like **She's the ideal person whom to invite* and **I'm looking for an essay question which to challenge the brighter students with* (where the preposition is stranded rather than being part of the relative phrase). Condition [ii] rules out **She's the ideal person in whom for you to confide*, and the like. *There is no evident explanation for the first restriction*, but the second is predictable from the properties of *wh* relative clauses and infinitivals taken together: infinitivals allow subjects only when introduced by the subordinator *for*, but this cannot occur in *wh* relatives since both it and the relative phrase require to be in initial position. <Emphasis added: AJG>. [from Huddelston et al. 2002: 1067]

²⁹ Bianchi (1999) explains the case of infinitival clauses in a way that is coherent with her proposal: infinitival clauses do not have a Topic Phrase, a projection which is needed in *wh*-relatives' derivation.

In this paper I would like to argue that the asymmetry in (18)-(19) does have to do with a parameter, but not with Bianchi's (1999) *Topic Parameter*. The gist of the analysis I want to put forward runs as follows: subject DPs in Romance languages (and those of English infinitival clauses) can never be moved to SPEC-C to check C's [uT] because their own [uT] has already been deleted (that is, it has not been just marked for deletion, but actually expunged). If attracting a subject DP is not an option, then attracting a PP is the most economical alternative to delete both [$uRel$] and [uT]. The reader may now wonder how a PP can help delete C's [uT]; in this respect, I assume, with Pesetsky & Torrego (2004a), that prepositions are a species of T, a claim that should not be controversial, since, after all, prepositions have usually been taken to be Case-checkers:³⁰

It is also a common observation that elements of the prepositional vocabulary are found in C. This led Emonds (1985: chap. 7) to suggest that the category C be understood as a species of P. Our treatment of English for, however, suggest that such elements are actually instances of T whose presence in C is due to movement—a hypothesis that might be plausibly extended to similar phenomena in other languages. What common property unites members of the supercategory that contains both prepositions and traditional instances of T? We suggest [...] that this supercategory unites those predicates that situate events and individuals in time and space. [from Pesetsky & Torrego 2004a: 510]

If this reasoning is tenable, then there are three candidates to delete C's [uT] in infinitival clauses with a PP that contains a relative D, as indicated in (47):

(46) *Infinitival Relative Clauses*

[_{CP} C [_{$uRel$} , EPP] [_{uT} , EPP] [_{TP} PRO [_{v^*P} ... [_{PP} [_{DP} D_{REL} ...]] ...]]]

(47) *Candidates to Delete C's [uT]*

1. The subject DP (i.e., PRO)³¹
2. T (being spelled-out as a preposition, unless *Agree* is invoked)
3. The PP containing a relative D (assuming Ps are a species of T)

The problem for the first option is rather murky: it seems that PRO (unlike subject DPs in matrix interrogatives and embedded declaratives) cannot be used to de-

³⁰ A reviewer asks a tough question: if P is a species of T, why do we get *do*-insertion even with PP-*wh*-questions? This is true: T-to-C movement occurs in English in those cases too (e.g., *To whom did you give the flowers?*). As I argue in Gallego (2006) this follows from the very analysis I put forward here: if v^*P is a phase, then *wh*-phrases must stop at its *edge* (that is, v^* 's SPECs) in their way to the CP layer, given successive cyclic movement; note that, once in SPEC- v^* , a P pied-piped by a *wh*-phrase is not closer to C than T itself, so T-to-C movement (that is, *do*-insertion) is still compulsory. A completely different scenario is at stake in the case of Null Subject Languages: since TP qualifies as a phase—as I will claim by the end of this section—a P pied-piped by a *wh*-phrase is closer to C than T because it stops at SPEC-T (i.e., the phase *edge*), not SPEC- v^* , so T-to-C movement can be blocked, and it is indeed, as I show in Gallego (2006). The facts, therefore, provide additional support for my analysis.

³¹ Due to space limitations I cannot consider the issue of whether a raising analysis of control (cf. Hornstein 2003) is relevant for the facts under discussion.

lete C's [μ T]. In fact, this might be related to the general impossibility of moving the subject of an infinitival clause, in both Spanish and English:³²

- (48) a. *No sé quién comprar los libros. (Spanish)
 Not know-PRS.1SG who buy-INF the books
 'I don't know who to buy the books'
 b. *I wonder who to solve the problem.

The second option (i.e., moving T) is also useless: it would require the appearance of the prepositional complementizer *for*, which, in turn, seems to force the presence of an overt subject DP, conflicting with PRO and its 'Null Case' (or whatever is responsible for its special behaviour; cf. fn. 30):

- (49) a. [_{CP} For_i C_[μ T, EPP] [_{TP} Mary to_i win the lottery]] would be great.
 b. * [_{CP} For_i C_[μ T, EPP] [_{TP} PRO to_i win the lottery]] would be great.

The remaining candidate is the only possibility left: moving the oblique relative phrase is the only option for infinitival relatives to converge. But why? I want to argue that the answer lies in the economy principles that rule the computational system. If a PP is attracted to C, all its uninterpretable features can be deleted at once: P, being a species of T, deletes C's [μ T], while the relative D deletes C's [μ Rel].

Let us shift our attention to Romance languages, and, more specifically, to Spanish. Consider the relevant asymmetry one more time: *wh*-relatives must be introduced by a preposition.

- (50) a. *El hombre {quien/el cual} habló. (Spanish)
 The man {who/the which} talk-PST.3SG
 'The man who talked'
 b. *El libro el cual Juan leyó. (Spanish)
 The book the which Juan read-PST.3SG
 'The book which John read'
- (51) a. El hombre con quien habló. (Spanish)
 The man with whom talk-PST.1SG
 The man I talked to'
 b. El hombre a quien vi. (Spanish)
 The man to who see-PST.1SG
 'The man who I saw'

As I see it, there are three possible causes for this:

- (1) The relative DP *quien hombre* or *el cual hombre* (Eng. {*who/the which*} *man*) cannot be generated.

³² Cf. Pesetsky & Torrego (2001: 416, fn. 69) for discussion. I put to the side facts like (i), noted by Torrego (1996), since they deserve a more careful consideration:

- (i) No sabemos {quiénes/cuáles/cuántos} leer este libro. (Spanish)
 Not know-PRS.1PL {who/which-ones/how-many} read-INF this book
 'We do not know {who/which ones/how many}-of us read this book'

- (2) *Quien hombre* can be generated, it moves to SPEC-C, but then *hombre* cannot be subextracted.
 (3) *Quien hombre* can be generated, but it never reaches SPEC-C.

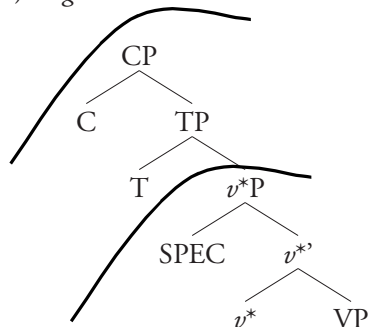
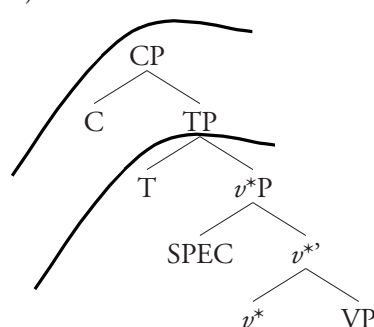
Here I argue that (3) correctly describes the problem. In Pesetsky & Torrego's (2001) system subject DPs' [uT] features can remain 'alive' until the CP is built up. I argue that that of Spanish subject DPs cannot; this would explain why Spanish lacks the patterns in (52), since they both involve merging the subject DP in SPEC-C to check C's [uT] (note that in (52a) this implies that *How intelligent* is in an outer-SPEC-C; as for (53b), cf. (11) in section 1).

- (52) a. [_{CP} How intelligent_i [_{iWh}] [_{CP} Mary_j [_{iTT}] C_[_{iWh}, EPP] [_{iTT}, EPP] [_{TP} t_j is t_i]]] !
 b. I know [_{CP} John_i [_{iTT}] C_[_{iTT}, EPP] [_{TP} t_i called her]]

Compare (52) with their Spanish word-by-word translations in (53) -as expected, they are impossible, for the Case feature of *María* and *Juan* cannot delete C's [uT]:

- (53) a. *_iQué inteligente María es! (Spanish)
 b. *Sé Juan la llamó. (Spanish)

If all this is on the right track, then T must be the *Locus* of the asymmetry. If the T feature of subject DPs in Spanish become derivationally 'dead' by the time the CP is being assembled, this must mean that TP, and not v^*P , is a strong phase in Romance languages. Consequently, the phase systems of English and Romance languages would be as depicted in (54):

(54a) *English*(54b) *Romance*

Note that (54) does not claim that Romance has more phases than English, but rather that the v^*P phase is somehow 'pushed up' to the TP level.^{33 34} If correct, phe-

³³ Therefore, all languages have two strong phases. This said, it is not so clear that being 'propositional' is what defines 'phasehood' (cf. Chomsky 2000). Actually, the facts seem to support a view under which morphological 'convergence' is the relevant criterion (cf. Uriagereka 1999a); in particular, note that TP is the minimal domain in which all case features are assigned a value. Accordingly, TP is, case-wise, a convergent domain. Cf. Gallego (2006) for a development of this idea.

³⁴ A similar conclusion was reached by Rizzi (1982), who phrased his claim in terms of 'bounding nodes'. For more related proposals that ultimately signal to a similar parameter, cf. Gallego (2006), where I argue that this 'pushing up' is related to head movement, a controversial claim given the alleged phonological status of this operation (cf. Chomsky 2001).

nomena like *that*-trace effects, clitic climbing (as discussed in Kayne 1989),³⁵ subject inversion, and, interestingly, the relativization patterns of Romance languages regarding *wh*-relatives can receive a unitary account: since subject DPs cannot be attracted to C to delete its [*u*T] in Romance, the only way for the intended derivations to converge is by moving a T element: either T itself or a preposition.

All in all, we can stop this section here. The main goal of the preceding lines was to provide an analysis of relative clauses under Pesetsky & Torrego's (2001) system, paying special attention to the asymmetry in (18)-(19). As I have tried to show, those facts are not as isolated as one might think: on the contrary, they are closely related to an important parameter which has T (and Case Theory) as its *Locus*.

4. Conclusions and possible extensions of the analysis

In this paper I have put forward a minimalist analysis of relative clauses that endorses Pesetsky & Torrego's (2001) proposal concerning the syntactic interaction between C and T, and the nature of Case. I have reviewed the main aspects of the so-called 'raising analysis' of relative clauses, focusing on Kayne's (1994) treatment and Bianchi's (1999) subsequent modifications. It has been claimed that relative clauses do involve *internal-head-Merge*, and an extra functional head creating predication as well (i.e., a little *c*). The analysis has departed from government and cartographic based approaches, arguing that Chomsky's (2001) generalization about empty categories can explain the absence of *that*-trace effects in relative clauses. Finally, a new account for two long-standing asymmetries between English and Romance languages has been presented, one that capitalizes on the notion of *phase* (a hallmark of the Minimalist Program); in particular, I have argued that Case convergence obtains in an earlier derivational stage in Romance languages than it does in English: the [*u*T] of subject DPs is marked for deletion in SPEC-*v**, not SPEC-T (cf. Uribe-Etxebarria 1992), which renders it inactive for computational affairs in the CP phase. The strongest conclusion which one can arguably draw, therefore, is that phases (or computationally convergent domains) do not behave in a uniform way cross-linguistically. Note that although this may in principle seem problematic, it is still sound within minimalism, for it is consistent with the possibility that, besides Chomsky's (2000) conceptual motivation (i.e., reduction of computational load), phases may emerge by bare output demands, hence supporting the *Strongest Minimalist Thesis* that language is an optimal solution to interface conditions.

To conclude, I would like to speculate one possible extension of the analysis which concerns the 'island' status of relative clauses. Consider first the facts:

- (55) a. *Which boy_i did Mary talk to [_{DP} the [_{CP} person [_{CP} who saw t_i]]]?
 b. *Where_i did you see [_{DP} the [_{CP} boy [_{CP} who works t_i]]]?

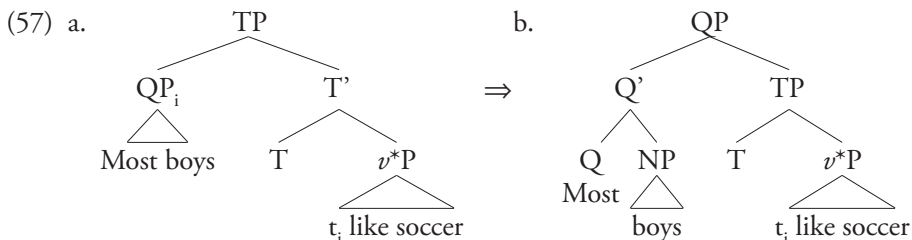
As (55) shows, relatives behave as 'strong islands' (cf. Cinque 1990 and Stepanov 2001), hence barring all types of movement. Under Chomsky's (2004) analysis,

³⁵ Recall that Kayne (1989) argued that T (at that time, INFL) was able to L-mark the VP in Romance so that 'barrierhood' of VP was eliminated, allowing clitic climbing. As far as I can see, this is perfectly coherent with what I am saying, since 'barrierhood' shares obvious properties with 'phasehood'.

the problem in (55) would trivially follow from adjuncts being placed in a “parallel plane” within Narrow Syntax: since pair-Merge (the operation dealing with adjunction) is designed in such a way that it eliminates all canonical dependencies (e.g., dominance, c-command, etc.), no Probe-Goal dependency can be established, and, therefore, extraction from within the relative clause becomes impossible. Although this account seems plausible at first glance, it must be noted that it is rather unlikely that the argument-adjunct asymmetry plays a real role with regards to extraction. I say this because movement out of a complement CP is also barred:

- (56) a. *Who_i did John like [_{DP} the idea [_{CP} that people should vote t_i]]?
 b. *Why_i will Mary listen to [_{DP} the proposal [_{CP} that John must be killed t_i]]?

(55) and (56) clearly suggest that the relevant factor is the ‘nominal’ nature of the structures.³⁶ In this respect, I would like to speculate that the internal-head-Merge analysis which I have assumed in this paper might shed some light on these facts. In particular, I would like to suggest that once the relative head has reached the *cP* layer, it triggers a process of syntactic type-shifting which Hornstein & Uriagereka (2002) dub *Reprojection*. In their proposal, Hornstein & Uriagereka (2002) focus on binary quantifiers (e.g., *all*, *most*, etc.), which, for the right semantics to obtain, must provoke a ‘relabelling’ at LF by which they are able to take the TP as their second argument (i.e., their nuclear scope). Roughly, the details are as in (57): first the QP *Most boys* raises to SPEC-T, and then the Q head ‘relabels’ the whole structure.



Given the logic of (57) one might wonder why *Reprojection* applies only at LF. Hornstein & Uriagereka (2002) argue that this derivational delay prevents computational tampering concerning (a) Chain Uniformity and (b) Checking Domains (in Chomsky’s (1993, 1995) sense). Furthermore, since *Reprojection* is a semantics-driven operation, it is sound for it to apply at LF (i.e., the SEM component), where effects like (58), the ones Hornstein & Uriagereka (2002) want to capture, take place:

³⁶ Admittedly, this says nothing about why simple nominal complements (i.e., all non-specific object DPs) are not islands.

[_{CP} Of which city_i did John buy [_{DP} a picture t_i]]?

I know of no explanation for the contrast between the data in (56) and (i). Plausibly, Richard’s (2005) analysis on extraction provides a solution: in his system, only complement CPs which *Agree* with a higher *v** allow extraction. Obviously, relative CPs differ from both complement CPs and regular object DPs in that they do not agree with any functional category, remaining “opaque” for extraction.

(58) *Nobody gave most children a red cent.

[from Hornstein & Uriagereka 2002: 110]

In (58) the desired LF licensing relation between *Nobody* and the NPI *a red cent* is blocked by the QP *most children*, which induces an intervention effect. Importantly, note that the problem goes away if we use a unary quantifier such as *two*, for it does not need to trigger *Reprojection* (cf. (59a)). Also, as (59b) indicates, the process does qualify as a covert one, for otherwise extraction of *What* should be out:

(59) a. Nobody gave two children a red cent.

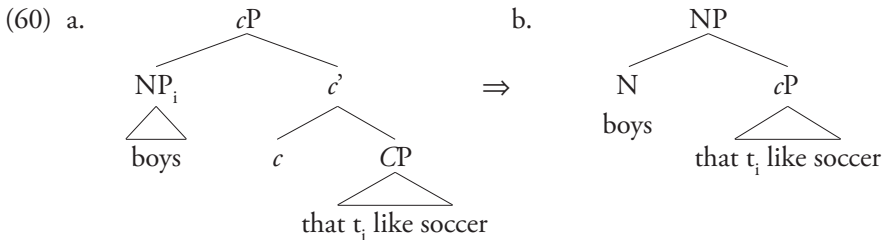
b. What_i did nobody give most children t_i ?

[from Hornstein & Uriagereka 2002: 110]

(58) and (59), then, support an analysis of LF-islands along the lines of Hornstein & Uriagereka's (2002): when binary quantifiers remerge as SPECs of T, we get a configuration in which they can take the TP as a regular dependent by means of *Reprojection*, turning it into a complex SPEC which is rendered out-of-sight for LF dependencies (e.g., *Agree*, *Attract*, XP covert movement, etc.).

There are grounds, however, to disregard the technical problems which force Hornstein & Uriagereka (2002) to delay *Reprojection* until LF. On the one hand, Chain Uniformity could be obviated if the system is strongly derivational and can 'forget' about immediately previous steps; on the other hand, the preservation of Checking Domains is no longer needed once their primitive status has been rejected (cf. Chomsky 2000).

This said, suppose *Reprojection* can apply as the derivation unfolds. At some point, the crucial step would be as indicated in (60), irrelevant details omitted:



Note that the overt *Reprojection* in (60) can buy us what we want: the *cP* becomes a complex SPEC (an island; cf. Huang 1982 and Uriagereka 1999b) within Narrow Syntax. If the technical problems Hornstein & Uriagereka (2002) note can be put aside as I just said, then we arrive at a quite clean account for why relatives are strong islands; the analysis, moreover, has the advantage of accounting for the fact that the relative clause (like any other adjunct) behaves as if it was not there apart from semantic interpretation, which allows us to capture the fact that the entire construction has a nominal nature. There is, nevertheless, an important problem for (60): the process is unmotivated. In the cases Hornstein & Uriagereka (2002) discuss, this matter does not even arise, for binary quantifiers must always take their second argument, so *Reprojection* is welcome (actually, it is needed). It is tempting to argue that the head-raising analysis of relative clauses only involve heads, for then we could claim that there is a process of 'projection' of the element that undergoes internal-Merge (in Donati's 2004 terms; cf. also Chomsky 2005); however, as we see, NP can

also do the job. I leave this question unsettled with no useful comment. The possibility of resorting to overt *Reprojection* processes seems to me to be a fair move (within certain limits, of course), but I realize that, in (60), a coherent motivation for it to apply is lacking.

References

- Arregui, K., 1998, «Spanish *el que* Relative Clauses and the Doubly Filled COMP Filter». Ms., MIT.
- Belletti, A., 2004, «Aspects of the Low IP Area», *The Structure of CP and IP*, ed. by L. Rizzi, 16-51. Oxford (New York): OUP.
- Bhatt, R., 1999, *Covert Modality in Non-Finite Contexts*. Ph.D. Dissertation, U.Penn.
- 2002, «The Raising Analysis of Relative Clauses: Evidence from Adjectival Modification», *Natural Language Semantics* 10. 43-90.
- Bianchi, V., 1999, *Consequences of Antisymmetry. Headed Relative Clauses*. Berlin: Mouton de Gruyter.
- 2000, «The Raising Analysis of Relative Clauses: A Reply to Borsley (1996)», *LI* 31. 123-140.
- 2002a, «Headed relative clauses in generative Syntax - Part I», *Glott International* 6. 197-204.
- 2002b, «Headed relative clauses in generative Syntax - Part II», *Glott International* 6. 235-247.
- Boeckx, C., 2003a, *Islands and Chains*. Amsterdam: John Benjamins.
- 2003b, «Case Matters and Minimalist Concerns», *Harvard Working Papers in Linguistics* 8. 161-197.
- 2004, «Long-distance agreement in Hindi: some theoretical implications», *Studia Linguistica* 58. 1-14.
- Bowers, J., 2001, «Predication». *The Handbook of Contemporary Syntactic Theory*, ed. by M. Baltin & C. Collins, 299-333. London: Blackwell.
- Brame, M., 1968, *A new analysis of relative clauses: Evidence for an interpretive theory*. Ms., MIT.
- Bresnan, J., 1972, *Theory of Complementation in English Syntax*. Ph.D. Dissertation, MIT.
- Brucart, J. M., 1992, «Some Asymmetries in the Functioning of Relative Pronouns in Spanish», *Catalan Working Papers in Linguistics* 1. 113-143.
- Carlson, G., 1977, «Amount Relatives», *Lg* 53. 520-542.
- Chomsky, N., 1965, *Aspects of the Theory of Syntax*. Cambridge (Mass.): MIT Press.
- 1986, *Barriers*. Cambridge (Mas.): MIT Press.
- 1993, «A Minimalist Program for Linguistic Theory». *The View from Building 20: Essays in Linguistics in Honour of Sylvain Bromberger*, ed. by K. Hale & S. J. Keyser, 1-52. Cambridge (Mass.): MIT Press.
- 1995, *The Minimalist Program*. Cambridge (Mass.): MIT Press.
- 2000, «Minimalist Inquiries: The Framework». *Step by Step. Essays on Minimalist Syntax in Honour of Howard Lasnik*, ed. by R. Martin, D. Michaels, & J. Uriagereka, 89-155. Cambridge (Mass.): MIT Press.
- 2001, «Derivation by Phase». *Ken Hale: A Life in Language*, ed. by M. Kenstowicz, 1-52. Cambridge (Mass.): MIT Press.

- 2004, «Beyond Explanatory Adequacy». *Structures and Beyond*. ed. by A. Belletti, 104-131. Oxford/New York: OUP.
- 2005, «On Phases». Ms., MIT.
- Cinque, G., 1982, «On the Theory of Relative Clauses and Markedness», *The Linguistic Review* 1. 247-294.
- 1990, *Types of A' Dependencies*. Cambridge (Mass.): MIT Press.
- Citko, B., 2001, «Deletion under Identity in Relative Clauses», *Proceedings of NELS* 31, ed. by M. Kim & U. Strauss, 131-145. Amherst (Mass.): GLSA.
- Den Besten, H., 1983, «On the interaction of root transformations and lexical deletive rules». *On the Formal Syntax of Westgermania*, ed. by W. Abraham. Amsterdam: John Benjamins.
- Donati, C., 2004, «On *Wh*-head-movement», Ms., U. Urbino.
- Frampton, J. & S. Gutmann, 2000, «Agreement is Feature Sharing». Ms., Northeastern University.
- Gallego, Á. J., 2006, «Phase Sliding», Ms., UAB/UMD.
- Gutiérrez-Rexach, J. & E. Mallén, 2003, «Expletive Definites in Relative Clauses». *Selected Papers from the 5th Hispanic Linguistics Symposium*, ed. by F. Ordóñez. Somerville (Mass.): Cascadilla Press.
- Hiraiwa, K., 2001, «Multiple Agree and the Defective Intervention Constraint in Japanese». *MIT Working Papers in Linguistics* 40. 67-80.
- Hornstein, N., 2003, «On Control». *Minimalist Syntax*, ed. by R. Hendrick, 6-81, Oxford: Blackwell.
- & J. Uriagereka, 2002, «Reprojections». *Derivation and Explanation in the Minimalist Program*, ed. by S. Epstein & T. Seely (eds.), 106-132.
- Huang, J., 1982, *Logical relations in Chinese and the theory of grammar*. Ph.D. Dissertation, MIT.
- Huddelston, R. & G. K. Pullum, 2002, *The Cambridge Grammar of the English Language*. Cambridge: CUP.
- Kayne, R. S., 1989, «Null Subjects and Clitic Climbing», *The Null Subject Parameter*, ed. by O. Jaeggli & K. Safir, 239-261. Dordrecht: Kluwer.
- 1994, *The Antisymmetry of Syntax*. Cambridge (Mass.): MIT Press.
- 2000, *Parameters and Universals*. New York: OUP.
- Koopman, H., 1983, «ECP effects in main clauses», *LI* 14. 346-350.
- Koster, J., 2003, «All languages are tense second». Ms., University of Groningen.
- Law, P., 2000, «On Relative Clauses and the DP/PP Adjunction Asymmetry». *The Syntax of Relative Clauses*, ed. by A. Alexiadou *et al.*, 161-199. Amsterdam: John Benjamins.
- Lebeaux, D., 1991, «Relative clauses, licensing, and the nature of the derivation», *Syntax and Semantics 25: Perspectives on Phrase Structure*, ed. by S. Rothstein, 209-239. New York: Academic Press.
- Longobardi, G., 1994, «Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form», *LI* 25. 609-665.
- Nissenbaum, J., 2000, *Investigations of covert phrase movement*, Ph.D. Dissertation, MIT.
- Ojeda, A. I., 1992, *Los sintagmas relativos en inglés y en español*. Oviedo: U. Oviedo.
- Pesetsky, D., 1998, «Some Optimality Principles of Sentence Pronunciation». *Is the Best Good Enough?: Optimality and Competition in Syntax*, ed. by P. Barbosa *et al.*, 337-383. Cambridge (Mass.): MIT Press.

- 2000, *Phrasal Movement and Its Kin*, Cambridge (Mass.): MIT Press.
- & E. Torrego, 2001, «T-to-C Movement: Causes and Consequences». *Ken Hale: A Life in Language*, ed. by M. Kenstowicz, 355-426. Cambridge (Mass.): MIT Press.
- & — 2004a, «Tense, Case, and the Nature of Syntactic Categories». *The Syntax of Time*, ed. by J. Guéron & J. Lecarme, 495-537. Cambridge (Mass.): MIT Press, 495-537.
- & — 2004b, «The Syntax of Valuation and the Interpretability of Features». Ms., MIT/U. Mass Boston.
- Richards, N., 2005, *Person-Case Effects in Tagalog and the Nature of Long Distance Extraction*, Ms., MIT.
- Rizzi, L., 1982, *Issues in Italian Syntax*, Dordrecht: Foris.
- 1990, *Relativized Minimality*. Cambridge (Mass.): MIT Press.
- 1997, «The Fine Structure of The Left Periphery». *Elements of Grammar*. ed. by L. Haegeman, 281-337. Dordrecht: Foris.
- Sauerland, U., 2000, «Two Structures for English Restrictive Relative Clauses». Ms., Tübingen University.
- Schachter, P., 1973, «Focus and Relativization», *Lg* 49. 19-46.
- Stepanov, A., 2001, *Cyclic Domains in Syntactic Theory*. Ph.D. Dissertation, U. Conn.
- Svenonius, P., 2002, «Case is uninterpretable aspect». Ms., University of Tromsø.
- Szczegielniak, A., 2004, *Relativization and Ellipsis*. Ph.D. Dissertation, Harvard University.
- Toribio, A., 1992, «Proper Government in Spanish Subject Relativization», *Probus* 4. 291-304.
- Torrego, E., 1996, «On Quantifier Float in Control Clauses», *LI* 27. 111-126.
- Uriagereka, J., 1999a, «Minimal Restrictions in Basque Movements», *NLLT* 17. 403-444.
- 1999b, «Multiple Spell-out». *Working Minimalism*, ed. by N. Hornstein & S. Epstein, 251-282, Cambridge (Mass.): MIT Press.
- Uribe-Etxebarria, M., 1992, «On the Structural Positions of the Subject in Spanish, their Nature and their Consequences for Quantification». *Syntactic Theory and Basque Syntax*, ed. by J. Lakarra & J. Ortiz de Urbina, 447-491. San Sebastián: ASJU.
- Vergnaud, J. R., 1974, *French Relative Clauses*. Ph.D. Dissertation, MIT.

HOW STRONG ISLANDS ARE DERIVED FROM THE WAY A TOP-DOWN DERIVATION IS LINEARIZED

Valérie Gautier

University of Nantes

The goal of this paper is to show that strong islands can be derived from the way derivation is linearized, as long as we assume that the derivation proceeds in a top-down fashion.

To begin with, I will present one advantage of adopting a top-down approach regarding linearization issue: the Linear Correspondence Axiom (henceforth, LCA, Kayne 1994) can be reformulated in a more derivational and minimalist way. In particular, Kayne assumes that the notion that derives precedence is the asymmetric c-command. Because of the asymmetric c-command, the LCA rules out the head-complement configuration in bare phrase structures. I show that with a top-down derivation, the problematic asymmetric c-command relation can be eliminated. With our *Top-down LCA*, precedence relations are derived from the way phrase structures have been built: roughly, if Y enters the derivation right after X, then X precedes Y. In a second part, I will argue that this non-standard approach to derivation¹ and linearization can capture CED effects: it will be demonstrated that subjects and adjuncts are islands because they have to be built in a parallel derivation.

1. A top-down derivation

In the generative framework, it is taken for granted that derivations proceed bottom-up. Roughly, a bottom-up derivation starts with the structurally lowest constituent in a clause and ends with the structurally highest one (such as subject or fronted elements). As an illustration, consider the derivation of a simple sentence like (1).

- (1) La fille danse.
The girl dances.

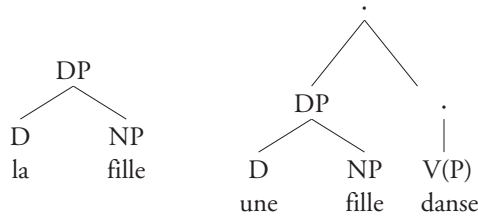
¹ The idea that a top-down derivation can derive CED effects has been first proposed by Boeckx (1999).

(2)

V(P) is the first
element entering
the derivation
V(P) danse

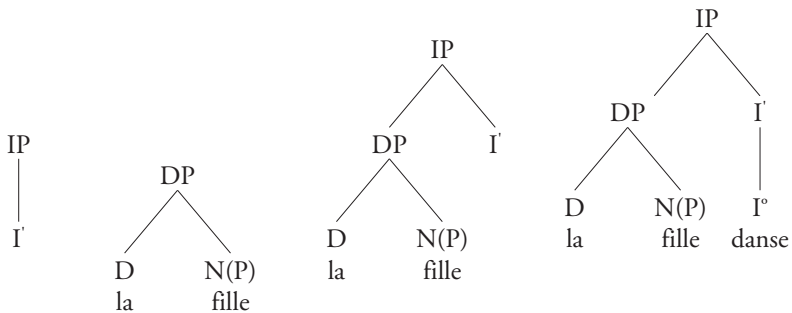
The subject DP is
assembled in parallel.

DP is merged with VP



Following Phillips (1996, 2001), I claim that the derivation should proceed in a top-down fashion. This non-standard assumption entails that the derivation of (1) starts with I(P) and ends with V(P). The top-down derivation of (1) is given below.

- a) IP-I' is the first element entering the derivation
 b) To check I' nominative features, DP subject is needed. [la fille] is then built in parallel
 c) DP is merged with I'. Nominative features are checked
 d) The inflexion enters the derivation.



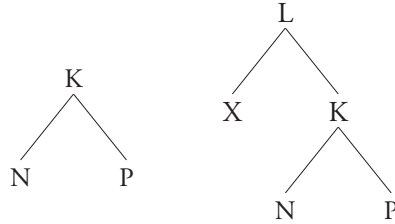
Bottom-up vs Top-down Merge

From a minimalist perspective, the phrase structure of a given sentence is built by the two structure-building operations *Merge* and *Move*. In a bottom-up approach, *Merge* is assumed to apply at the root² of the tree, that is, *Merge* is a cyclic operation. (3) is the definition of cyclic *Merge*.

² « Merge always applies in the simplest possible form: at the root ». (Chomsky 1995)

(3) *Cyclic Merge*

Merge X and K yields the new constituent L. L dominates the tree.

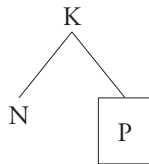


In a top-down derivation, *Merge* crucially does not increment the tree at the root: each time a new item is inserted in the derivation, it replaces something in the structure. *Merge* in a top-down derivation is then a counter-cyclic operation. I define this counter-cyclic operation as *Top-down Merge*. *Top-down Merge* is given in (4) and illustrated in (5) and (6).

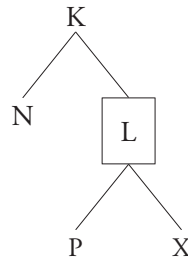
(4) *Top-down Merge*

- a) Applied to α and β , Merge forms the new object $\{L, \{\alpha, \beta\}\}$ by merging α and β .³
- b) L replaces the last terminal node merged in the structure.

(5)



(6)



As an illustration, let us see how the VP in (7) is concatenated.

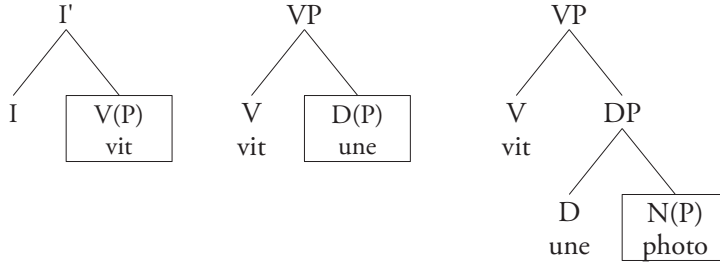
(7) [vit une photo de Marie]
saw a picture of Mary

1
At this step, V(P) is the last terminal node merged

2
D(P)*une* enters the derivation
a) VP [vit une] is built,
b) VP [vit une] replaces V(P) [vit]

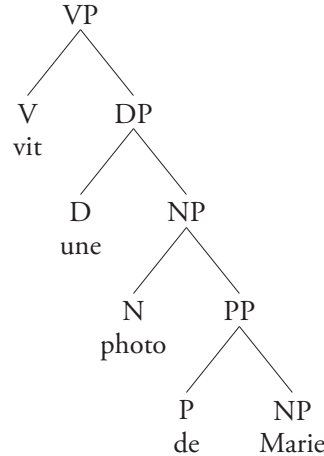
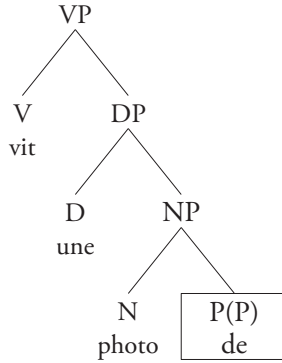
3
N(P)*photo* enters the derivation
a) DP [une photo] is built
b) DP [une photo] replaces D(P)*une*

³ This definition has been inspired by the definition of counter-cyclic *Merge* given by Kitahara (1995).



- 4
 P(P)*de* enters the derivation
 a) NP [photo de] is built,
 b) NP [photo de] replaces N(P) [photo]

- 5
 N(P)*Marie* enters the derivation
 a) PP [de Marie] is built,
 b) PP[de Marie] replaces P(P)*de*



Several arguments have already been given in favor of a top-down syntax.⁴ In particular, Phillips (1996, 2001) showed how a top-down derivation⁵ explains why constituency tests sometimes yield contradictory results. As Phillips pointed out, building the sentence in a top-down way entails that the insertion of an item in the derivation can destroy the preceding constituent to create a new one.⁶ That is, contrary to a bottom-up approach, constituent structure is not permanent. As a consequence, when the constituency test's contradictions are mysterious in a bottom-up approach, there are expected in a top-down derivation.

As an illustration, consider (8) and (9) below.

⁴ See Boeckx (1999), Richards (1999), Guimaeres (2004), Drury (2005), for instance.

⁵ Note that Phillips also assumes that the derivation proceeds from left-to-right. I am not making this assumption here

⁶ For instance, in the derivation of (7), the insertion of the N(P)*Marie* destroys the constituent NP*photo de*.

(8) John *talked to* and *gossiped about* the kid who sprayed paint on his car.

(9) Helen *talked to* Jonathan and Alice did * (to) Matthew.

Building the structure in a top-down way entails that in (8), [talked to] is still a constituent as it has not been merged with a argument. Thus, [talked to] can be coordinated. In contrast, in (9), the addition of *Jonathan* causes destruction of the constituent [talked to]. As the verb and the preposition can no longer form a unit, [talked to] cannot be deleted in (9) (for details, see Phillips 2003). In the next section, I'm going to present another strong argument for a top-down syntax. One advantage of a top-down derivation is that it provides us a way to reformulate the LCA into an axiom which fits in a minimalist approach.

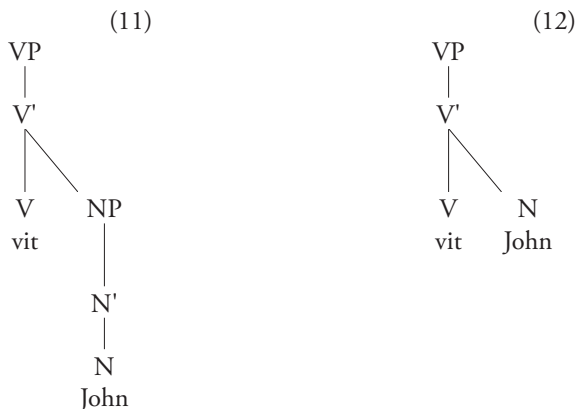
2. Why the LCA should be reformulated

Kayne (1994) argues that there should exist a correspondence between hierarchical relations in a phrase structure and the linear order between terminal nodes. Kayne's idea is that this correspondence is given by the asymmetric *c*-command relation. He formalized this intuition with the LCA.

(10) Linear Correspondence Axiom (LCA)

A lexical item α precedes a lexical item β if α asymmetrically *c*-commands β .

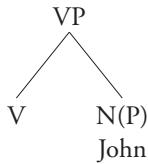
It should be reminded to the reader that Kayne's proposal has been elaborated in the Government and Binding framework. Kayne argues that (10) derives the properties of one of the central GB's module: X-bar theory. Crucially, the LCA justifies the fact that in X-bar schemata, the complement of a head cannot be another head. As an illustration, let us compare (11) and (12). In (11), the complement of the verb is the maximal projection NP, with the three X-bar levels. Note that V *c*-commands N, but N does not *c*-command V. By (10), V precedes N. (11) is a licit configuration for the LCA. In (12), the complement of the verb is the head N° *John*. Here, no order can be established between V and N because V and N are not in an asymmetric relation (V and N being dominated by the same node). (12) is illicit for the LCA.



Since GB model, progress in the field of constituency description has been made, leading to a discussion of the adequacy of the X-bar template. In particular, the relevance of intermediate vacuous projections, such as N' in (11), is questioned.

As a consequence, in the minimalist framework, X-bar theory is replaced by the bare phrase structure. In this new approach to phrase structure, lexical items selected in the numeration, have a double status: they are minimal and maximal projections. This means that the head-complement configuration in (11) is replaced by the head-complement configuration in (13), where *John* is dominated by a minimal and maximal projection node.

(13)



Note that in adopting (13), the head-complement configuration becomes a phrase structure ruled out by the LCA. That is, (13) is (12): the complement of the head V is another head.

In brief, in the X-bar theory, a head's complement has to be a maximal projection. With the asymmetric c-command relation, the LCA derives this property. In the bare phrase structure, a head's complement can be another head. By ruling out this syntactic configuration, the LCA does not derive this property. I conclude that the LCA is not an appropriate principle for the new phrase structure formalism adopted by the generative model. To solve this compatibility problem, I suggest the reformulation of the LCA. The reasoning is the following: as X-bar theory has been replaced by bare phrase structure, the LCA should be replaced by a new Linear Correspondence Axiom, which fits with the bare phrase structure formalism.

3. Top-down LCA

To see what should be changed in the LCA, we need to know what in Kayne's axiom, repeated in (14), makes the structure in (13) illicit.

(14) Linear Correspondence Axiom (LCA)

A lexical item α precedes a lexical item β if α asymmetrically c-commands β .

(13) is bad because V and N stand in a mutual c-command relation. It just so happened that the relation which derives precedence is the asymmetric c-command relation. What needs to be changed seems quite obvious: the asymmetric c-command should be eliminated from the LCA. What should be then the relation that maps hierarchical relations onto linear order? Note that the simple c-command relation is not good either since V c-commands N and N c-commands V. I suggest that in order to reformulate the LCA into a satisfying minimalist principle, we need a deriva-

tional axiom which says that V precedes N because V entered the derivation before N. The only way to get such an axiom is to assume that the derivation proceeds in a top-down way. I propose (15) as the reformulation of the LCA.

(15) *Top-down LCA*

At the derivational step n , a lexical item α precedes immediately a lexical item β iff

α c -commands β

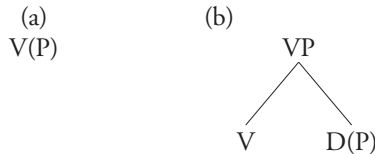
β is the last terminal node merged at step n

α was the last terminal node merged at step $n - 1$.

NB: β and α are terminal nodes

(15) entails that c -command relations map onto precedence relations in a derivational way: at each derivational step, the last terminal node α merged in the tree precedes the new terminal node β inserted in the derivation. With (15), we do not need the asymmetric c -command anymore: when a term β arrives in the derivation after a term α , then β cannot precede but will automatically follow α (even if, representationally, α and β are in a mutual c -command).

It is shown in a) and b) below how (15) derives the precedence relation for a head-complement configuration: The complement $D(P)$ follows the head V , because the lexical item $D(P)$ has been inserted in the derivation after the lexical item $V(P)$.



As a concrete illustration, let's go back to the derivation of [vit une photo de Marie], to see how the new LCA works.

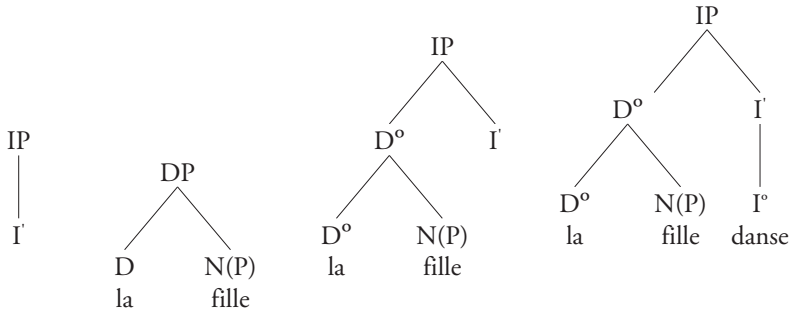
Step 2: the terminal node $D(P)une$ is merged with the terminal node $V(P)vit \rightarrow D(P)une$ follows $V(P)vit$. Step 3: the terminal node $N(P)photo$ is merged with the terminal node $D(P)une \rightarrow N(P)photo$ follows $D(P)une$. Step 4: the terminal node $P(P)de$ is merged after the terminal node $N(P)photo \rightarrow P(P)de$ follows $N(P)photo$. Step 5: the terminal node $N(P)Marie$ is merged after the terminal node $Pde \rightarrow N(P)Marie$ follows Pde . At this point of the discussion, we have demonstrated how (15) derives linear order for terminal nodes in a head-complement configuration. I left aside the question of linearization of specifiers and adjuncts. In the second part of this paper, this problem will be considered. In particular, I am going to show that specifiers and adjuncts have a special status regarding linearization, that is, when they enter the derivation, (15) cannot apply. I will argue that the Constraint on Extraction Domain (CED, Huang 1982) can be derived from this failure.

4. Why adjuncts and subjects are islands: a top-down explanation

The specifier position will be the first case examined here. Let us go back to the top-down derivation of (1), repeated in (16).

- (16) Une fille danse.
The girl dances.

- a) IP-I' is the first element entering the derivation b) To check I' nominative features, DP subject is needed. [la fille] is built in a parallel derivation. c) DP is merged with I'. Nominative features are checked d) The inflexion enters the derivation.



Note that before entering the derivation, DP has been spelled-out to become a D°. The “reduction” of the maximal projection into a terminal node is necessary given the assumption that only terminal nodes can enter a top-down derivation.⁸

At step c), the ordering relation between DP and I' is unspecified. Remember that (15) gives the precedence relation between two terminal nodes: since I' is not a terminal node, (15) cannot apply. The spec-head order is finally obtained at step d), once I° enters the derivation: D° c-commands I° and I° has entered the derivation after D°, then D° precedes I°. I claim that the way the subject is built in a top-down derivation explains subject's islandhood. It is well-known, that DPs cannot be extracted from a subject position. As an illustration, consider (17).

- (17) *Who_i does a picture of t_i upset Mary ?

Remember that the subject is built in a parallel derivation. This entails that the wh-phrase “who” and the subject DP “a picture of” belong to two distinct derivational workspaces. “who” belongs to (i) and “a picture of” belongs to (ii).

⁷ This assumption is in the spirit of Uriagereka and Nunes (1999)'s Multiple Spell-out Model.

⁸ The introduction of IP-I' being an exception in this respect.

- (i) */Who does /*SUBJECT*/ upset Mary/
- (ii) /a picture of/

I suggest that DPs must check their thematic features in their own derivational workspace.⁹ This proposal is formulated in (18).

- (18) A DP must check its theta-role (see Hornstein 1998) in its workspace. Otherwise, the derivation crashes.

In (17), —“who” cannot check its theta-role with “upset”, since “upset” already has an argument. The only possibility for “who” to check its theta-role is then with the preposition “of”. As “who” and “of” do not belong to the same workspace, the derivation crashes. The ungrammaticality of (17) is correctly predicted by (18). Note that (18) predicts that (17) should be good if “upset” has no overt argument. This prediction is correct: (19) is grammatical because the complement position of “upset” has been filled with a trace. The reason why a parasitic gap construction legitimates an extraction out of a subject is then intuitively explained.

- (19) Who_i does a picture of t_i upset t_i ?

As there is no overt argument for “upset”, a copy of “who” can be inserted in complement’s position of “upset”. “who” can then check its theta-role in its derivational workspace, the derivation does not crash. A copy of “who” is created in the complement position of “of” to check the selectional properties of the preposition.

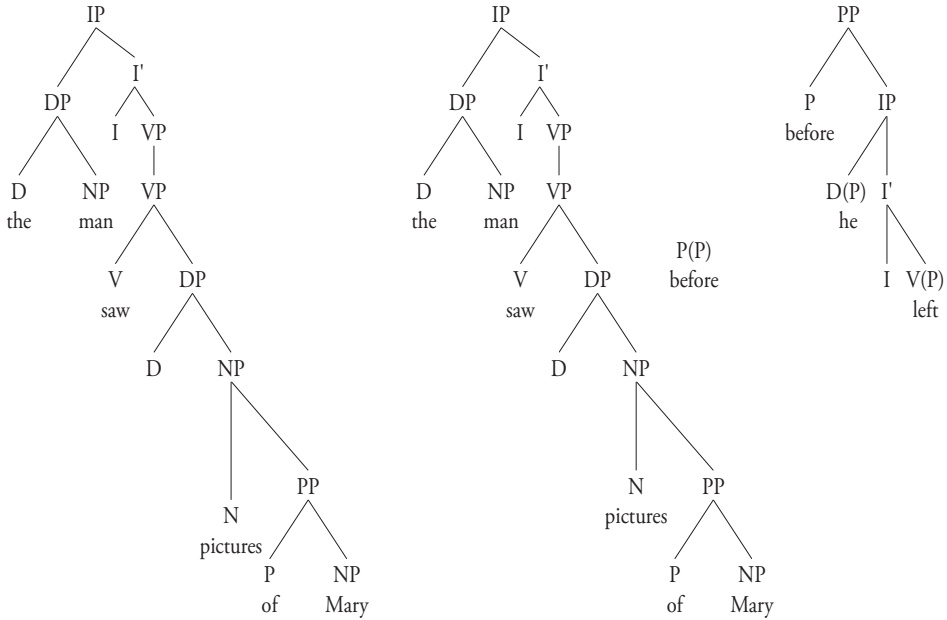
- (i) /Who does /*SUBJECT*/ upset *whol*
- (ii) /a picture of *whol*

We have just seen how the way subjects are introduced in a top-down derivation explains why they are strong islands. Let us examine now how (15) deals with adjuncts. For the sake of the presentation, we will consider only cases relevant to the island paradigm: that is, the discussion will be limited to right adjunctions. Note that, contrary to Kayne’s proposal, right adjunction is a legitimate operation here. (20) is an instance of right adjunction.

- (20) The man saw pictures of Mary before he left.

- | | | |
|--|--|--|
| a) VP has been built. The last terminal node merged in the structure is N(P) <i>Mary</i> | b) P(P) enters the derivation. P(P) cannot replace N(P). | c) In order to prevent of <i>Top-down LCA</i> ’s violation: → PP is built in parallel. |
|--|--|--|

⁹ Crucially note that here, sideward movement (along Nunes’ 1995 lines) is not allowed.



At step b) “before” is the new element to be inserted in the derivation, after [the man saw pictures of Mary] has been built. The terminal node $P(P)_{\text{before}}$ cannot be merged with the last lexical head arrived in the derivation (“Mary”). As a consequence, no ordering can be established between between $P(P)$ and $D(P)$ “Mary” (since $D(P)$ does not c-command $N(P)$). As a result, the adjunct is built in parallel.¹⁰ The fact that the adjunct is built in a derivational workspace distinct from the main derivation explains sentence like (21). (21) is bad because a *wh*-phrase has been extracted from the adjunct [before filing].

- (21) *Which paper_i did you read Don Quixote before filing t_i ?
 (i) */Which paper did you read Don Quixote/
 (ii) /before filing/

[which paper] belongs to the derivational workspace (i).¹¹ Since “read” already has an argument, [which paper] cannot check its theta features in its own derivational workspace. As a consequence, (18) is violated and the derivation crashes. Adjunct’s islandhood is then justified. As for subject cases, (21) becomes good with a parasitic gap construction.

¹⁰ The question of how the adjunct is finally linearized will not be discussed in this paper.

¹¹ Even if the DP, as a spec-CP, has been built in a parallel derivation. As the end, it belongs to the “main” derivation.

- (22) Which paper did you read t_i without filing PG_i ?
 (i) /Which paper did you read *which paper*/

As there is no overt external argument for “read”, “which paper” can check its thematic features: the derivation does not crash. A copy of “which paper” is created in (ii) and inserted to check selectional properties of “filing”.

- (ii) /before filing *which paper*/

Conclusion

I showed that adopting a top-down derivation allows us to reformulate the LCA in a derivational way, and to eliminate the asymmetric c-command relation. The way our Top-down LCA applies provides us with a story to derive straightforwardly subjects and adjuncts’ islandhood: nothing can be extracted from these constituents because they are built in a parallel derivation. The reason why movement across islands becomes acceptable when a “copy” is inserted in the parallel derivation (i.e. parasitic gap) is then intuitively justified.

References

- Boeckx, C., 1999, *Grammatical Notes for a Top-down Syntax*.
- Chomsky, N., 1995a, *Bare phrase structure*. In *Government and binding theory and the Minimalist Program*, ed. G. Webelhuth, 343-439. Oxford: Blackwell. [Reprinted in *Evolution and revolution in linguistic theory: Essays in honor of Carlos Otero*, ed H. Campos & P. Kempchinsky, 51-109. Washington D.C.: Georgetown University Press].
- 1995b, *The Minimalist Program*. MIT Press, Cambridge Massachusetts. Linguistics.
- 2000, «Minimalist Inquiries: The framework». In *Step by step: Essays in minimalist syntax in honor of Howard Lasnik*, ed. R. Martin, D. Michaels & J. Uriagereka, 89-155. Cambridge, Mass.: MIT Press.
- Drury, J., 2005, *Alternative Directions for Minimalist Inquiry: Expanding and Contracting phases of the derivation*. Doctoral Dissertation. University of Maryland.
- Fox, D. & Pesetsky D., 2004, «Cyclic linearization of syntactic structure». To appear in *Theoretical Linguistics, special issue on Object Shift in Scandinavian*; Katalin E Kiss, ed.
- Gautier, V., 2000, *Dérivation linéaire et Cyclicité*. Master thesis. University of Nantes.
- Guimaraes, M., 2004, *Derivation and representation of syntactic amalgams*. Doctoral dissertation. University of Maryland.
- Huang, C.T. J., 1982, *Logical relations in Chinese and the theory of grammar*.
- Kayne, R., 1994, *The Antisymmetry of syntax*. MIT Press. Cambridge Massassuchets.
- Kitahara, H., 1995, «Target α : deducing Strict Cyclicity from Derivational Economy», *LI*.
- Murgia, E., 2004, *Syntactic Identity and Locality Restrictions on Verbal Ellipsis*. Doctoral Dissertation. University of Maryland.
- Nunes, J., 1995, *The copy theory of movement and linearization of chains in the minimalist program*. Doctoral Thesis. University of Maryland.
- Phillips, C., 1996, *Order and Structure*. Doctoral Dissertation. MIT.
- 2003, «Linear Order and constituency», *LI* 34, 37-90.

- Richards, N., 1999, «Dependency Formation and Directionality of Tree Construction». In *MITWPL 34: Papers on Morphology and Syntax*, Cycle two, Vivian Lin, Cornelia Krause, Benjamin Bruening and Karlos Arregi, eds. Cambridge: MIT Working Papers in Linguistics.
- Stepanov, A., 2001, «Late adjunction and minimalist phrase structure», *Syntax* 4, 94-125.
— Doctoral Dissertation. MIT, Cambridge, Massachusetts.
- Uriagereka, J. & Nunes, J., 2001, «Cyclicity and Extraction domains», *Syntax* 3, Number 1. April 2000.

ADJECTIVES AND PROPER NOUNS IN ROMANCE AND ENGLISH

Ion Giurgea
Université Paris VII

0. Abstract

My contribution examines some problems raised by adjectival restrictive modification of proper nouns. I begin with a contrast between the way in which this modification is achieved in English and in Romance, particularly in Romanian: while English uses the same structure for proper and common nouns, in Romance, in the case of the definite article, other structures are preferred (and in Romanian obligatory), which have the form PN+Art+Adj (structures which I call “identificatory appositions”). After briefly examining the types of conversion of PNs to CNs and the behaviour of determiners with PNs, I discuss some possible solutions to the contrast between Romance and English, and conclude that Romance uses the identificatory apposition for a special type of restrictive modification of PNs, one in which a selection among familiar entities is involved. In Romanian instead of preference we find obligatoriness for morphological reasons. Then I sketch an analysis of the structure of identificatory apposition, associating it to a special semantic rule from which we can derive most of its properties. In the end I present some further applications of the structure I have proposed.

1. The problems

In the standard use, proper nouns directly refer to individuals, coming from the lexicon with the semantic type <e>, so they don't admit restrictive modification. However, sometimes a selection has to be made among entities sharing the same name, in which case a restrictive modification is needed. When the restrictive modifier is an adjective, in English it suffices to insert a definite article (or another determiner) before Adj+N. However, Romance languages, in these cases, prefer another strategy, illustrated in (3)-(4). In Romanian, this is more than a preference: a construction of the type (1)b is completely excluded.

(1) a. the tall boy b. the tall John

(2) a. (rom.) băiatul înalt (it.) il ragazzo biondo/alto (fr.) le garçon grand/blond
 boy.the tall the boy blond/tall the boy tall/blond

- | | |
|---|--|
| b. (rom.) *Ionul blond / înalt
Ion.the blond / tall
*cel Ion blond/înalt
the Ion blond /tall | (it.) ?(?) il Gianni biondo / alto
the G. blond tall
(fr.) ?(?) Le Jean blond / grand
the J. blond/tall |
| (3) a. (rom.) Ion blondul
I. blond.the | b. Ion cel blond/înalt
I. the blond/tall |
| (4) a. (it.) Gianni il biondo / alto
G. the blond / tall | b. (fr.) Jean le blond/grand
J. the blond / tall |

The problems I want to address are:

- (i) Where does this difference come from, and
- (ii) How are the structures in (3) to be analyzed?

2. Preliminary observations

Before addressing these problems, some observations are in place about the general issue of restrictive modification of proper nouns, and the behavior of determiners with proper nouns.

2.1. Proper nouns and restrictive modification

In their standard use PNs (proper nouns), functioning as rigid designators, lacking descriptive content, take neither determiners (ex. 5)¹ nor restrictive modifiers (ex. 6):

- (5) a. *(The/a) boy came in b. (*The/a) John came in
- (6) a. * John whom I met in Bilbao is ugly
b. John, whom I met in Bilbao, is ugly
c. (fr.) Le journaliste se promenait dans Sarajevo dévasté
 "The journalist walked through Sarajevo devastated"
d. I don't like John mad

Examples (6)b-d show various modifiers of proper nouns which have in common the property of not being restrictive: an appositive relative clause in (6)b, adjuncts presumably having the structure of small clauses with a PRO subject coindexed with the N in (6)c and d. Notice that the adjective in (6)c is not used to contrast a devastated Sarajevo with other parts of that town or other towns called like that which are not devastated, but simply says that at the time of the main clause event, that city was devastated.

¹ More precisely, they don't take meaningful determiners. They allow at most an expletive determiner, which is a definite or a specialized, "proprietary" article. Some languages make extensive use of such an article (definite in colloquial German and Portuguese, propriety in Catalan). In other languages the definite article is taken by subclasses or individual PNs marked as such in the lexicon. These PNs are never "prototypical" (the most prototypical PNs being antroponyms), and always take the article. Ex. eng. London vs. The Hague (individual items), fr. Paris, Londres etc., vs. la France, le Mexique (a class contrast).

In special cases (always marked), PNs come to behave as common nouns, admitting determiners, restrictive modifiers, plural. Then they are no longer rigid designators, but translate as a predicate. Researchers have identified several types of meaning that can be obtained in these contexts. The list of types that I present here is based on the French school, especially Gary Prieur (1994):

(i) “denominative”: the PNs is converted to the predicate “entity called N” (this is the prototypical case of PN-conversion):

- (7) a. There are three Maries in our class
 b. (fr.) Le Muller que j’ai connu à Bonn
 “The Muller which I met in Bonn”

(ii) spatial/temporal parts (sections, or stages): a class is formed from the spatial or temporal parts of an individual designated by N in its standard use:

- (8) a. (fr.) Le vieux Paris
 “Old Paris”
 b. Young Mozart
 c. (fr.) La Rome antique
 “Ancient Rome”
 d. (sp.) La España medieval
 “Medieval Spain”

(iii) images of the referent: in some cases, the parts of the entity that constitute the class are harder to define. They may be subjective images of that referent, or types of situations linked to that referent—for example, socially defined urban environments or urban life in (9)b, or emotional states of a person in (9)d:

- (9) a. The Paris that I read about
 b. (fr.) Le Paris populaire, le Paris des pauvres
 “The Paris of the common people, of the poor”
 c. Le Pierre que j’aime n’est plus
 “The P. I love doesn’t exist any longer”
 d. Ce jour-là, j’avais vu un Arsène Lupin que j’ignorais, faible, abattu, les yeux las de pleurer...
 “That day I saw an Arsène Lupin which I never knew, weak, depressed, the eyes tired of crying”
 e. (sp.) Esta mañana me encontré con una María muy rejuvenecida
 this morning (me) met. 1sg with a Mary more rejuvenated

(iv) metaphorical: here, the predicate formed may be translated as “person sharing a set of relevant properties with N, being, from a certain point of view, an equivalent of N”; in this case, the modifier doesn’t refer to the bearer of the name in its standard use, but to the referent metaphorically defined as a counterpart of it:

- (10) a. (fr.) Le Gorbatchev albanais
 “The Albanian Gorbatchev”
 b. (fr.) La Christophe Colomb des temps modernes
 “The(fem.) Cristopher Columbus of modern times”
 c. (sp.) el Sinatra español
 “The Spanish Sinatra”

(v) metonymical: on names of famous creators or producers, predicates can be formed meaning “object produced / created by N”:

- (11) a. I bought a Van Gogh and several Picassos
 b. (fr.) J'ai écouté du Bach
 "I heard some Bach"²

2.2. Modified proper nouns and determiners

We have seen that in their different uses as CNs (common nouns), PNs generally receive determiners like ordinary CNs, the difference pointed out in (1)-(4) concerning only the definite determiner and the denominative use of PNs.

However, examples (12)-(13) show another difference between Romance and English: while in Romance even non-restrictive prenominal adjectives, when used with a standard PN, require the definite article, in English non-restrictive prenominal adjectives don't require any determiner, thus opposing to the restrictive ones:

- (12) a. I saw tall Mary Appositive, depictive
 b. I saw the tall Mary Restrictive, denominative
- (13) a. (fr.) *J'ai vu belle Marie b. J'ai vu la belle Marie
 I have seen beautiful M. I have seen the beautiful M.

An explanation for this contrast may be found in Longobardi (1994). This case, as well as other differences between Germanic and Romance, is covered, according to him, by the following parameter: D is strong in Romance and weak in Germanic. It is commonly assumed that D is the locus of referentiality in the noun phrase. The referential element in (12)a and (13) is, obviously, the N. In (12)a D can check referentiality by covert movement, while in Romance, as seen in (13), this movement has to be overt. When no prenominal adjective is present, no determiner is necessary because the PN moves to D:

- (14) J'ai vu (*la) Marie.
 I have seen (the) M.

Of course, this explanation, which uses the theoretical apparatus of the early 90's, can be translated in current minimalist terms by putting an EPP-feature on D in Romance.

Longobardi offers the following arguments for his view:

(i) The absence of the determiner is possible in Romance with modified PNs provided that the PN occupies the first position:

- (15) a. (it.) *antica Roma b. l'antica Roma c. Roma antica
 ancient Rome the ancient Rome Rome ancient

² In (11)b the partitive article, impossible with PNs in the standard use, signals the type shifting (compare to (i)). A similar contrast appears in Romanian—in the metonymical use the N may appear as a mass noun ((ii)a), while in the standard use a PN has to receive, in this context (direct object), a prepositional accusative marker ((ii)b-c):

- (i) (fr.) J'ai écouté Jean
 "I listened to John"
- (ii) a. (rom.) Am ascultat Bach b. *Am ascultat Vasile c. L-am ascultat pe Vasile
 have.I listened Bach have.I listened V. him-have.I listened Ac. V.

For (15)c N-to-D is assumed.

(ii) In the construction PN+Adj, the adjective may have a reading available only prenominally with CNs:

- | | | |
|--------------------------|--------------------------|---|
| (16) a. la sola ragazza | b. la ragazza sola | c. Maria sola = la sola Maria |
| the <i>solo.fem</i> girl | the girl <i>solo.fem</i> | M. <i>solo.fem</i> the <i>solo.fem</i> M. |
| "the only girl" | "the lonely girl" | "only Mary" |

This fact too would be explained by assuming raising of the PN to D in (16)c.

(iii) The same D-strength parameter explains the distribution of argumental bare nouns in Romance and Germanic: thus, in Romance, they appear only in lexically governed position, and have only existential interpretation, while in Germanic they are unrestricted, and they may also acquire a generic reading, functioning as "names of kinds":

- | | | |
|---------------------------------|-----------------------|--------------------|
| (17) a. I bought apples | b. I like apples | c. Apples are good |
| (18) (rom.) a. Am cumpărat mere | b. Îmi plac mere*(le) | c. *Mere sunt bune |
| have.1 bought apples | me.D like apples(the) | apples are good |

The explanation relies on the following assumptions: in argumental positions only DPs are allowed (the D-level must be projected). A null D, as all empty heads, must be lexically governed. That's why the null D specified for mass/(weak) indefinite plural can appear only in object position in Romance. Generics, as a special type of PNs —"names of kinds"—, are based on a chain between N and D, like PNs, the difference being that in their case the base position (N) is interpreted, while for PNs D is interpreted. Thus generics may appear bare in Germanic, where D is weak. In Romance a definite article is necessary since D is strong and the noun, lexically specified as CN, lacks the +ref feature that determines raising to D in the case of PNs. Hence the following contrast:

- | | |
|----------------------------------|-----------------------------------|
| (19) a. (*The) wolfs are mammals | b. (it.) *(I) lupi sono mammiferi |
| | the wolfs are mammals |

This explanation is not without problems. First, the construction PN+Adj, which was taken to show N-to-D, is very restricted in Italian: it appears only with possessives, ordinals and the adjectives *vecchio*, *giovane*, *antico*, *solo*. For other adjectives it is very marginal with a restrictive reading and impossible with an appositive reading:

- | | |
|---------------------------------|-----------------------------------|
| (20) ?? <i>Gianni simpatico</i> | (judgement from Longobardi 1994). |
| G. friendly | |

Secondly, in other Romance languages all postposed modifiers require a determiner:

- | | |
|-------------------------------------|----------------------------------|
| (21) a. (fr.) Je pense à (*la) Rome | b. Je pense à *(la) Rome antique |
| I think of the Rome | I think of the Rome ancient |

It is not clear why movement of the PN to D is blocked in (21)b while it is possible in (21)a.

A possible answer, inspired from the analysis to be presented in 3.1, is that (ordinary) adjectival modification requires an $\langle e, t \rangle$ type, intersective adjectives combining with the nouns (NPs) by the rule of Predicate Modification (Heim and Kratzer 1998, see (22) below), while only $\langle e \rangle$ type nouns raise to D. This would explain why PNs converted to CNs require a determiner, but not why even non-restrictive adjectives with standard PNs do so, as we have seen in (13). The answer lies perhaps in the position of the non-restrictive adjectives with regard to the N: in Romance, they are generally preposed. We could suppose that their position relative to the N must be preserved in order to maintain their special meaning (non-restrictive). So the N must not overpass them. Then the only solution remains the insertion of an expletive article. Formally, this could be represented by assigning prenominal adjectives to special functional projections which the noun should not be able to overpass. It has been in fact proposed (Bernstein 1993, Coene 1999) to treat these adjectives as heads, a proposition which we will not adopt, since such adjectives don't show the usual properties of functional heads (for instance, they are an open class, which is typical for lexical items), and the structures we talk about are obviously nominal extended projections, not adjectival ones.

3. Solutions to the problem in (1)

3.1. A syntactico-semantic solution (Cornilescu 2004a,b)

Turning back now to the problems presented in section 1, I will present the account for the facts illustrated in (2)-(3) (reproduced below for convenience) given by Cornilescu (2004) for Romanian:

- | | | | |
|-----|-------------------------|-------------------------|----------------------|
| (2) | a. (rom.) băiatul înalt | b. *Ionul blond / înalt | Ion.the blond / tall |
| | boy.the tall | *cel Ion blond/înalt | the Ion blond /tall |
| (3) | a. (rom.) Ion blondul | b. Ion cel blond/înalt | |
| | I. blond.the | I. the blond/tall | |

Cornilescu starts by distinguishing two types of adnominal adjectives: NP-adjectives and DP-adjectives. NP-adjectives combine with an $\langle e, t \rangle$ type (which is the denotation of NPs) as follows: intersective adjectives (type $\langle e, t \rangle$) combine by the rule of Predicate Modification (Heim and Kratzer 1998) (or Theta-Identification; Higginbotham 1985):

$$(22) [[\alpha \beta]] = \lambda x [[\alpha]](x) \text{ and } [[\beta]](x)$$

Non-intersective (intensional and relational) adjectives, which are of the type $\langle e, t \rangle$ (predicates of predicates of individuals) combine by the normal Functional Application rule.

Since all NP-adjectives require an $\langle e, t \rangle$ type to combine with, they can't modify PNs. Adjectives in (2) are NP-adjectives, hence the ungrammaticality of (2)b.

DP-adjectives combine with an $\langle e \rangle$ type, in a predicative construction (small clause). PNs are of type $\langle e \rangle$, so they allow DP-adjectives. The structures in (3) reflect DP-level adjectives, in a predicative structure:

(3^b)_b [_{DP} DP [_{D'} +def [_{CP} τ_{DP} [_{C'} [_{PredP} τ_{DP} [_{Pred'} *cel* [AP]]]]]]]]]

Cornilescu bases her analysis of (3)_b on the following arguments: first, relational adjectives, which typically apply to the intension of the N-function and not to the entity denoted by the DP, are impossible in the structure (3)_b:

(23) a. societatea (cea) bogată b. societatea (*cea) astronomică
 society.the (the) rich society.the the astronomical

Secondly, the structure (3)_b requires a definite determiner:

(24) a. muntele cel înalt b. * un/fiecare/alt munte cel înalt
 mountain.the the high a/each/another mountain the high

This second constraint is explained by analyzing *cel* as a predicative head with a deictic feature (hence its interpretation, see section 4, (39)-(40)), requiring a definite subject.

As for structure (3)_a, which is not available for CNs, she assumes that the adjective checks definiteness and the PN moves to SpecDP_{max}, as ordinary PNs do (see Longobardi 1994, presented in 2.2. above).

This analysis has several problems. First, why is the structure Det+Adj+NP possible?

(25) frumoasa Maria
 beautiful.the Maria

To see if we can answer this, we have to give a closer look to prenominal adjectives. As other Romance languages, Romanian has three kinds of prenominal adjectives (putting aside determiner-like adjectives like *prim* “first”, *ambii* “both”, which we won’t discuss here, since they don’t involve restrictive modification):

(i) most adjectives appearing in this position are non-restrictive (when they are restrictive, they appear after the noun); they usually express inherent, familiar properties, and often have an affective, emotional connotation (ex. (25));³

³ Various tests indicate that ordinary quality adjectives are non-restrictive in Romance when anteposed:

— impossibility of contrastive focus:

(i) (it.) * Il BIONDO ragazzo è venuto, non il bruno
 the blond boy is come not the dark-haired
 (rom.) *BLONDUL băiat a venit, nu cel brunet
 blond.the boy has come not the dark-haired

— impossibility of generic use:

(ii) (rom.)a. * Îmi plac înaltele case b. Îmi plac casele înalte
 me.D like high.the houses me.D like houses.the high

— impossibility to appear in quantificational DPs:

(iii) (rom.) Fiecare (*frumoasă) zi (frumoasă) mă bucură
 each beautiful day beautiful me delights “Each beautiful day delights me”

— impossibility to be used as an answer to a which- question:

(iv) (it.) — Quale ragazza è venuta? # — La bionda ragazza — La ragazza bionda
 which girl is come the blond girl the girl blond

See Zamparelli (1993), a.o.

(ii) a small number of ordinary quality adjectives may be restrictive (function as selectors):

(26) a. *tânărul Petre*
young.the Petre

(iii) some are non-intersective modalizers (or intensional adjectives) (type “former, alleged, possible, mere, true”), which are always restrictive and appear only in this position:

(27) a. *pretinsul Grigore*
alleged.the Grigore

For type (i) we may say, given the meaning, that they are DP-level adjectives, in the left-periphery of DP, either topical or focal.

In (ii) we may have to deal with an appositive structure, since *tânărul* “the young” may also function as a noun (these are adjectives that allow (lexical) nominalization):

(28) *Tânărul a intrat în cameră*
young.the has entered in room

But what about (iii)? The adjective is surely NP-level (being of the type $\langle e, t \rangle \langle e, t \rangle$). We could say that the PN is converted to a CN. But why can't this conversion apply in the case of postnominal adjectives, as we have seen in (2)-(3)?

Here we should observe that this conversion is available in Romanian, as for other Romance languages, for most types of determiners and meanings of the converted N, as shown throughout 2.1. The only problematic case, presented in (2)-(3), involves a different type of restrictive modification, one in which an acquaintance with the members of the class PN is supposed, which is not the case in (27), for example, where the N is purely “denominative”. So I would propose the following generalization

(29) Romanian applies conversion $PN \rightarrow CN$ when PN must denote a class except for the case of unique selection from familiar individuals; in this latter case, a special structure (which I will call “identificatory apposition” - see ex. (3)) is used

For a usual conversion of the denominative type, see also:

(30) *Sunt un Ion si trei Marii la cursul nostru*
are a Ion and three Maries at course.the our

A further problem for Cornilescu's analysis is that PNs which are specified in the lexicon as requiring a definite article do admit the structure $D+PN+Adj$:

(31) *Parisul vechi*
Paris.the old

Also the analysis of examples (3) as instances of raising reduced relatives faces a series of problems, which I will present in section 4, when I discuss the analysis of these structures, proposing an alternative view.

3.2. A morphological solution

Since the impossibility of having a determiner concerns only the definite article, and does not arise when this article comes from the lexicon, as shown in (31), we could simply assume that the reason for the impossibility of (2)b is the fact that some PNs don't allow an enclitic definite article. This is plausible because the Romanian enclitic definite article has affixal status —forming a true definite declension:

- (32) a. lupul (m.sg.N-A.) b. câinele (m.sg. N-A.)
 wolf.the dog.the

In (32) we see how the forms of the article for the same gender, number and case may vary according to the declensional type of the N, which is a indisputable sign of the inflectional status of the article.

So the ability of having or not having definite forms is a morphological matter.

However, in other cases where an enclitic article is possible, Romanian uses a proclitical form, which is clitic but not affixal (see (33)). Then why isn't this form used in the case under discussion (34)?

- (33) a. *doiii copii c. *maii mari copii
 two.the children more.the big children
 b. cei doi copii d. cei mai mari copii
 the two children the more big children "the biggest (eldest) children"
- (34) a. *Ionul blond b. *cel Ion blond
 Ion.the blond the Ion blond

3.3. Conclusion: a compromise

(33)-(34) show that a purely morphological explanation is difficult to maintain. So I will adopt (29), which also covers other Romance languages (see ex. (4)), with the amendment that the preference in (29) is stricter for Romanian due to the affixal status of the ordinary definite article in this language.

In conclusion, when speaking of restrictive modification of PNs and type-shifting from PN to CN, we have to distinguish two types of selection from a class denoted by the N:

- (35) (i) PN = {(the.PN)₁, (the.PN)₂, ...} (selection among familiar individuals)
 (ii) PN = {x: entity called "PN"} (denominative), or
 = {x: (spatial/temporal) part of PN}, or
 = {x: image of PN}, or
 = {x: entity resembling to PN} (metaphorical), or
 = {x: entity produced by PN} (metonymical) etc.

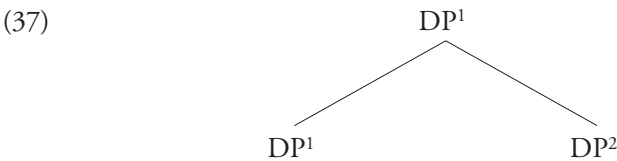
The representation in (35)(i) gives us a clue for the analysis of the structures in (3), to which we may proceed now.

4. The structure used by Romanian to avoid (1): the "identificatory" apposition (IA)

The structures illustrated in (3) are an instance of a wider type, which can be encountered in many languages, including Germanic:

- (36) a. Richard the Lion-Hearted b. (fr.) Philippe le Bel
 P. the handsome

Based on the results reached in 3.3 and formalized in (35), I propose the analysis in (37) for this structure, with the associated semantic rule (38):



- (38) $[[DP^1 DP^2]] = \text{the unique } x. x=[[DP^1]] \text{ and } x=[[DP^2]]$

The rule (38) explains why DP¹ may be only a PN or a definite description (see (24)), and DP² is always a definite description (see (39)), and why this structure is used for selection among *familiar* referents (hence it is not felicitous for generics, see (40)):

- (39) (rom.) a. Mihai cel blond c. Mihai profesorul
 Mihai the blond Mihai professor.the
 b. *Mihai un blond d. * Mihai un profesor
 Mihai a blond Mihai a professor

- (40) (rom.) a. Apa caldă nu e bună de băut
 water.the warm not is good to drink "Warm water is not good to drink"
 b. Apa cea caldă nu e bună de băut
 water.the the warm not is good to drink #generic, OK selection from
 contextually salient bottles of water, for ex.
 c. Îmi plac casele (??cele) înalte
 me.D like houses.the the the high

Notice that (37) is not a symmetrical structure: since according to standard minimalist view the object formed by Merge has to be the projection of one of the merged elements, we have an asymmetry between DP¹ and DP², in that DP¹ is the head, while DP² is a non-head (an adjunct, unless further refinement of the structure is introduced, a problem which I will not further develop here).

The order between DP¹ and DP² is pragmatically established: DP² identifies a referent from all the possible DP¹. Thus, (41)a below is appropriate as an answer to the question "Which Dumitru?", while (41)b is appropriate as an answer to the question "Which professor?":

- (41) (rom.) a. Dumitru profesorul b. Profesorul Dumitru
 Dumitru teacher.the teacher.the D.

I will now present some reasons to prefer my analysis to that proposed by Cornilescu, and presented in section 3.1.

First, Cornilescu’s analysis of the structure in (3)a (PN Adj+Art) cannot explain why the article is obligatory in this case. She considers that the PN and the Adj start

from a SC, and then the adjective inflected for definiteness checks definiteness on D while the PN, being referential, moves to SpecDP:

- (42) $[_{DP} \text{PN} [_{D} \text{Adj-Art}] [_{CP} \tau_{PN} [_{C+\text{rel}}] [_{\text{PredP}} \tau_{PN} \text{Pred}^0 \tau_{\text{Adj-Art}}]]]]$
 $[_{DP} \text{Ion} [_{D} \text{blond-ul}] [_{CP} \tau_{\text{Ion}} [_{C+\text{rel}}] [_{\text{PredP}} \tau_{\text{Ion}} \text{Pred}^0 \tau_{\text{blond-ul}}]]]]$

But if the article is necessary to check some feature on D in this case, despite the raising of the PN, why isn't it necessary with unmodified PNs (where, let us recall, it is forbidden)?

- (43) a. Mihai Viteazul b. Mihai(*ul)
 Mihai brave.the Mihai the

Secondly, according to her analysis of (3)b (see (3')b above), *cel* would have to be split into two words: a Pred^0 when preceded by a N, a D^0 elsewhere (that is in cases of N-ellipsis, cardinals, superlatives, see (33)).

Thirdly, structure (3)a is in fact possible only with adjectives that can be nominalized (ex. (44)-(45)):

- (44) a. Maria blonda b. M-am întâlnit cu blonda
 Maria blond.the me-have.1 met with blond.the(fem) "I met the blonde"
- (45) a. *Maria roșia c. *M-am întâlnit cu roșia
 Maria red.the me-have.1 met with red.the(fem)
 b. Maria cea roșie
 Maria the red

This follows from our analysis, in which the definite adjective is part of a second DP. In Romanian only nominalized adjectives can take the enclitic definite article. In the elliptical constructions ($[_{N}e]$ +Adj) only the proclitic definite article *cel* may be used. So the second DP of the IA may either appear as *cel*+Adj, if the adjective is not nominalized (case (3)b), or as Adj+Art, with nominalized adjectives (case (3)a). In Cornilescu's analysis, where the adjective is the predicate of a SC, the restriction of the structure (3)a to nominalized adjectives remains a mystery.

The existence of two structures for this sort of PN modification, (3)a and (3)b, is thus better explained by our analysis than by Cornilescu's. Moreover, our analysis unifies structure (3)a to another structure, not discussed by Cornilescu, PN CN-Art:

- (46) a. Maria profesoara
 Maria teacher.the

We see that in the structure PN X-Art, X must always be nominal, a N or a nominalizable adjective, which directly follows from our analysis of X-Art as a separate DP.

5. Problems for our analysis

Our analysis doesn't explain why in Romanian CNs too are allowed in the structure (3)b, while this is impossible in other Romance languages, and also in Romanian for the structure (3)a:

- (46) (rom.) a. fata cea blondă c. (fr.) *la fille la blonde
 girl.the the blond the girl the blond
 b. *fata blonda d. (it.) *la ragazza la bionda
 girl.the blond the girl the blond

An economy principle could easily rule out (46)b-d. The difficulty remains the structure (46)a. Cornilescu explains this structure by the selectional properties of the predicative head *cel*: it takes as a subject any definite DP (hence also CNs). In (3)a only PNs are allowed because only they can raise to SpecDP_{max}, above the determiner -L.

An answer in the lines of our analysis (37)-(38) would be that Romanian has further developed the marking of selection among familiar entities, as opposed to ordinary restrictive modification.

The agrammaticality of (46)b can further be related to the impossibility of having CNs in the second member of IAs after a CN (see (47)). If the adjective in (46)b is nominalized, as we proposed, the two cases can be covered by the same rule.

- (47) a. Maria profesoara b. * Femeia profesoara
 Maria teacher.the woman.the teacher.the

As to where this rule comes from, I propose that the answer is that in an IA-structure involving CNs the N must be interpreted as the same in the two DPs:

- (48) a. the x {x=the {y:N(y)} and x= the {y: N(y) and A(y)}}
 b. * the x {x=the {y:N(y)} and x=the {y: M(y) and A(y)}}, with M≠N

This is a natural restriction, which simply says that *a thing cannot be identified as two sorts of things*. It doesn't operate on PNs because there only one sort of things is involved (only one of the DPs contains a predicate).

The structure (3)b, as all structures of type *cel*+Adj, relies on an ellipsis:

- (49) [_{DP} cel [_{NP} [e] AP]

In IA, [e] in the second DP is identified with the N in the first DP.

Thus the structure (3)b, with N-ellipsis in the second member, is the only way for an IA involving CNs to satisfy the condition stated in (48).

A further structure apparently involving IA is the structure Def+CN+PN:

- (51) a. (rom.) profesorul Popescu b. (fr.) le professeur Ducange
 professor.the Popescu the professor Ducange

In this structure the determiner has to be definite, which seems to indicate an IA:

- (52) a. (rom) *un profesor Popescu b. (fr.) *un professeur Ducange
 a professor Popescu a professor Ducange

But, unlike in the IA cases treated above, DP₁ is not familiar. Moreover, D may be absent in English, which recalls the behavior of modified PNs:

- (53) professor Smith

Stowell (1991) and Giusti (2002) have proposed that the CN is an (adjectival) modifier in such cases. But that doesn't explain why it appears only with PNs. I leave this problem for further study.

The strong limitations on IA (for instance, only PNs and personal pronouns in Romance languages other than Romanian) could be seen as a general problem for my analysis. The answer I can give is that wherever ordinary restrictive modification is available (by Predicate Modification or Functional Application), it is preferred.

6. Further applications of the structure proposed for IA

The same IA could be represented in the structure *personal pronoun+definite DP*, which appears in languages where the personal pronoun can't be a transitive D (a-b below vs. c-d):

- | | |
|-------------------------------|------------------------|
| (54) a. (rom.) noi lingviștii | c. we linguists |
| we linguists.the | d. (it.) noi linguisti |
| b. (fr.) nous les linguists | we linguists |
| we the linguists | |

The structure of IA apposition could explain the arousal of structures with multiple determiners, which are found in many languages:

- | | |
|-------------------------------------|-----------------------------|
| (55) (old rom.) locul cela strimtul | (got.) jains wigs sa raihta |
| place.the that-one narrow.the | that way the right |
| (alb.) burri i madh ⁴ | (norv.) skogen den grønne |
| man.the Agr.msgN big | forest.the the green |
| (gr.) o anthropos o kalos | |
| the man the good | |

The structure in (37)-(38) explains why in multiple determiners constructions the determiner is always definite. No other theory of multiple determiners, as far as I know, does this (usually, two D-layers are posited, but why should those layers always be definite?).

A further interesting related problem is the arousal of Germanic adjectival weak declension, which appears when the adjective is preceded by definite determiners. In the first attested stages of Germanic languages, the adjective took weak declension only when combined with the definite article (*sa, so, þata* everywhere except ON, which has *inn*), and the definite article was only anaphorical. Historical grammarians trace back the weak declension to nominalizing suffix. The Germanic weak declension is formed by adding the suffix *-n-* to the adjectival stem, and a similar *-n-* suffix (in the form *-o:n-*) is used, in Latin and Greek, to build nouns meaning "entity (usually person) having the property P".

⁴ In Albanian the so-called "adjectival article" has lost all its determiner properties, functioning as a mere agreement marker (it appears, with most adjectives, in all positions —predicative, in indefinite or quantified DPs, when the adjective is nominalized, etc.). However, as the example shows, it reproduces the forms of the definite article, so the two probably have the same source.

Putting all this together, we could explain the origin of the weak declension as follows: when having to select among familiar entities, Germanic used a form of IA, nominalizing the adjective, and also using a deictic determiner. As the deictic determiner evolved in the direction of the definite article, the *-n-* suffix on the adjective, continuing to be associated with the determiner, was reinterpreted as a simple definiteness agreement marker, losing its nominalizing capacity:

- (56) a. sa goda b. ains gods wairs
 the good-N(ominalizer) one good man
 c. wairs sa goda selection from familiar entities, hence IA allowed
 man the good-N
 d. *sa gods wairs →(by ways of analogy) sa goda wairs

Instead of (56)c we find in the attested Germanic languages (56)d as a result of the generalization of the order Adj-N in Germanic, which was used with all the other determiners.

References

- Bernstein, J., 1993, *Topics in the Syntax of Nominal Structure Across Romance*, CUNY Diss.
 Coene, M., 1999, *Definite Null Nominals in Romanian and Spanish*, diss., Universiteit Antwerpen.
 Cornilescu, A., 2004a, «Modes of semantic combination: modification, predication and the syntax of the Romanian adjective, paper presented» at the *Annual Conference of the English Department of the University of Bucharest, june 3-5 2004*.
 —, 2004b, «Modes of semantic combination: NP/DP adjectives and the structure of the Romanian DP», conference held at LLF/Paris VII, December 2004.
 Gary P., M.-N., 1991a, «Le nom propre constitue-t-il une catégorie linguistique?», *LFr* 92, 4.26.
 — 1991b, «La modélisation du nom propre», *LFr* 92, 4.26.
 — 1994, *Grammaire du nom propre*. Presses Universitaires de France, Paris.
 Giusti, G., 2002, «The functional structure of noun phrases: a bare phrase structure approach». In Cinque, G. (ed.) *Functional Structure in DP and IP*, Oxford U. P., 54-90.
Gramática descriptiva de la lengua española, 1999, Bosque, I. and Demonte, V. (eds.), Espasa Calpe, Madrid. Vol. I chap. II *El nombre propio* by J. Fernández.
 Heim, I. and Kratzer, A., 1998, *Semantics in Generative Grammar*. Oxford, Blackwell.
 Higginbotham, J., 1985, «On semantics», *LI* 16: 547-593.
 Jonasson, K., 1991, «Les noms propres métaphorique: construction et interprétation», *LFr* 92, 64-81.
 —, 1994, *Le nom propre. Constructions et interprétations*, Louvain, Duculot.
 Krahe, H. and Meid, W., 1969, *Germanische Sprachwissenschaft*, Sammlung Goeschke, Berlin.
 Longobardi, G., 1994, «Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form», *LI* 25:4, 609-665.
 — 2001, «The structure of DPs.» In Baltin, M. and C., C. (eds.) *Handbook of Contemporary Syntactic Theory*.
 Marit, J., 2003, «Double definiteness in Scandinavian», *Nordlyd* 31.1:230-244
 Noailly, M., 1991, «L'énigmatique Tombouctou: nom propre et position de l'épithète», *LFr* 92, 104-113.
 Stowell, T., 1991, «Determiners in NP and DP» in Leffel, K. and Bouchard, D. (eds.), *Views on Phrase Structure*, Dordrecht, Kluwer, 37-56.
 Zamparelli, R., 1993, «Prenominal Modifiers, Degree Phrases and the Structure of the AP», *University of Venice Working Papers in Linguistics* 3, 138-161.

EXTRACTING INFORMATION FROM PARTICIPIAL STRUCTURES

Enikő Héja, Kata Gábor

Linguistic Institute, Hungarian Academy of Science and Eötvös Loránd University

Abstract

Our applied linguistic research aims at increasing the efficiency of a rule-based information extraction (IE) system by enhancing it with further grammatical knowledge. The input of the IE system is made up of sentences of business news. The event of the piece of news is identified through the main verb of the sentence, while participants and circumstances of the event through arguments and adjuncts of the main verb. Our objective was to unfold the hidden information, contained by NPs within which non-finite verbs (e.g. participles) appear. Thus, we invented a rule-system to transform participial structures into sentences with a finite verb, so that they could serve as input of the IE system. To tackle this task we had to be able to distinguish between real participles and adjectives. According to us there are some distributional criteria which can be used as the basis for creating the right classification.

1. Introduction

In what follows we would like to present our applied linguistics research which aims at increasing the efficiency of an already existing rule-based information extraction system by enhancing it with further grammatical knowledge. Our work concentrates on the NewsPro information extraction system (Prószéky 2003), developed jointly by MorphoLogic Ltd., Institute of Informatics at Szeged University, and Linguistics Institute of Hungarian Academy of Sciences. The system was developed and tested on a corpus of short business news.

Firstly, NewsPro performs a shallow syntactic analysis on the input text, then it matches pre-defined semantic patterns —so-called ‘event frames’— to the text. In case of successful pattern matching, slots of event frames are filled by the elements of the text, thus the output identifies the main event of the piece of news as well as its participants and circumstances. Semantic patterns are centered around finite verbs while their complements and adjuncts represent participants and circumstances, respectively. Thus, pattern matching is based on the *finite* verb previously recognized as predicate, and its argument structure. This method relies on the supposition that in short news it is always the verbal predicate that expresses the main event. Although this approach proves to be working in most cases, it has the disadvantage of

omitting secondary information (frequently indicated as the cause or the antecedent of the main event) from pattern matching. The reason is that secondary information is represented grammatically by non-finite verbal forms such as participles or deverbal nouns.

For example:

‘*[_{part} A cég által kedden meghozott döntés] nyomán sokan keresnek új munkahelyet.*’
 ‘Due to [the decision made by the firm on Tuesday], many people are looking for a new job.’

In the sentence above, NewsPro is able to identify the main event (i.e. looking for new jobs), but not the bracketed constituent, which expresses an earlier event, conceived as the cause of the main event. However, the user may be interested to learn about the antecedents and the connection between the two pieces of information.

This phenomenon is supposed to be handled by a preprocessing module within NewsPro. Preparation of transformational rules as well as other tasks related to the preprocessing of the text were performed by Intex (Silberztein 1993). The module transforms input participial structures into complete sentences with a finite verb as their predicate. Further steps of the processing, such as syntactic parsing and semantic pattern matching may run on the transformed sentences without any modification.¹ Moreover, as Hungarian constituent order is relatively free, we expect the system to yield better results on automatically generated sentences, as their constituent order is homogeneously SVO.

2. The Corpus Annotation Tool

The implementation and testing of the transformation rules, as well as any task the preprocessing of the text involved were carried out using Intex, a powerful corpus processing tool freely available for research purposes. Intex is particularly suitable for implementing lexicalist approaches to language processing, as it makes wide use of several types of structured dictionaries. The feature we took particular advantage of is that morphosyntactic and semantic description of words are available at every level of the analysis. This allowed us to create transformation rules which referred to the base verb of participles (that we also coded in the dictionary), and the semantic-syntactic properties of the base verb.

3. The Outlines of the Problem

The success of the preprocessing module on one hand depends on the *grammatical and semantic well-formedness* of the output (theoretical requirement) and on the other hand on *the degree of informativity* of the transformed sentences (practical requirement). We made an attempt to elaborate an algorithm for filtering out supposedly informative participial structures on the basis of solely grammatical information. Below we give a brief description of the method we used.

¹ We have not dealt with the morphological aspect of the elements in the resulting sentences yet, the morphological module of Hungarian Intex is presently under development.

3.1. Participles and Participial Structures in Hungarian

The focus of our research consists in past participles. Hence, we need to introduce the main features of Hungarian past participles. Although linguists did not reach a consensus about the exact status of participles in Hungarian, it is widely accepted that participles originate from verbs, either by derivation or by inflection. From our point of view the precise nature of this process plays no role, in tandem with the consideration that there might be no strict boundary between derivational and inflectional suffixes. Past participles can be derived freely from verbs if the verb has an argument the thematic role of which is patient or theme.

Below we present the form of Hungarian past participles:

Verb - (Vowel) - (t) - t

The form of Hungarian past participles coincides with the past tense form of the corresponding verb. Some examples: *ad - ott* 'given'; *megérdemel - t* 'deserved'.

The expression of the form *verb - suffix* is the head of the participial structure. The characteristics of participial structures which enable the production of well-formed sentences by means of our rules are on one hand that the participle preserves the meaning of its base verb, and on the other hand that the arguments of the base verb can be derived from the internal structure of the NP containing the given participial structure. As the internal structure of a Hungarian NP is rather strict, our rules are able to recognize the constituents of it and identify them as adjuncts and complements of the base verb.

3.2. The Problem

First, let us have a look at some examples, which might be the output of our transformation rules, but are not able to serve as input of the pattern matching process:

'a <i>jegyzett</i> tőke'	[particip Valaki jegyzett tőke -t]
The subscribed capital	Somebody subscribed capital - ACC
'a <i>nyomott</i> hangulat'	[particip Valaki nyomott hangulatot -t]
The depressed mood	Somebody depressed mood - ACC
'a <i>mérsékelt</i> PC-chip kereslet'	[particip Valaki mérsékelt PC-chip kereslet-t].
The moderated PC-chip demand	Somebody moderated PC-chip demand - ACC
a <i>nyomtatott</i> sajtóban	[particip Valaki nyomtatott sajtóban -t]
The printed media - INE	Somebody printed media - ACC
'a <i>ragozott</i> szóalakok - ból'	[particip Valaki ragozott szóalakokból -t]
The inflected word forms - ELA	Somebody inflected word forms
'a <i>kerekített</i> euróár - ak'	[particip Valaki kerekített euróárak -t]
The rounded Euro price - PL	Somebody rounded Euro prices - ACC
'a <i>használt</i> ingatlan - ok'	[particip Valaki használt ingatlanok -t]
The used property - PL	Somebody used properties - ACC

The sentences above are in some way semantically ill-formed. For instance in the second example *nyomott hangulat* 'depressed mood' has nothing to do with the fact that somebody depresses something. There are also cases which are not as bad as this.

Taking *ragozott szóalakok* ‘inflected word forms’, we find that the resulting sentence is improper in another way. Namely, this NP denotes rather a state than the process itself, so *ragozott szóalakok* does not mean that there is somebody who really did inflect those word forms, instead it means that those word forms are in a certain state, i. e. they have certain suffixes.

So, we have to be able to tell apart participial structures that result in well-formed sentences from those that do not. Our solution is based on the fact that there is an adjective-participle homonymy in Hungarian. Actually, we state that those transformations that contain *adjectives* produce ill-formed sentences, and the structures containing *participles* can serve as the input of our rules.

3.3. Adjectives and Participles

As we mentioned above there is no consensus among Hungarian linguists about the exact nature of participles. One fact which is responsible for this insecurity is homonymy between adjectives and participles. As some authors state, there are certain cases when the nature of a given expression is absolutely undecidable (e.g. *kedvelt*, ‘much liked’) (Kömlősy 1992 and Kenesei 2000). Thus, the question arises: on what ground is it possible to distinguish between adjectives and participles? We follow Kömlősy’s (Kömlősy 1992) suggestion according to which there are syntactic —mainly distributional— tests, which we can rely on to make our decision.

Before we list them, we have to note that semantic facts also support our hypothesis above. Namely, as participles keep the main characteristics of a verb—for example its argument structure and the ability to place the time of the event denoted by the participle in relation to the time of the main verb of the sentence (Kiefer 2000)— they are capable of preserving the event structure of the original verb, as opposed to adjectives. According to us this is the most important feature of participles which guarantees the well-formedness of the resulting sentence.

Now, coming back to our main train of thoughts we list the syntactic tests, mentioned above:

- (1) *comparison*: only adjectives can undergo comparison,
- (2) *deriving adverbs*: only adjectives can serve as input of adverbial derivation,
- (3) *predicative use*: only adjectives have predicative use,
- (4) *preverb detachment*: only preverbs in participles could be detached when there is a negation.

Unfortunately, the criteria above are not able to help us directly in telling apart adjectives from participles, since we cannot take the possible transformation of the texts’ elements into consideration when applying our rules.

3.4. Our Solution

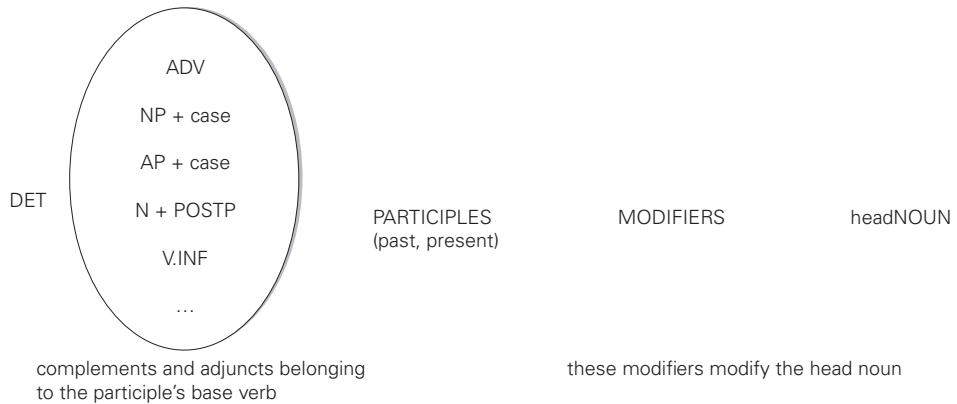
Thus, the question remains: how could we distinguish between adjectives (the transformations of which supposedly result in ill-formed sentences) and participles (with which the rules output well-formed sentences) *in situ*?

We found that the three criteria below are sufficient to make the right classification:

- (1) If at least one of the base verb's *complements* is present, then it is a participle (Komlósy 1992).
- (2) If at least one of the base verb's *adjuncts* is present, then it is a participle.
- (3) If at least a *preverb* is present, then it is a participle.

The acceptability of the (1) criterion follows from the fact that participles are “closer” to the verb from a derivational aspect and they keep the arguments of the base verb (or at least some of them.)

To support the (2) criterion, we have to elaborate the statement above according to which the internal structure of the Hungarian NP is quite bound. The following illustration represents the constituent order within the NP.



This means that the constituents within the ellipse are always attached to the participle, because they were complements and adjuncts of the base verb, as the following example illustrates:

$[_{part} A \text{ cég eladás} - a \text{ kapcsán} \text{ felmerül} - t] \text{ hamis} [_{N} \text{ híresztelés}]$
 DET firm sale - POSS in-connection-with emerge - PAST false rumor
 'A false rumor that emerged in connection with the selling of the firm'

In this NP *A cég eladása kapcsán* is an adjunct of the base verb, while *hamis* is a modifier of the head noun *híresztelés*.

The (3) condition correlates with the observation that while participial phrases put emphasis on the course of events denoted by the base verb, adjectives express states of processes. As Hungarian preverbs' main function is to express aspect, it is conceivable that the presence of a preverb supports the 'course of events' reading, which means that there is a participle. Beside these intuitions, the above mentioned distributional facts also supply confirmation for our hypothesis.

(1) comparison:

**a [tegnap mérsékel]-t-ebb kereslet*
 the [yesterday moderate]- SUFF- COMP demand

(2) predicative:

**Ez a szóalak [az órán ragoz]-ott.*'

this word form [during the class inflect]- *SUFF*

(3) ADV formation:

**[1000 Ft-ra mérsékel]-t-en'*

[1000 Ft-onto reduce]- *SUFF-ADV*

(4) preverb detachment:

**A házak [fel nem újít]-ott-ak.'*

the houses [re- not store]- *SUFF-PL*

Consequently, our main hypothesis is proved: only those expressions which do have complements, adjuncts or preverbal modifiers are participles, and those which have neither of them in their left context are adjectives. Since our original aim was to filter out structures bearing information, we have to examine how syntactically separated groups of participles and adjectives can be compared with the informative-uninformative partition. In fact, it turns out that informative structures coincide with participial structures.

4. The Grammar

When setting up our transformational rules, we used the following presuppositions:

- (1) it is possible to derive participles both from transitive and intransitive verbs,
- (2) if the verb is intransitive the head of the noun phrase is the subject of the participle's base verb,
- (3) if the verb is transitive the head of the noun phrase is the object of the participle's base verb,
- (4) the complements and adjuncts of the base verb appear before the participle,
- (5) as the past participle usually expresses anteriority in time, the transformed finite verb is in past tense.

In addition, we dealt only with NPs which begin with a determiner. We decided to do so because complements and adjuncts appearing before participles might be extremely diverse, which makes the exact recognition of the NP's left boundary hard. Using a determiner as the left boundary of the NP enables us to identify all constituents between the participle and the beginning of the NP as complements and adjuncts of the base verb, while expressions appearing between the participle and the head noun can be identified as the modifiers of the noun (in accordance with the illustration above).

Firstly, on the basis of the (1)-(3) conditions we divided the NPs into two groups: one of them consists of participles derived from transitive verbs and the other from intransitive verbs. The practical purpose of that was to encode the feature of transitivity in a dictionary, by means of which local grammars are able to perform the transformation.²

² We used the verbal argument structure database prepared by Corpus Linguistic Department to develop our dictionary (i. e. to encode the relevant syntactic characteristics of verbs).

4.1. Transitive Verbs

We considered a verb transitive if it had at least one transitive occurrence in our database. To transform this kind of participial structures into finite sentences we used the following algorithm:

$$\text{Det (V_compl/V_adj) V}_{\text{past_part}} \text{ N} \rightarrow \text{Valaki V}_{\text{past_part}} \text{ Det N - ACC (V_compl/V_adj)}$$

The rule above says that if an NP consists of a sequence of a *determiner* ‘Det’, *complements* and/or *adjuncts* of the base verb (‘V_compl/V_adj’), a *past participle* (‘V_{past_part}’) and a *head noun* (‘N’), it has to be transformed into a string which is made up of *Valaki* (the Hungarian counterpart of ‘somebody’), the *past participle*, a *determiner*, the *head noun* with accusative case and finally, the *complements* and/or *adjuncts* of the base verb. Here parentheses mark optionality. The reason for the acceptability of the past participle’s use at the right hand side of the rule is the fact that in Hungarian the forms of past participles and the forms of the corresponding past tense verbs coincide. The transformation is illustrated by the example below:

[_{part} A bővítés után - ra tervez - ett] munkaerővándorlási [_N korlátozás]
 DET expansion after - SUB plan - SUFF migration-of-labor restriction
 ‘The restriction of migration of labor planned to be introduced after the expansion’

‘Valaki tervezett munkaerővándorlási korlátozást
 ‘Somebody planned to introduce the restriction of migration of labor
 bővítés utánra.’
 after the expansion.’

In other words, the first step of filling the slots in the argument structure of the base verb is to identify the head of the NP as the object of the resulting sentence. Secondly, the subject position —since in most cases the subject itself does not appear in this kind of participial structures— is occupied by the expression *Valaki* ‘Somebody’. This solution is made possible by the fact that in these cases the subject is usually an agent. Actually, though not frequently, the subject might also appear in the participial structure. When this happens, it could be expressed by the postposition *által* ‘by’. Such structures are handled by the following rule:

$$\text{Det N}_{\text{subj}} \text{ által (V_compl/V_adj) V}_{\text{past_part}} \text{ N} \rightarrow \text{N}_{\text{subj}} \text{ V}_{\text{past_part}} \text{ N - ACC (V_compl/V_adj)}$$

As in the example below:

[_{part} A budapest - i cég által rendszeresen közzéte (sz) - tt] eredmény - ek
 DET Budapest - ADJ firm by regularly publish - SUFF results - PL
 ‘Results published by the firm in Budapest regularly’
 ‘A budapesti cég közzétett eredményeket rendszeresen.’
 ‘The firm in Budapest published results regularly.’

There are cases, when the subject of the base verb appears in the participial structure, but it is not expressed by an *által* postpositional phrase. In such cases the NP denoting the subject of the base verb is in nominative case morphologically, and it is the possessor of the head noun of the main NP. However, this construction does not

contradict our hypothesis, according to which the constituents preceding the participle are complements and adjuncts of the base verb, since the possessor plays the role of the subject in the transformed sentence:

‘_{[part} A svéd Networks tervez - ett] adósságátalakítás - i program - já - ban’
 DET Swedish Networks plan - SUFF debt-conversion - ADJ program - POS - INE
 ‘In the debt conversion program planned by the Swedish Networks’
 ‘A svéd Networks tervezett az adósságátalakítási programot.’³
 ‘The Swedish Networks planned the debt-conversion program’

4.2. Intransitive Verbs

We considered a verb as intransitive if our lexical database did not contain any transitive occurrence of it. Regaining the original argument structure, i.e. that of the base verb, was quite simple: the head of the NP is identified as the transformed sentence’s subject. Just as in the case of transitive verbs, complements and adjuncts of the base verb precede the participle. In the resulting sentence they follow the finite past tense verb. It might be interesting that the subject of base verbs belonging to this class is usually a patient.⁴ We used the rule below to transform such structures:

Det (V_compl/V_adj) V_{past_part} N → Det N V_{past_part} (V_compl/V_adj)
 ‘_{[part} A bécs - i kereskedelmi bíróság - on tegnap lezajl - ott] tárgyalás’
 DETVienna - ADJ mercantile court - SUP yesterday pass-off - SUFF trial
 ‘The trial which passed off yesterday at the mercantile court in Vienna’
 ‘A tárgyalás lezajlott tegnap a kereskedelmi bíróságon Bécsben.’
 ‘The trial passed off yesterday at the mercantile court in Vienna.’

As the example shows, the argument structure of intransitive verbs can be fully reconstructed from the NPs internal structure. Nevertheless, we have to note that intransitive verbs are less useful regarding information extraction. This is because they express less implicit information, since these participles are usually derived from verbs with only a vague semantic content. For instance: *bekövetkezik* ‘come true’, *beindul* ‘start up’, ‘be launched’, *létrejön* ‘come into existence’, *kialakul* ‘take shape’. Hence, the identification of their arguments adds probably no extra information to our existing knowledge. Still, such structures could be worth dealing with, since they might help us in bringing hidden relations to light.

5. Evaluation

We scrutinized a total of 7058 sentences, i.e., 43% of the whole corpus. As the corpus is split up in smaller texts according to their topic, we decided to evaluate

³ Unfortunately, some occurrences contradict this consideration. Occasionally we can only rely on our knowledge of the world when deciding whether the head noun’s possessor is equal to the subject.

⁴ There is an other use of this structure, the so-called newspaper language use. In such cases the subject of the base verb may be also an agent. (e.g. ‘[part A tegnap lemondott] elnök.’-‘The president, who resigned yesterday’). This use lays emphasize on the anteriority in time (Laczkó 2000).

the results on a set of randomly selected 500-sentence long fragments to avoid the useless repetition of the same patterns and the eventual lack of different kinds of structure.

The number of hits in the test corpus is 798, with a precision of 64%. We were not able to count recall values as there is no manually annotated Hungarian corpus which contains annotation of NP-internal participial structures. However, as this application is supposed to improve the efficiency of an IE system, where correctness is more important for the users than the amount of the output, we have reasons for focusing on precision.

Most of the improper hits are due to one of the following deficiencies:

1. Improper morphological analysis, hence inaccuracies in the dictionary are responsible for most of the false hits.
2. As it has been already mentioned, we cannot handle NPs without a determiner. However the more informative (and at the same time longer) NPs relatively often begin with a determiner.
3. Ensuing from the nature of the texts there are a lot of peculiar nouns in the corpus. Unlike a usual Hungarian text, the occurrence of firms' and trades' names is quite common, such as compound words with NN internal structure, the recognition of which is also difficult.
4. There are lexicalized participles, too. Although they might have preverbs, adjuncts or complements in their context they behave as adjectives (e.g. *elmúlt* 'bygone', *ismert* 'known').

As it is obvious from the error types, this rule system is a high-level language processing application, where most of the errors are due to problems of the lower level analysis such as morphology, dictionaries or named entity recognition. On the other hand, as it is a rule-based application, the deficiencies that its evaluation brings into light can be directly turned into useful information for outlining the directions of future development.

References

- Kenesei, I., 2000, «Szavak, szófajok, toldalékok». In: *Strukturális magyar nyelvtan* 3. ed. by Kiefer, F. pp. 75-135. Budapest. Akadémiai Kiadó. ('Words, Lexical Categories, Suffixes'. In: *Structural Grammar of Hungarian* 3.)
- Kiefer, F., 2000, *Jelentélmélet*. Budapest. Corvina ('Semantics').
- Komlósy, A., 1992, «Régensek és vonzatok». In: *Strukturális magyar nyelvtan* 1. ed. by Kiefer, F. pp. 299-529. Budapest. Akadémiai Kiadó ('Predicates and Complements'. In: *Structural Grammar of Hungarian* 1.)
- Laczkó, T., 2000, «A melléknévi és határozói igenévképzők». In: *Strukturális magyar nyelvtan* 3. ed. by Kiefer, F. Budapest. Akadémiai Kiadó. ('The Suffixes of Adjectival and Adverbial Participles'. In: *Structural Grammar of Hungarian* 3.)
- Prószéky, G., 2003, «Information Extraction from Short Business News Items». In: *Proceedings of the First Hungarian Computational Linguistics Conference*, ed. by Alexin, Z. and Csentes, D. pp. 161-167. Szeged. Szeged U. P.

- Silberztein, M., 1993, *Dictionnaires électroniques et analyse automatique de textes: Le système Intex*. Paris. Masson.
- Vajda, P. et al., 2004, «A ragozási szótártól a NooJ morfológiai moduljáig». In: *Proceedings of the Second Hungarian Computational Linguistics Conference*, ed. by Alexin, Z. and Csendes, D. pp. 183-190. Szeged. Szeged U. P.

THE ACQUISITION OF BASQUE ERGATIVE CASE AN EXPERIMENTAL STUDY

Maidier Huarte

University of the Basque Country

Basque children omit ergative case markers for about five months before the production of this case mark becomes adult-like. This has been considered a problem in the acquisition of Basque and has been related to the ergative character of the language. The aim of this work is to present the results of a picture selection task done by 24 Basque bilingual children which show that the comprehension of Basque case marking and, more precisely, of the ergative case precedes the production of it.

Based on production data, Basque acquisitionists have pointed out that children have some trouble in acquiring the case marking system of the language. The difficulty would be induced by the ergative character of the language. However, and from a Universal Grammar point of view, the type of language to be acquired should not impose any extra burden on the learner. In other words, from a learnability perspective, both accusative and ergative languages should be equally difficult or simple to acquire. So, what is the problem? Is it really that ergative languages are more complex to acquire or is it just a matter of how to look at the data?

The goal of this paper is to present the comprehension data of 24 bilingual children, data which will show that children have internalized Basque case marking system long before they are able to produce case marks in an adult manner. In order to do so, first, I will explain the features of Basque case marking, and then, in section two, the findings in the acquisition of Basque on which I base my study. Next, in section three, I will outline Gerken and McIntosh (1993), another study that informs mine. Finally, in section four, I will present the experimental task designed to isolate what has been considered a problem in the acquisition of Basque, as well as the results and some discussion of these.

1. Basque is an ergative language from the point of view of morphology

Languages can be accusative or ergative. Being accusative at the morphological level implies that both the subjects of transitive and subjects of intransitive verbs bear

* This research has been supported by a grant from the Basque Government BFI.078. I would like to thank Cecile McKee, Pat Bolger, Itziar Idiazabal and Xabier Artiagoitia for their helpful comments on earlier versions of this work. Also this research would not have been possible without the nice collaboration of all these kids.

the same case, *nominative*, whereas the object of transitive verbs is assigned a different case, *accusative*. English is an example of a morphologically accusative language:

- (1) a. He_{NOM} left b. He_{NOM} phoned him_{ACC}

Basque is an ergative language at the morphological level and assigns different cases to both types of subjects. Subjects of intransitive sentences, as well as objects of transitive sentences, bear *absolutive* case. Subjects of transitive sentences are, in turn, assigned *ergative* case. Finally, second objects are assigned *dative* case. The null morpheme $-\emptyset$ corresponds to the absolutive case, $-k$ morpheme corresponds to the ergative case and $-(r)i$ corresponds to dative case. All these morphemes are attached at the end of the argument phrase bearing case, as the table and examples in (2) illustrate:

(2) **Table 1**

	<i>Subject</i>	<i>Direct Object</i>	<i>Second Object</i>
a. Intransitive V	$-\emptyset$ abs	—	—
b. Transitive V	$-k$ erg	\emptyset abs	—
c. Ditransitive V	$-k$ erg	\emptyset abs	$-(r)i$ dat

- (3) a. Mutil-a- \emptyset bizikleta-z etorri da
 Boy-det-abs bike-by come INT-aux
The boy came by bike
- b. Krokodilo-a-k oinetako-a- \emptyset jantzi du
 Crocodile-det-erg shoe-det-abs put on TRA-aux
The crocodile put on the shoe
- c. Krokodilo-a-k mutil-a-ri oinetako-a- \emptyset jantzi dio
 Crocodile-det-erg boy-det-dat shoe-det-abs put on DIT-aux
The crocodile put the shoe to the boy

The verbal system in Basque also reflects agreement with the three types of DP arguments. The verb, depending on its argument structure, selects an intransitive, a transitive or a ditransitive auxiliary. The verb bears aspectual markers whereas the auxiliary is specified for the person and number of the arguments in the sentence as well as for tense information. Since the focus of this paper is on case marking of DPs and, more specifically, on ergative morphemes, I will not give a detailed description of how Basque verbal system works. Let us now turn to how the case system is acquired.

2. The acquisition of Basque case marking system

The data I will present first was collected for the research project HEGEHJ-BUSDE, conducted by the University of the Basque Country in cooperation with the University of Hamburg.¹ The project aimed to describe how monolingual and

¹ HEGEHJ stands for *Haur Euskaldun eta Gaztelaniadun Elebidunen Hizkuntz Jabekuntza* ‘The acquisition of language by Spanish and Basque bilinguals’. BUSDE stands, in turn, for *Baskisch und Spanisch: Doppelter Erstspracherwerb* ‘Basque and Spanish: the acquisition of two first languages’.

bilingual children acquired Basque, Spanish, or both languages simultaneously. In order to do this description, one Basque monolingual child and three bilingual children were video-taped while playing either with their peers or with their parents. The recording sessions lasted 30 minutes and had a frequency of every other week since children were 1,07 until they were 5,00 years of age. Some years later, Zubiri (1997) added to the description the study of another two Basque monolingual children. Before I go on to explain how these children develop Basque, I would like to draw the reader's attention to the fact that the bilingual children studied developed the language in the same way as the monolingual children.²

2.1. Stages in the acquisition of Basque case marking

As happens with the acquisition of other languages, Basque children, too, seem to go step by step in showing their competence in the language. All authors agree in stating that Basque children go through three different stages in acquiring case marking.

The productions from the first stage are characterized as being two-worded and caseless. All the words used in this stage belong to lexical categories, that is, children use adjectives, nouns, verbs or even adverbs but do not produce functional elements yet: there are no case morphemes; no aspect morphemes (born by verbs) and no auxiliaries are used:³

- (4) Jurgi atara (5) Aitita aputu
 Jurgi take out Granpa break
 Jurgi takes it out *Granpa broke it*
- (6) (*Adult*) Ta hemen zeñek itten do lolo?
 And who sleeps here?
 (*Child*) Egos
 Egoitz- erg missing (Egoitz is a Basque masculine name)
- (7) (*Adult*) Ta nori esango dotseu etortzeko gurekin
 And who will we ask to come with us ?
 (*Child*) Amane-dative missing
 To Amane (Amane is a girl's name)

In example (4), *Jurgi*, the subject of the transitive verb 'take out', should bear the ergative morpheme but the child has not produced the case mark. In the same man-

² In all the longitudinal studies conducted by HEGEHJ-BUSDE, it was concluded that the monolingual and bilingual children developed Basque in the same manner. All kids went through the same stages with some differences in the time of appearance of certain structures. However, and due to the small amount of children studied, these differences could be paired more with individual differences (found in the course of acquisition of all languages) rather than be taken to be caused by the different modes in the acquisition of Basque (bilingual or monolingual). A recent study carried out by Ezeizabarrena et al (2005) confirms the diagnosis that Basque monolingual children and Basque dominant bilingual children have a parallel development in the process of acquisition of the lexicon, and also, of the grammatical cases: absolutive, dative and ergative.

³ I will interpret the examples and assign cases where children have not produced them based on the context where these have been uttered.

ner, *Aitita* in (5), which is the subject of the transitive verb ‘break’ *apurtu* should be marked with the ergative case but the child has not produced it. The answer to the question posed by the adult in example (6) requires the ergative case mark but the child produces the name without any morpheme. In (7), the answer to the question ‘who will we ask to come with us’ needs to be marked with the dative morpheme. However, the child produces the answer without the dative marker.

The turning point in the acquisition of case marking in Basque comes when children are around 2,04 years of age. From a morphological point of view, children’s productions are longer and also more complex. Kids combine more than two words in their productions and also start using case marks though not in all the contexts where these are needed. Many of the contexts that require the use of a case morpheme (ergative, mostly) remain unmarked in children’s productions. Because Basque is a pro-drop language, not all arguments need to be overtly produced. However, whereas transitive subjects are left unmarked in many contexts during this second stage, whenever an argument requiring dative case is produced it is also overtly marked, as opposed to what happens with the ergative case. Let us analyze some examples:

- | | |
|--|---|
| <p>(8) eba(g)I nik
cut I-erg
<i>I cut it</i></p> | <p>(9) Asunek ekarrita
Asun-erg brought
<i>Brought by Asun</i></p> |
| <p>(10) Egoitzeri emaman hau
Egoitz-dat bring this
<i>Bring this to Egoitz</i></p> | <p>(11) Ni jan dut
I-erg-missing eat TRANSaux
<i>I have eaten</i></p> |
| <p>(12) Ni ez to bota
I-erg-missing neg TRANSaux throw
<i>I have not thrown it</i></p> | |
| <p>(13) Ni kantatuko dut
I-erg-missing sing-fut TRANSaux
<i>I will sing</i></p> | |

Whereas in examples (8-9) children produce the ergative morpheme required by the transitive subjects (*ni-k*, *asun-ek*), in the same stage, other transitive subjects (examples 11-13) are not properly marked. In contrast, children properly mark all overt arguments needing the dative morpheme [ri].

It is important that I bring here the conclusions drawn by the acquisitionists who studied the development of Basque. First, these authors found out that, whereas children frequently omit the ergative mark on transitive subjects, they rarely use the ergative mark on other arguments requiring either the absolutive or the dative case. Therefore, the errors made are errors of omission but never of commission (at least not at a significant level). In addition to this, if any error is to be found in the corpora, the error is always related to the ergative case, no dative or absolutive cases are mis-assigned by the children studied. As we will see, these data are extremely relevant so as to test the hypotheses that have been proposed in order to explain the ‘ergative dance’. But let us first describe the third stage in the acquisition of case marking.

Five or six months after the first case mark is produced, and following Brown’s criterion of 90% use, children’s use of grammatical cases increases and reaches adult

levels.⁴ The stabilization of the use of grammatical cases limits the beginning of the third stage:

- (14) hori amatxok aukin dau
 that mum-erg have-progress TRANS-aux
 mum has had that
- (15) beste kotxiei ipini bi otzegu pegatina
 other car-dat put must DITR-aux sticker
 We have to put the stickers on the other cars
- (16) amali paxatxen txio hau nahi dula amak
 mum-dat happens INT-aux this want TRANS-aux-rel mum-erg
 what happens to mum is that she wants this

In examples (14) through (16), we can see how children have adequately produced the ergative marks on the DPs. Also, all the DPs requiring dative case have been properly marked with dative case [ri].

Summing up, three different stages have been identified in the acquisition of Basque grammatical cases. In the first stage, children do not produce any case mark. In the second stage, children start using case marks but not in all contexts where these are required. Finally, in the third stage, children's use of grammatical cases reaches adult levels. Let us now turn to how scholars have explained this staged production of the three grammatical cases.

2.2. Explanations to the staged acquisition of case marking

All authors agree on suggesting that the acquisition of the ergative case is problematic for the children observed. This has been explained in two different ways. Barreña (1993, 1999), Zubiri (1997), Ezeizabarrena & Larrañaga (1996), basing their interpretations on the maturational hypothesis of language acquisition (Radford 1986, Meisel 1992), defend the view that the functional projections assigning case may not have been fixed yet during the second stage. Elosegi (1998) suggests that the absence of some ergative morphemes might be explained by the phonological context where they should have occurred.

The problem we face is that there are no data in the corpora favoring one or the other working hypotheses. Let me develop this idea a bit more, starting with the consequences of the first hypothesis, i.e., the syntactic hypothesis. If it were true that at this second stage the functional projections assigning case were not fixed yet, then we would expect that children would use ALL cases in a random way; in other words, we would expect to find commission errors in the corpora. As we have seen, and crucially, during this second stage of their language development, all errors made by the children are errors of omission but never of commission. Furthermore, the omission

⁴ Two criteria have been used in order to assess the acquisition of a certain grammatical case. First the case mark has to be used ON different arguments, i.e., ON different DPs, to avoid the possibility that the DP has been lexicalized together with the case mark. Second, at the time one case mark is found, other different case marks should also be used.

of case marks is restricted to the ergative case and it cannot be assessed for the absolutive case or the dative case.

On the other hand, if the problem were of a phonological nature, then we would expect to find other [k] ending morphemes (such as the one corresponding to plural or the one at the end of the partitive case [rik]) missing in children's productions.⁵ During this stage there are some instances of plural markers but these are certainly not comparable to the number of contexts where an ergative marker is needed. With regard to the partitive case [rik], its usage is not attested until the third stage. Interestingly, and contrary to what is believed to happen with the absolutive, dative and ergative cases, the acquisition of this case mark is not gradual. It is acquired some months later than the other three but its production is error free from the beginning. Therefore, and taking into account that during this second stage we find almost no instances of these phonologically similar markers, the comparison is difficult.

3. Gerken and McIntosh

Already in 1969, Shipley et al. claimed that production patterns do not necessarily reflect children's grammatical knowledge. In other words, these authors defended that the lack of certain elements in children's productions does not necessarily imply a lack of these elements in their grammatical competence.

Gerken & McIntosh (1993) proved that children as young as 2 who do not produce function morphemes are indeed sensitive to the linguistic contexts where these functors occur. This sensitiveness lead them to defend the idea that the functors these children were not producing did in fact belong to their grammatical system since children were using them in sentence processing tasks.

In addition to semantic or prosodic cues, it has been demonstrated that adults use function words to process the incoming speech stream (Greenberg 1963, Clark & Clark 1977). Gerken & McIntosh proposed that children and adults might share the same representation of functors. So if it were the case that children were using these same cues in an adult manner, then children could also be using functors to segment and label the incoming speech stream. Following this assumption, if children were actually using functors for segmentation purposes, then they should be able to identify phrases. Let us imagine, for instance, that children know that "the" and "was" are function words and that functors can either introduce or close phrases. Then the presence of these functors in the input would help them separating the speech stream into phrases. If in addition to this, children distinguished among the different types of functors and were sensitive to the specific contexts where they occurred, then identifying the functor "the" would automatically lead to at least partial recognition of a noun phrase. In the same manner, identifying the functor "was" would lead to the recognition of a verb phrase.

⁵ Even though the discussion of the status of the partitive case is not relevant for the purpose of this paper, I would like to note that some authors (Laka 1995, de Rijk, 1972) consider that the so-called partitive case is just a polar determiner restricted to cases of absolutive case.

These authors designed a picture selection task to test whether children were sensitive to the specific contexts where functors occur. The target words were tested in the following four conditions:

1. presence of grammatical morpheme before the target word
 - a. Find *the* bird for me.
2. presence of a grammatical morpheme but not the one required by the context -ungrammatical from now on
 - b. Find *was* bird for me
3. presence of a nonsense morpheme
 - c. Find *gub* bird for me.
4. no morpheme
 - d. Find * bird for me

16 experimental item types were tested. Each sentence was presented together with a choice of four pictures, one of which was related to the target word, the other three being distractors. The place where the correct picture appeared was balanced across all pages.

Results:

Children chose the correct picture more times when the target word was preceded by the grammatical morpheme “the” than when it was preceded by the ungrammatical “was” or the nonsense morpheme “gub”. The conclusion drawn from these results is that children identify functors and the specific contexts where they occur.

There was nevertheless an unexpected result. Children did not seem to differentiate between the presence of a grammatical morpheme and the absence of it. These authors give two possible explanations to this phenomenon. One reason why children did not make any difference between the two type of sentences could be based on the prosodic similarity between the two conditions (synthesized speech was used to make sentences with words and nonwords as uniform and natural as possible). The second reason they proposed was that it might be the case that the omission of the determiner is not a strong syntactic violation for children. They speculate with the fact that only singular count nouns must be preceded by an article in English and, therefore, children could be treating the determiner as an optional element in the structure.

4. My experimental study

I based my study of Basque children’s ergative markers on Gerken and McIntosh’s (1993) experiment. Recall that Basque and English are different in that the functional elements tested by Gerken and McIntosh were free morphemes, whereas the ones to be tested in Basque are bound morphemes (attached at the end of the argument DP). So, I tested full sentences since this is the context where grammatical cases and the morphemes associated to them occur. The conditions tested were:

1. ERG: a transitive sentence where the subject bears the ergative morpheme [k] -a function morpheme in its corresponding position:
 krokodilo-a-**k** oinetako-a jantzi du
 crocodile-the-**erg** shoe-the put on TRANS-aux
the crocodile put on the shoe
2. ABS: a transitive sentence where the subject bears a grammatical morpheme but not the one required by the context: [-Ø], i.e., the absolutive case marker:
 krokodiloa-**Ø** oinetakoa jantzi du
 crocodile-the-**abs** shoe put on TRANS-aux
the crocodile put the shoe on
3. NONS: a transitive sentence where the subject bears an ungrammatical morpheme: [-l], nonsense morpheme (NONS) corresponding to a sound in the language:
 Krokodiloa-**l** oinetakoa jantzi du
 crocodile-the-**nons** shoe put on TRANS aux
the crocodile put the shoe on

Before I go on, I would like to point out the fact that I am not proposing that children use ONLY syntax to process the information they receive. My sole intention is to show that syntactic competence may be there from the beginning even when children's productions seem to be indicating the opposite. I will be keeping the semantic and prosodic cues constant across experimental sentences. The only information that will vary from sentence to sentence will be the grammatical morpheme used to mark the subject of the sentence. If we find that this minimal morphosyntactic variation has an effect on children's comprehension and depending on the kind of effect we find, we might be in a position so as to defend that the syntactic competence is already there.

The predictions are:

1. Children should perform better on the sentences with a grammatical morpheme (be it the ergative morpheme or the absolutive morpheme) than on those marked with a nonsense morpheme (ERG and ABS conditions tested against NONS condition)
2. Assuming that children are aware of the specific contexts where function morphemes occur, children's performance on sentences with a grammatical morpheme should be better than the performance on sentences with a grammatical morpheme other than the one required by the linguistic context (ERG condition tested against ABS condition)

4.1. Method

Subjects. A total of 29 children all ranging in age from 2,04 to 2,09 years of age were tested at their schools. Five of the children failed to meet the criterion for inclusion so I will be reporting on the results of all other 24 children.⁶

⁶ The second stage in the acquisition of grammatical cases has been taken to start at 2,04. Thus, the subjects of my experiment are right in the middle of this stage when the grammatical system has not been fixed yet. This is why it is so interesting to take a look at how these children behave at the comprehension level.

My observation of these children in their classes supports the claim that they were at the second stage of acquisition reported by Basque acquisitionists. All of these children were Basque-Spanish bilinguals. 16 of them spoke Basque at home with both their parents; the rest acquired the language from just one parent: 4 from their mothers and 4 from their fathers. I won't report on this aspect here, but I would like to underline the fact that there are no differences in these children's responses based on the source of acquisition of the language.⁷

Stimuli. Twelve experimental sentences were created (see appendix 1). The items resulting from applying the ERG; ABS and NONS conditions to the experimental sentences were distributed in three different lists (see appendix 2):

<i>Experimental type:</i>	Krokodiloak oinetakoa jantzi du
<i>ERG condition:</i>	Krokodiloa-K oinetakoa jantzi du → List 1
<i>ABS condition:</i>	Krokodiloa-Ø oinetakoa jantzi du → List 2
<i>NONS condition:</i>	Krokodiloa-L oinetakoa jantzi du → List 3

This way, I came up with 4 stimuli marked with the ergative morpheme (grammatical and adequate in the context), 4 stimuli marked with the absolutive morpheme (grammatical but inadequate in the context) and 4 stimuli marked with the invented nonsense morpheme in each of the lists. In addition to this, the subject of the sentence bearing the target case mark preceded a word beginning with a vowel to facilitate children's perception of the case mark. Children were randomly assigned to one of the three lists (eight subjects for each list). Four intransitive filler sentences were added to make sure that the children were paying attention to the task. Whenever a child failed to choose the correct picture after at least three of these four filler sentences, she was rejected. Also, two training sentences were included at the beginning of the battery to make sure that the child understood the task. These training sentences were intransitive too. The training and filler sentences were same for all three lists and occupied the same position in the battery across lists (see appendix 1).

A big book with 4 pictures in each page was presented to the child. The pictures represented in each of the pages corresponded to the following situations, with the target verb always represented:

The crocodile put on the shoe

- target sentence - *the crocodile put on the shoe*
- same subject, different object - *the crocodile put on the glove*
- different subject, same object - *the mouse put on the shoe*
- different subject and different object. - *the mouse put on the glove*

The place on the page where the correct picture appeared was balanced across the 16 pages (experimental + filler). Also, the number of subjects across lists was balanced.

⁷ As it has already been pointed out, although these children are Basque-Spanish bilingual, the dominant language for them is Basque. I find it worth looking at the data from a comprehension point of view also (see footnote 2) to see whether the quantity of the input in a given language has an influence on the comprehension of certain structures which, from a production point of view, have not been accounted for.

Procedure. I spent quite a long time in the schools with the children before I proceeded to do the test. Since the subjects to be included in my study were young, I thought it important to spend some time with them in their classroom, playing and helping in their daily routines, so that these children became familiar with me. Once the children felt comfortable, the task of taking them out of class to do the test wasn't problematic.

Piloting demonstrated that it was important that I took some time to show the testing place to the younger kids, first, to give them the chance to explore the testing room before the real test took place, which would help to avoid distractions on the testing day. Second, by taking children to the testing room, I made sure that the kids got used to being out of class and that they felt comfortable being away from the safety of their teachers. A third goal of these visits was to "train" the instructions I would be using in the test as well as to make children familiar with the task I would be asking from them. Whenever I got the kids to the testing room, I brought a book with me so that I could play a game with them. I told them they had to find what I told them in the book. This way, the day of the test I just had to repeat the same game with the kids. The testing instructions were as follows:

Today I brought a big book with lots of pages, and in each of these pages there are four pictures. Let's count them (the kids are learning to count in class). *And have you noticed that the pictures are different? What's this?* (Signalling to all pictures in the training board)(...)

Once I had made the child realize what was different in each of the pictures I proceeded to explain what I expected them to do). *So, we will be playing the same game we played yesterday: I will tell you something and you will show where this is in the pictures, OK?*

4.2. Results

To begin with, and taking into account that children selected the correct picture on average 72% of the time (chance is at 25%), I would say that these children have performed very well in the task.

Secondly, just looking at the right answers and as happened with English-speaking children, Basque children, too, seem to perform better if the stimulus they hear is the one with the grammatical morpheme in the required context. These are the mean percent correct picture choices:⁸

1. after *-k* ergative morpheme: 84%
2. after *-Ø* absolutive morpheme: 72%
3. after *-L* nonsense morpheme: 59%

Several Wilcoxon's tests reveal that the differences between these means are significant. The difference between children's responses to the sentences marked with the ergative morpheme (1) (grammatical and adequate) and the ones marked with the absolutive morpheme (2) (grammatical but inadequate) is statistically significant ($Z=-2.527$;

⁸ In Huarte (2007), I include data from another 42 children ranging in age from 2;10 to 4;01. See chapters 5 through 7 for further discussion.

$p < 0.05$). The difference between children's responses to the sentences marked with the ergative case (1) and the ones marked with a nonsense morpheme (3) is also statistically significant ($Z = -3.447$; $p < 0.01$). Finally, the difference between children's responses to sentences marked with the absolutive morpheme (2) (grammatical but inadequate) and the ones marked with the nonsense morpheme (3) is statistically significant too ($Z = -2.144$; $p < 0.05$).

The type of list given to the children did not have any significant effect on these children's responses. It is also worth noting that the position of the drawings did not have an influence on how children did the task ($U = 5.121$; $p = 0.163$) either.

5. Discussion

Children's good performance on the test lead to think first that this kind of task is easily carried out by children this young and that it is therefore suitable to test different levels of the acquisition of grammar.

Second, the results of this test show that Basque children who are not reliably producing case markers do actually seem to know where these should occur. The fact that children respond better if the stimulus they hear is marked with a grammatical morpheme (be it ergative or absolutive) than if the stimulus they hear is marked with a nonsense morpheme suggests that kids are able to distinguish between functional elements belonging to their language and nonsensical elements.

Third, the fact that children respond better to stimuli if adequately marked (with the ergative) than if inadequately marked (with the absolutive) suggests, in turn, that children distinguish between these two functional elements.

These results go against the syntactic account of the ergatives missing since the semantic and prosodic information do not vary across sentences, and the only information varying being the morpheme in the sentence (ergative, absolutive or nons), it seems that children already know this syntactic information and also the type of element expected in each context. So, we may conclude that children already have the syntactic competence even though they do not seem to be as good performers as they should be.

In this particular aspect of the language, comprehension precedes production. Now, we are left with the task of explaining why children are not producing ergative morphemes in a consistent manner. The fact that the ergative marker is a word-final voiceless stop, a marked option in the sonority hierarchy and a possibility which is restricted to word-final coda positions suggests that a phonological explanation might be in order.

References

- Barreña, A., 1993, *Haur euskaldunen hizkuntzaren jabeakuntza-garapena: INFL eta KONP funtzio kategorien erabileraz* 'The acquisition and development of Basque. On the use of INFL and COMP categories'. University of the Basque Country: Doctoral dissertation.
- , 1999, «Ausencia de reglas gramaticales en las primeras producciones lingüísticas de los niños vascos», *Infancia y aprendizaje* 86. 91-117.
- , M. J. Ezeizabarrena and I. García, in press, *Influence of the linguistic environment in the development of lexicon and grammar of Basque bilingual children. A portrait of the young in*

- the new multilingual Spain*, ed. by C. Perez-Vidal, M. J. Garau and A. Bel. Multilingual Matters.
- Clark, H. and E. V. Clark, 1977, *Psychology of language. An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Elosegi, K., 1998, *Kasu eta preposizioen jabekuntza-garapena haur elebidun batengan* 'The acquisition of case and prepositions by a bilingual child'. University of the Basque Country: Doctoral dissertation.
- Ezeizabarrena, M. J. and M. P. Larrañaga, 1996, «Ergativity in Basque: a problem for language acquisition», *Linguistics* 34. 955-991.
- Gerken, L. A., and B. J. McIntosh, 1993, «Interplay of function morphemes and prosody in early language», *Developmental Psychology* 29. 448-457.
- Greenberg, J. H., 1963, «Some universals of grammar with particular reference to order of meaningful elements». *Universals of grammar*, ed. by J. H. Greenberg, 73-113. Cambridge, MA: MIT Press.
- Huarte, M., 2007, *Ergatiboaren jabekuntza: ikerketa experimental bat eta beraren emaitzak*. University of the Basque Country: Doctoral dissertation.
- Meisel, J. M., and N. Muller, 1992, «Finiteness and verb placement in early child grammar: Evidence from simultaneous acquisition of French and German» in *Bilinguals. The acquisition of verb movement*, ed. by J. M. Meisel, 109-138. The Netherlands: Kluwer Academic Publishers.
- Shiple, E. F., C. S. Smith, and L. R. Gleitman, 1969, «A study in the acquisition of language: Free responses to commands», *Lg* 45. 522-542.
- Zubiri, J. J., 1997, *Izen sintagmaren determinazioa eta kasuen jabekuntza eta garapena hiru urte arte* 'The acquisition and development of cases and NP determination'. University of the Basque Country: Doctoral dissertation.

Appendix 1

Here is the battery presented to the child:

Training sentences:

'Neska zuhaitzaren ondoan dago'

The girl is next to the tree

'Txoria mutilaren eskuan dago'

The bird is on the boy's hand

Filler sentences:

'Mutila puztukiarekin jolasten dabil'

The boy is playing with the balloon

'Neska negarrez dago'

The girl is crying

'Sagarra mutilaren buruaren gainean dago'

The apple is on the boy's head

'Neska bizikletaz etorri da'

The girl came by bike

Experimental sentences:

'Krokodiloak oinetakoa jantzi du'

The crocodile put on the shoe

'Behiak irratia zapaldu du'

The cow stepped on the radio

'Txerriak azenarioa jan du'

The pig ate the carrot

'Tximinoak akordeoia apurtu du'

The monkey broke the accordion

'Elefanteak euritakoa zabaldu du'

The elephant opened the umbrella

'Saguak atea margoztu du'

The mouse painted the door

'Oiloak arrautza ipini du'

The hen laid on the egg

'Sugeak aulkia harrapatu du'

The snake trapped the chair

'Katuak eskularrua izkutatu du'

The cat hid the glove

'Txakurrak hezurra topatu du'

The dog found the bone

'Zaldiak hegazkina ikutu du'

The horse touched the plane

'Pinguinoak etxea egin du'

The penguin built the house

Appendix 2

Here are the three lists children can be assigned to:

<i>First list</i>	<i>Second list</i>	<i>Third List</i>
Neska zuhaitzaren ondoan dago Txoria mutilaren eskuan dago Mutila puztukiarekin jolasten dabil	Neska zuhaitzaren ondoan dago Txoria mutilaren eskuan dago Mutila puztukiarekin jolasten dabil	Neska zuhaitzaren ondoan dago Txoria mutilaren eskuan dago Mutila puztukiarekin jolasten dabil
Krokodiloak oinetakoa jantzi du	Oiloa arrautza ipini du	Oiloa arrautza ipini du
Zaldia hegazkina ikutu du	Krokodiloa oinetakoa jantzi du	Katuak eskularrua izkutatu du
Txakurra hezurra topatu du	Txerriak azenarioa jan du	Elefanteak euritakoa zabaldu du
Neska negarrez dago	Neska negarrez dago	Neska negarrez dago
Sugeak aulkia harrapatu du	Elefanteak euritakoa zabaldu du	Tximinoa akordeoia apurtu du
Oilok arrautza ipini du	Katua eskularrua izkutatu du	Txakurrak hezurra topatu du
Katuak eskularrua izkutatu du	Txakurral hezurra topatu du	Zaldia hegazkina ikutu du
Sagarra mutilaren buruaren gainean dago	Sagarra mutilaren buruaren gainean dago	Sagarra mutilaren buruaren gainean dago
Sagual atea margotu du	Behial irratia zapaldu du	Behiak irratia zapaldu du
Elefantea euritakoa zabaldu du	Sugea aulkia harrapatu du	Txerria azenarioa jan du
Tximinoal akordeoia apurtu du	Tximinoak akordeoia apurtu du	Sugeal aulkia harrapatu du
Neska bizikletaz etorri da	Neska bizikletaz etorri da	Neska bizikletaz etorri da
Txerrial azenarioa jan du	Pinguinoak etxea egin du	Krokodiloal oinetakoa jantzi du
Pinguinoal etxea egin du	Zaldial hegazkina ikutu du	Sagua atea margotu du
Behia irratia zapaldu du	Saguak atea margotu du	Pinguinoal etxea egin du

ON THE INTERACTION OF VELARS AND LABIALS

Daniel Huber

Eötvös Loránd University

Abstract¹

The paper treats data from a wide range of languages to show, on the one hand, that there is pervasive direct interaction between labials and velars to the exclusion of coronals, and, on the other hand, that the drive behind these phenomena is simply the presence of labiality in labials and the lack of any place specifications in velars. These data then further support the view that velars lack place specifications altogether, a view presented in Huber 2004b contra Paradis and Prunet 1991. Most importantly, the paper clearly shows that all these phenomena are in fact prosodic phonologically conditioned and absolutely regular, rather than random or unprincipled changes. On this basis, the paper sets up a new typology of the phenomena, which better captures the phonological conditions underlying them.

0. Introduction - the wider perspective of the research on velars

The present paper treats only one aspect of the thesis that all the observable phenomena related to plain velar consonants (notably /k g x ɣ ŋ/) can be accounted for if no phonologically relevant place of articulation is assumed in velars. In other words, the phonological representation of velar segments lacks any place specifications. This stance is not the mainstream opinion (cf. Paradis and Prunet 1991). However, arguments in favour of this view come from a number of directions such as issues of markedness, interactions between velars and palatals as well as issues of epenthesis, among others—all of them are topics already discussed elsewhere at some length (see Huber 2002, 2003a, 2003b, 2004a and 2004b). The topic of the present paper, the interaction of velars and labials, provides, in particular, a surprisingly rewarding area where the thesis can be affirmatively tested.

The data to be discussed below illustrate, on the one hand, that there is pervasive direct interaction between labials and velars, crucially excluding any coronal (or dental) involvement in these phenomena. Since the coronal (dental) space is excluded, these changes cannot be attributed to any place assimilation effects on the production side. Indeed, some authors (e.g. Ferreiro 1999: 116 and Schmidt 1993: 68)

¹ This paper is a considerably expanded chapter of (Huber 2002), of which Huber 2004b is a shortened version.

stress the acoustic similarity of velars and labials, that is, they attribute a central role to perception. On the other hand, the data presented below also support the view that the drive behind these phenomena is simply the presence of labiality in labials (expressed as some feature or element in phonological theories) and the lack of any place specifications in plain velars (features or other). The evidence presented here (and their theoretical account) clarifies an important point in the discussion, namely, what supports that velars can be said to lack a place of articulation.

The paper is structured as follows. In section 1, a preliminary typology of the relevant phenomena is presented (largely based on Huber 2002: 31-35 and Huber 2004b: 27-30). Section 2 presents an extensive range of labial-velar interactions attested mainly in the diachronic changes of quite a number of languages. Section 3 is an attempt to analyze the data in the framework of Government Phonology (with only a brief explanation of the necessary theoretical assumptions, though). The last section summarizes the findings, and references close the paper.

1. About the typology

There seem to be at least two ways to classify the various phenomena that show interactions between labials and velars: (1) whether they occur frequently in natural languages; and (2) whether there is some phonological motivation behind the phenomena. The second approach is admittedly and conspicuously more phonological, while the frequency approach will turn out to be the result of mere lack of data and their unsatisfactory understanding. In Huber (2002, 2004b) no typology had been set up, although a mixture of these two approaches is implicit. The combined approach seemed promising there since the primary emphasis was on drawing attention to the facts themselves, hardly ever described systematically in the phonological literature, while trying to give a theoretical account for most of the phenomena concerned at the same time and pointing out more problematic cases. Consequently, cases that could be handled more easily in the theory were termed 'typical' phenomena while others were termed 'atypical'. This is presented in (1) below. Since this typology is a convenient point of departure for the following discussion of the individual cases, reference to the later sections of this paper is also indicated below.

- (1) The typology (to be modified):
 - 2.1 Typical (=frequently attested) phenomena:
 - 2.1.1 Non-phonologically conditioned phenomena
 - 2.1.2 Phonologically conditioned phenomena
 - 2.2 Atypical (= less frequently attested) phenomena

Interestingly enough, the various interaction phenomena have turned out to show a biased combination of these two perspectives in that atypical phenomena tend to be phonologically conditioned while typical phenomena are either so conditioned or not. The most important conclusion here will be that all these phenomena are in fact phonologically conditioned and regular: certain changes systematically occur in pre-vocalic, others in preconsonantal and word-final contexts. This will lead to a reconsideration of the initial typology shown in (1).

The most important point, however, in the subsequent argumentation is that all the observed phenomena can be directly explained by the presence or absence of a labiality element (the equivalent of [+labial] in featural terms), which is assumed in labials anyway (cf. Harris and Lindsey 1995: 65-73, Cyran 1997: 24). Velars, it will be shown, do not need to be assumed to have place specifications at all.

2. The data and their analyses

2.1. Typical phenomena

2.1.1. *Non-phonologically conditioned phenomena*

2.1.1.1. Among the Indo-European (IE) languages, members of the Celtic branch have been cited most often to show a well-known phenomenon where an IE labio-velar consonant (for instance /k^w/) turns either into a plain velar or labial. Of particular interest here is, of course, the change to a plain labial. Two paradigm examples for the change in Celtic are shown below with some other IE cognates for the sake of comparison:

- | | | | |
|-----|--------------------------------|---|--|
| (2) | IE *ek ^w o- ‘horse’ | > | Ogam Irish <i>ech</i> /ex/ versus Welsh <i>ebol</i> ‘colt’
(Schmidt 1993: 68) |
| | also | > | Latin <i>equu-</i> /-kw-/ , Old English <i>eoh</i> /-x/ |
| | | > | AncGrk <i>hippo-</i> /-pp-/ |
| | IE *-k ^w e ‘and’ | > | Lepontic Celtic <i>-pe</i> /p-/ (Eska-Evans 1993: 44) |
| | also | > | Latin <i>-que</i> [kwe] > Spanish <i>que</i> [ke] |

As can be seen, Indo-European *k^w turned into /p/ in the so-called P-Celtic languages such as Welsh and Lepontic, while it remained a velar, even if later simplified to a plain /k/, in Q-Celtic languages such as Ogam (Old) Irish. Based on this dialectal feature, Celtic languages fall into two types as charted below (Schmidt 1993: 68, also cited in Huber 2004b: 28):

- | | | | |
|-----|-----------------|---|--|
| (3) | *k ^w | > | /k ^w / (> /kw/) in Celtiberian, Ogam Irish, Archaic Gaulish |
| | | > | /k/ in Goidelic: Modern Irish, Scottish Gaelic |
| | | > | /p/ in Brythonic: Welsh, Breton; Lepontic (Gaulish) |

What is significant phonologically at this point is the observation that these changes to /p/ are not conditioned by a triggering segment in the environment of the labio-velar consonant. In other words, the development to /p/ is not the result of any kind of place assimilation or other. At the same time, it is equally obvious that the labial glide /w/ in the labio-velar is the only possible source of the labiality. It is a case of reconfiguration then within an initially complex segment. Although there is no contextual triggering segment, there is phonology behind the curtains—but let us see some other examples first.

2.1.1.2. Some examples for the very same kind of splits are also attested in other IE languages, which shows that the change is far from being irregular, exceptional or

rare in any sense. For instance, Latin, as we have just seen, had retained the labio-velars. From the Italic languages, however, Latin is the only such variety, since neighbouring Osco-Umbrian varieties came up with plain labial reflexes uniformly. This means that the Italic branch showed exactly the same kind of dichotomy as did the Celtic branch, with Latin retaining labio-velars while other Italic languages turning them into labials. The correspondences are regular between Latin /kw/ and Osco-Umbrian /p/ as far as the scarcity of Osco-Umbrian data allows us to see. The following is a brief illustration:

(4)	Latin	Oscan	Umbrian	
	<i>quis</i> /kw-/	<i>pis</i> /p-/	<i>psi</i> /p-/	‘who?’ (Fodor 2000: 1494)
	a sentence in Oscan:	<i>status</i>	<i>/p/us set hurtin</i>	(Fodor 2000: 1122)
		statues	which are in the garden	

Continuing with the discussion of Latin, it is noteworthy that some Romance languages turned Latin labio-velars into plain labials. Romanian and Sardinian have this feature, again without any contextual restrictions. The Romanian data also reveal that both voiceless /k^w/ and voiced /g^w/ were affected. Here are some examples from Romanian and Sardinian with the corresponding Latin items (data from Tamás 1976):

(5)	Latin	Romanian	Sardinian	
	/kw/	/p b/	/p b/	
	<i>aqua</i>	<i>apă</i>	<i>abba</i>	‘water’
	<i>equa</i>	<i>iapă</i>		‘mare’
	<i>lingua</i>	<i>limbă</i>		‘language’
	<i>adaquare</i>	<i>adăpă</i>		‘to take to water’
	<i>quattro</i>	<i>patru</i>	<i>battoro</i>	‘four’
	<i>qui</i>	<i>pe</i>		‘that <conj.>’
	<i>cinque</i>		<i>kimbe</i>	‘five’

As for other Indo-European languages, the Germanic branch, for instance, preserved the IE labio-velars, which show, of course, later effects of Grimm’s Law: IE *k^w > /x^w/: as in OE *hwa* ‘who’, OE *hwat* ‘what’, and IE *g^w > /k^w/: as in OE *cwicu* ‘alive’ (> E *quick*), Dutch *kwi(e)k* ‘quick, alive’, IE *g^wena > OE *cwena* ‘woman’ (> E *queen*), Southern Dutch *kween* ‘old woman’. Slavic languages merged labio-velars into plain velars, original palatal velars having become some sibilant (other satem languages had similar reflexes).

There are some sporadic alternations, nevertheless, whose theoretical importance seems to be little (data partly taken from Huber 2004b: 29):

(6) Sporadic correspondences

/f, v/	/k, kw/
wolf	Old Slavic * <i>wilku</i> > Czech <i>vlk</i> ‘wolf’, Polish <i>wilk</i> ‘wolf’, etc.
four	L <i>quattuor</i> , <i>quartus</i>
five	L <i>quinque</i>

The word *wolf* seems to be an isolated example. Since the word-final /f/ comes regularly from /p/ as derived by Grimm's Law, the change to /p/ must have preceded the Germanic Consonant Shift. It is interesting that Latin also has *lupus* 'wolf' with /p/, which is unexpected. The only explanation (if borrowing from, say, Osco-Umbrian can be excluded) is phonetic in nature: the neighbouring /u/ may have had an influence on the etymological /kw/. The Germanic word for *four* also goes back to an initial IE /k^w/. A standard explanation in this case, however, is the analogical influence of *five*, which regularly goes back to an etymological IE /p/. In this latter word, however, Latin shows analogical influence since the initial *kw-* of the cluster in *quinque* is not etymological. It is either the influence of Latin *quattuor* 'four', or the assimilation of the second /kw/ of *quinque*. Nevertheless, the examples in (6) are considered slightly deviant in form; they do not represent the regular state of affairs, which is that Latin and Germanic both preserved the IE labio-velars and Slavic (as well as other satem languages) came up with plain, but still velar, reflexes.

2.1.1.3. Ancient Greek is more revealing than it might seem at first sight, therefore it deserves attention. Ionic and Attic dialects of Ancient Greek show some remarkable changes of IE labio-velars *k^w, *g^w, *g^{wh}. On the one hand, there are regular plain labial reflexes of IE labio-velars: Lat *se[kw]i-* - Gr *(h)e[p]e-* 'follow', Gmc [k] *u* - Gr [b] *ous* 'cow', etc. However, the fate of these clusters seems in fact to have been determined by the following vowel: only when the vowel was one of the back vowels /a o u/, did the change to a plain labial ensue. When the following vowel was front /e i/, developments to dentals are found instead, which is truly remarkable. The data in (7) below show some examples for the changes to dentals:

(7) *k ^w	→ t	*k ^w e	> /t/e	'and'
		*k ^w is	> /t/is	'who?'
		*k ^w et ^w ores	> /t/ettares	
			or /t/essares	'four'
		*penk ^w e	> pen/t/e	'five'
		*k ^w ei/k ^w oi/k ^w i	<root of 'pay'>	
		*k ^w i-ti-	> /t/i-sis	
	but:	*k ^w oi-neh ₂	> /p/oiné	
*g ^w	→ d	* ṅ-g ^w en-	> a-/d/en-(os)	(cf. Lat. in-/gw/en 'hips, waist')
			(before /i/, however, often: *g ^w iyos > /b/ios 'life', */d/ios)	
*g ^{wh}	→ t ^h	*g ^{wh} en-je/jo-	<thematic impf. of 'kill'>	
		*t ^h en-jó	> 1sg. /th/einó	
	but:	*g ^{wh} on-o-s	> /ph/onos 'murder, killing'	

There are then morphological alternations between /p b p^h/ and /t d t^h/ in ancient Greek, but their actual morphophonological status is not investigated here. (It also has to be recalled that non-Attic varieties had regular developments such as IE *penk^we 'five' > pen/k/e.) The general developments in the Ionic and Attic dialects can be summarized as follows:

(8)	IE *k ^w -		/t-/
	IE *g ^w -	>	/d-/ / _____ [+front]
	IE *g ^{wh} -		/t ^h -/

Elsewhere: /p b p^h/, respectively.

It is truly noteworthy that in this case reference must be made to a following vowel, and also that all IE labio-velars are uniformly affected. This is going to gain importance in the following discussion. As for the actual motivation for this surprising change to dentals, some sort of palatalization is at work, but the details are not relevant in this discussion (see Rix 1976: 87).

There is a further complication in Ancient Greek, however. As early as Pre-Mycenean Greek, a change of the form *k^w > /k/ / _u or u_ took place, that is, *k^w became plain /k/ in the vicinity of /u/. In this variety, however, all other labio-velars remain intact. The word for 'shepherd' illustrates all the Ancient Greek changes particularly well. In Mycenaean Greek, there is /*gwoukolos*/ 'shepherd' rather than */*gwoukwolos*/: the only change here is the simplification of the middle complex labio-velar. If one compares this with the (later) Attic Greek form /*boukolos*/, then both the rather early change *k^w > /k/ and the later Ionic-Attic developments of the initial /g^w/ to /b/ can be seen.

A parallel development to that just described in Pre-Mycenean Greek, that is, delabialization under the influence of a neighbouring labial vowel, also occurred in Germanic languages where reflexes of IE *k^w have become simple /k/ before a labial vowel (and also at the end of words). Compare IE *g^wou- > Gmc *cu* > English *cow* /kau/, Gm *Kuh* /ku:/, D *koe* /ku:/ with simplification versus *cwicu* 'alive' with retained /kw/. Evidence for the original presence of /kw/ comes from Dutch, for instance, where the preterite form of the verb 'come' is still *kwam* (singular) -*kwamen* (plural) with /kw/ retained, while all other forms show the loss of the labial glide /w/: *komen* / **kwomen* 'to come' and *gekomen* / **gekwomen* 'come <past participle>'. This preterite form is also attested in Old English: *cwom* 'came' as opposed to the simplified form in, say, *cuman* 'to come'.

Returning to the discussion of Latin and Ancient Greek, one more important observation is in order here. In Ancient Greek simple /k/, which could occur either pre-consonantly or prevocally, does not undergo any changes comparable to those above:

(9)	Latin	Ancient Greek	
	se[ks]	he[ks]a	'six'
	de[k]em	de[k]a	'ten'
	[k]entu-	(he)[k]ato-	'(one) hundred'
but:	se[kw]i-	(h)e[p]e-	'follow'
	e[kw]u-	hi[pp]o-	'horse'

This observation is important because it shows that only complex labio-velars underwent the change, simple velars did not.

2.1.1.4. An important point in the theoretical analysis, to rush ahead a little, is the relevance of whether a sequence /kw/ or a single but complex phoneme /k^w/ is

assumed. The straightforward answer is that it does not matter. In fact, there is by and large agreement (see any handbook on IE comparative linguistics) that IE had labio-velar phonemes such as /k^w/ rather than sequences of a velar followed by a labial glide as in /kw/, for example. Evidence comes from metrical facts in diction, syllabicity facts and, of course, later historical developments. Now it will suffice to point out that assuming a sequence runs into a problem difficult to evade. Namely, if /k^w/ is really /kw/, then it has to be explained why /tw/, for instance, did not behave like /kw/ and, in particular, why it did not change into a plain labial in the course of time. This means in practice that the changes from a labio-velar (and exclusively from these) to a plain labial could only happen at a time when the original sounds were (still) a single phoneme.

2.1.1.5. Turning away from Indo-European languages, the following switches between Standard Chinese and Santai Chinese lend additional support for the view that only complex labio-velar segments, occupying one single timing unit, are capable of either splitting or switching. Duanmu cites (2002: 85) the minimal pairs in (10a) for such regular switches between the two Mandarin varieties, Standard Chinese (SC) and Santai Chinese. He also cites some words in (10b) for the lack of switches, to illustrate his point:

(10) Standard Chinese	Santai Chinese	
(a) [hwəi]	[fəi]	‘ashes’
[fəi]	[hwəi]	‘to fly’
[hwaŋ]	[faŋ]	‘yellow’
[faŋ]	[hwaŋ]	‘house’
(b) [hən]	[hən]	‘very’
[hau]	[hau]	‘good’

What is a labio-velar in SC is labial in Santai and vice versa (10a), while plain velars do not show such switches (10b). In analyzing the phoneme inventory of SC, Duanmu considers [h] to be one of the realizations of the velar fricative /x/ (2002: 27), which means that the above data are rightly considered to be labial-velar interactions. From the data above it is apparent that only the labialized velar [h^w] switches to [f] and vice versa, while plain [h] never does. He argues convincingly at great length (2002: 82-89) that a prenuclear glide (a /w/ in the case at hand) does indeed belong to the onset (it shares its timing slot). Consequently, the switches in (10) are only possible if [h^w] is in fact a single segment rather than a sequence [hw].

In addition, Duanmu has also confirmed (pc, 2005) that there are pairs of words with initial [f] in both dialects: “The only clear case where both dialects (SC and Santai) use [f] is when the vowel is labial, in particular for the syllable [fu1] ‘husband’ or [fu4] ‘father’.” And he goes on to say: “Also, there are words where both dialects use [h] [although one would expect alternation; addition mine]. This happens for the syllable ‘fire’, which is [hwo3] in SC and [ho] in Santai [not *fo; add. mine]. I believe the reason is that Santai does not have the syllable [fo].” What becomes clear from this comment is that the presence, in both dialects, of word-initial /f/ and the lack of expected alternation is due to the presence of a following labial vowel. Although the alternations in (10) support the view of the single segment analy-

sis, there is a problem: why does the switch work in the $f \rightarrow x^w$ direction as well (cf. 10a)? In this case, it seems, an original plain labial splits into a labio-velar —this is odd. Furthermore, it would be good to know whether there are such switches among other labials and velars as well, in particular with stops. Further investigation is needed here.

Notwithstanding these additional remarks, the data in (10) and the earlier Ancient Greek phenomena in (9) show that the relevant conclusion is that only single, if complex, labio-velars undergo a change to plain labials.

In connection with this latter claim, namely that only labio-velars can turn into labials while plain labials cannot turn into labio-velars, it has to be shown why the well-known diachronic change $/f/ > /b/$ in Spanish is not a counter-example. It is known that Latin initial $/f/$ - changed to $/h/$ - in Spanish as well as in some other neighbouring Romance varieties such as Gasconian (where it is much more consistent than in Castilian Spanish by the way). In this case, a plain labial $/f/$ turns into a plain velar $/x/$ (which was realized as $[h]$ and still later disappeared altogether) in exactly the same prevocalic position as did all the other phenomena treated so far. What is peculiar is that $/p/$, for example, does not undergo similar changes. Lapesa (1981: 38) attributes this change to a Basque substratum since Basque “seems to lack original $/f/$; in Latinisms it tends to omit it (*filu > iru*; *ficu > iko*) or substitute it with $/b/$ or $/p/$ (*fagu > bago*; *fiesta > pesta*). Moreover, Basque—including Vizcayan throughout the Middle Ages—used to have an aspirated $/h/$ which could also substitute $/f/$, with which it alternates.” Lapesa (*ibid.*) writes that “the initial focus of the phenomenon is limited in the ninth to twelfth centuries to the north of Burgos, La Montaña and Rioja.” What all this means for the present discussion is that this particular change happens to be a case of sound substitution, originally in Basque, from which it spread to areas under Basque influence—such is not the case in any of the phenomena discussed so far. In addition, this change is far from being as regular as any of the cases presented above.

All in all, these Spanish cases do not pose a serious objection to the claim that only complex labio-velars can undergo a change to plain labials, not the other way round.

2.1.1.6. A number of important conclusions emerge from the preceding discussion. First of all, although there is no contextual phonological motivation for the various phenomena, all the above changes seem in fact to be phonologically conditioned since they occur pre-vocalically and not pre-consonantly. This is true for all the phenomena discussed above: for the Celtic divisions into P-Celtic and Q-Celtic, Italic varieties (both Ancient and Romance), as well as the switches between Standard and Santai Chinese. In this way then, all velar-labial interactions are (prosodic) phonologically conditioned. This is a major observation, which has tended to be overlooked in the literature (including Huber 2002 and Huber 2004b). Second, the above changes provide considerable support for the view that only complex labio-velars can turn into plain labials (or plain velars, of course). Plain velars and plain labials cannot undergo any comparable changes: e.g. $k^w > p$ and $h^w > f$ are possible changes, while neither $*k > p$, $*x > f$, nor $*p > k$, $*f > x$ are attested prevocalically (recall that the Spanish change is irrelevant).

2.1.2. Phonologically conditioned phenomena

These phenomena, similarly to those in 2.1.1 above, also lack a triggering environment. The phonotactic environment is readily seen, however: these changes occur before a consonant or both before a consonant and a word boundary. As for the changes themselves, here plain labials lose their labiality and become plain velars. These are typical lenition cases. (The Dutch and English changes have been treated at some length in both Huber 2002 and Huber 2004b, while Huber 2002 treated Romanian in a preliminary way).

2.1.2.1. Dutch shows reflexes of a diachronic change where a labial turned into a plain velar in preconsonantal positions. Here are some comparative data that show cognates of Dutch words in English and German (in Dutch < ch > represents /x/):

(11) the rule:	Dutch: /f/ --> /x/ / __C	
the cognates:		
Dutch	English	German
kopen > <i>kocht</i> 'to buy, bought 3Sg'	cheap	kaufen 'to buy'
berucht 'notorious' related to <i>beroe[p]en</i> 'to be called'	—	berufen 'to be called'
gracht <type of channel> <D <i>gra[v]en</i> 'to dig out'	grave	graben 'to dig'
klucht <type of comedy; farce> related to D <i>kloo[f]</i> 'split, gap'	cleave	klaffen 'to gape'
achter 'behind'	after	
kracht 'power'	craft	Kraft 'power'
lucht 'air'	loft	Luft 'air'
stichting 'fund'	—	Stiftung
zacht 'soft'	soft	sanft

The data above reveal that the change occurred irrespectively of the nature of the preceding vowel, both front and back vowels could appear there. What is also shown by the data is that the change was likely to occur only before a /t/. In fact, van der Wal (1992: 30) gives the rule in the form: ft > cht (= [ft] > [xt]) in Old Dutch, more precisely in Old Hollands, not in other Netherlands varieties. The problem that immediately arises is why the change is restricted to this environment. This will have to be treated elsewhere, though.

2.1.2.2. Northern Russian has a similar phenomenon where a labial turns into a plain velar preconsonantly and word-finally. This change is then different from the Dutch cases above in an important respect: it is not restricted to preconsonantal environments, rather it applies at the end of words, too. (12) illustrates the reflexes of Old Slavonic *w in Standard Ukrainian, Standard Czech, Standard Russian and Northern Russian in word-initial, preconsonantal and word-final positions (data from Cyran-Nilsson (1998: 90), highlighted parts in IPA:

(12)	St Ukr	St Czech	St Russian	Northern Russian	gloss (St Russian)
	vOda	vOda	vAda	vAda	'water'
	ławka	la:fka	łafka	łaxka	'fixed bench'
	sliw	slOf	slOf	slOx	'word'

Without going into the peculiarities of the individual languages, suffice it to say that reflexes of *w show a tendency to strengthen to fricatives in more and more environments. East Ukrainian (not represented above) is the most conservative because it retains /w/ in all original environments. Standard Ukrainian has fricative /v/ word-initially, but /w/ elsewhere, Standard Czech and Standard Russian pattern alike since they have /v/ word-initially and /f/ in the other two environments (they differ in more special environments not cited above). Northern Russian went furthest in that it has /v/ word-initially but turned /f/ to /x/ when word-final or before a consonant. The problem that arises if one compares Dutch and Northern Russian is why Dutch does not have that change word-finally as well? A discussion will follow later.

To conclude this section, all the typical phenomena are phonologically conditioned, more precisely, they are prosodically conditioned (there is no contextual reason at all) since they occur before a consonant or both before a consonant and a word boundary. That these labial-velar changes occur in prosodically defined environments is exactly what has been found to hold for all the cases in 2.1.1 as well.

2.2. Atypical phenomena

These are all cases where a plain velar becomes a labial, which is in stark contrast to an earlier conclusion (in 2.1.1.6 above) that only complex labio-velars can undergo splits to a plain labial. (Recall that a possible case of plain labials turning into plain velars, in Spanish, has been refuted above). What is more, these atypical changes always occur in phonologically weak positions: in pre-consonantal and word-final positions (just like in 2.1.2.2 Northern Russian above). That these are called atypical is due to the initial difficulty in explaining them rather than their actual rarity in languages (cf. Huber 2004b). In fact, they are phonologically absolutely regular, but this time the nature of the preceding vowel does have a role to play here: these changes took place after labial vowels, and only later could they spread further.

2.2.1. Old English plain velars turned into labials before a consonant or at the end of a word in Middle English times. It is true, though, that the change only occurred after back vowels, never after front vowels where the original velar fricative vocalized and came to form diphthongs. Moreover, this change does not apply word-initially since there it regularly gave /h-/ as in house, home, etc. Here are some examples for the change to labials:

(13) the rule: (Middle) English: /x/ --> /f/ / __C/#

some examples with /-f/ (spelt <gh> today) and their Germanic cognates with velars:

clough	Scots <i>cleuch</i> /klu:x/
cough	Du <i>kuchen</i>
enough	G <i>genug</i> ; Du <i>genoeg</i>
laugh	G/Du <i>lachen</i>
rough	Du <i>ruig</i> (cf. G <i>rauh</i>)
trough	G <i>Trog</i> ; Du <i>trog</i>

and some others:

chough, slough (of a snake), tough

also preconsonantly:

laughter, draught (cf. dra[g], draw < drawe < drage; G *tragen*)

These (Middle) English developments are a mirror image of Northern Russian above in the sense that exactly the reverse change happens in exactly the same environment. This change, it has to be repeated, occurs after back vowels only, more precisely after labial vowels. There is only a handful of examples with /a/ and these can be analogical in fact.

2.2.2. In Romanian a plain velar turned into a labial before a consonant, but not at the end of words. Parallel developments are also attested in Dalmatian, an extinct language. That the change could originally be restricted to positions following a back (or more precisely, labial) vowel is indicated by the Dalmatian data: Latin *octu* gave Dalmatian *guapto* 'eighth', *cognatu* gave *commut*. Also, in Albanian, traces of the same development are restricted to positions following a back (labial) vowel: Albanian *lu/ff* *të* < Latin *lu/klta* (Tamás 1976: 67). Romanian, however, seems to have extended the rule as the following data testify (from Tamás 1976):

(14) the rule from Latin to Romanian: /k g ŋ/ > /p b m/ /__C

the data:

drea[pt]ă	'right'		< Latin dire[kt]-
dre[pt]	'straight, direct'		< Latin dire[kt]-
fa[pt]	'fact'		< Latin fa[kt]-
la[pt]e	'milk'		< Latin la[kt]-
lu[pt]ă	'fight'		< Latin lu[kt]a-
noa[pt]e	'night'		< Latin no[kt]-
o[pt]	'eight'		< Latin o[kt]u-
coa[ps]ă	'thigh'		< Latin co[ks]a
cu[mn]at	'male relative'	< [ŋn]	< Latin co[gn]atus
pu[mn]	'fist'	< [ŋn]	< Latin pu[gn]u-
se[mn]	'sign'	< [ŋn]	< Latin si[gn]u-

Notes: (1) the occasional diphthongs <ea, oa> are later regular Romanian developments; (2) the [gn] > [ŋn] is regular too.

The Romanian changes are a mirror image of Dutch above since word-finally no change occurs in either, but Romanian has exactly the reverse change. It might be worth recalling that Romanian also retained original preconsonantal /p/'s (cf. *șapte* < Latin *septem* 'seven'), which is unique among Romance languages, and that Romanian regularly turned labio-velars to plain labials anyway (see (5) above). All in all, there is quite some labial dominance in Romanian.

The intriguing problem in the English and Romanian data is where these labials could possibly get their labiality from. Probably it is not irrelevant that the changes are either still restricted to positions after a back (possibly labial) vowel (in English) or at least they used to be so restricted (in Romanian). A possible account for this phenomenon following King's idea will be presented in the following section.

3. The analyses

As had already been indicated in the Introduction, a major observation in connection with labial-velar interactions is that they cannot be easily attributed to assimilations on the production side. The acoustic similarity, that is, the perception side of the phonological component, however, has been, noted by a number of authors. Probably Ferreiro (1999: 116), writing about the history of Galician, had some similar observations in mind when he commented on this change to a labial as "being utterly natural". Schmidt (1993: 68) similarly notes that labials and velars are acoustically nearly equivalent. (It has to be noted here that labiovelars of the /kp gb/ type have been excluded altogether from the discussion. They will have to be treated elsewhere).

In works of Classical Generative Phonology, say in SPE, the feature [grave] had been introduced to subsume labials and velars as well (see Durand 1990 for an overview). This is a perception feature which is in opposition with [acute] positively specifying coronals. It has to be noted as well that, since every segment had to be specified in that framework, velars were defined in SPE as [-labial] and [-coronal], that is, no independent feature was assumed which could define velars positively. This observation should not be neglected.

Government Phonology, without going into the peculiarities of the framework, sees the various phonological phenomena to be deducible from a strictly limited number of possible interactions between strictly adjacent segments. In fact, the only possible effects are termed *licensing* and *government*: licensing makes the realization of a segment possible while government exerts various effects that reduce the capacity of a segment to appear in a given position and thereby to deprive segments from their inherent properties ("consonants are mute, vowels are loud"; recall the Latin grammatical term *mutae* for stops; cf. Szigetvári 2001: 56). Both these forces apply from right to left (at least in the standard version of the theory, cf. Charette 1992, Harris 1997, Szigetvári 2001). In Government Phonology, the binary features of earlier frameworks are replaced by privative elements (cf. Harris and Lindsey 1995). Labials, in particular, have a place element U which defines their lip-rounded pronunciation. Velars do not have an element of their own, which is the simple translation of the lack of labial and coronal properties expressed as [-labial, -coronal] in the earlier SPE

theory. The lack of an independent element defining velars naturally follows from SPE features and it will be the basis for the following analyses.

As has been established above, the Celtic and other changes from labio-velars to plain labials (e.g. /kw/ > /p/) in (2-10) do not have contextual conditioning, rather they are prosodically conditioned by the prevocalic environment. They can be analysed as a simple case of internal restructuring of a segment (the promotion of U labiality to head position):

$$(15a) \quad \begin{array}{l} k^w \implies p \\ [] \quad [U] \\ \backslash \\ [U] \end{array} \quad \text{versus} \quad (15b) \quad \begin{array}{l} k^w \implies k \\ [] \quad [] \\ \backslash \\ [U] \end{array}$$

At this point in the argumentation it is important to realize that it is not absolutely theoretically necessary that a velar lacks a place specification. Consider the possibility that there is indeed an element in the representation of velars. Either choice is possible for the representation in (15b). In (15b) all that happens is that the labiality of the secondary articulation disappears while nothing happens to the rest of the segment (and its representation). If an element were assumed in velars, it could still happily survive. In (15a), on the other hand, it does matter whether place specification is assumed in velars or not because in this case it has to be explained how the actual switch from velarity to labiality comes about since, as has been stated above, there is no phonological conditioning in the environment. In other words, there is no source for the labiality. Notice that the supposed velarity element has to be delinked (to use a well-known term) in the first step and the labiality element must be then promoted to the position it occupies in /p/. There is no theoretical motivation whatsoever for the delinking of the supposed velar element.

Notice at the same time that neither the promotion of the labiality to head position nor its deletion from the secondary position needs any special theoretical machinery: both phenomena are driven by the prosodic environment itself, namely the prevocalic position—a position which is phonologically strong. Here licensing makes segments stable, ‘licensed’ (see Harris 1997, Szigetvári 2000). Also recall that in this position only complex labio-velars could be shown to change; the single case, in Spanish, of labials turning to velars in this position does not ultimately figure here, and no data were found for a theoretically possible change of velars turning to labials before a vowel. To sum up then, there is no motivation for assuming an element in velars since it would not be used to account for any phonological phenomena (recall Occam’s Razor), while the fate of the labiality element—which has to be assumed in labials on independent grounds—in secondary position is readily accounted for by the prosodic environment: the prevocalic position.

The Northern Russian change in (12), which occurs preconsonantly and word-finally, is analysed as a case of phonologically conditioned lenition. Here the effect of government is seen to make consonants more like vowels (recall that government destroys the inherent properties of a segment). The labial place element is lost while all other elements like voicing and continuancy are unaffected:

(16) p ==> k similarly: f ==> x
 [U] [] [U] []

Notice that in earlier featural terms, there was no problem on the formal side of the explanation since a labial segment became a non-labial one. However, an explanation should be found for why, on losing its labiality, the segment in this position gains a velar place of articulation exactly. (If, however, a feature defining velars were assumed, then there must be a reason for that to surface in this context.) Apparently, neither the velarity nor the cause for the [-labial] specification is encoded in the environment. In the government approach, on the other hand, the explanation is straightforward and no appeal has to be made to the segmental environment: the labiality element U is either simply deleted from the representation through government or made impossible by not being licensed. In any case, the segmental environment has no role to play.

To conclude, so far the theoretically more straightforward cases have been accounted for. Two pervasive patterns have been identified. It has been found above that in prevocalic position the strengthening of a labio-velar either to a plain labial or a plain velar results in a more prominent, consonant-like consonant. This is the effect of licensing. In preconsonantal and word-final positions, the loss of the labial element resulted in lenitions of (plain) labials to plain velars. This then is the effect of government. What is important is that the two sets complement each other.

There remain more difficult cases, like the Dutch reduction of labials before /t/ on the one hand, and the Middle English and Romanian changes on the other. In these latter cases, there is indeed a contextual reason for the acquisition of labiality, at least in the original setting. Later changes, however, could result in the extension of this initial pattern to more environments. The basic idea (taken from King 1969) for their treatment is that they can be analysed in terms of well-formedness (phonotactic) constraints banning certain velars in certain environments, these constraints beginning to apply in consecutively more and more environments.

In connection with the Romanian change, Robert D King offers a plausible analysis (1969: 115) in terms of rule addition. (It has to be noted that King aimed at an SPE-type analysis of historical changes). He argues that the change from velar to labial (see data in (14) above) is surprising only if one views this as a change converting a velar segment into a labial one. He proposes instead that the actual change is in the rule component: the addition of a restriction on well-formed structures. While earlier in the history of Romanian there used to be no restriction on a sequence of a non-coronal (labial and velar) and coronal segments, now a rule was introduced of the following form:

(17) Rober D King's analysis (1969):

$$[-\text{continuant}] > [+ \text{anterior}] / \text{_____} \begin{matrix} [-\text{continuant}] \\ [+ \text{coronal}] \end{matrix}$$

What this rule does is to add a restriction to the system to the effect that before a coronal non-continuant only a [+anterior] segment is allowed. This rule crucially does not say that a velar becomes something else, but that before a non-continuant

coronal there can only appear a [+anterior] sound. Obviously, the rule applies vacuously to labials as well. It should be borne in mind that King also expresses the view that the labial in fact does not come from the velar. What is particularly attractive in King's analysis is that it can be extended to the (Middle) English and Dutch data as well: a rule can be added to exclude certain sequences. Although this solution is rather attractive, some questions remain. For instance, it is not immediately clear why the [+anterior] happens to be a labial, since an Italian-style solution with coronal gemination would also meet this restriction (cf. Italian *notte*, *fatto*, for instance). Of course, one could add a rule prohibiting geminate consonants in Romanian, so this is not a serious problem. It is more problematic, though, that the reasoning is difficult to test since there is no other set of segments other than the velars that would show the effect of this rule addition. Notice that the same objection cannot be raised against the Dutch restriction since there even stop /k/ turns into /x/ regularly before another non-continuant: *zoek-* [-k] - *zocht-* [-xt] 'look for <pres>, <past>'. Nevertheless, this approach is important since King excludes the alternative route along some /k/ > /k^w/ > /p/ trajectory, which is totally unsupported by the data anyway and it is absolutely unnecessary once one rejects the claim that it is sounds rather than grammar that changes. As a final remark on this analysis, this change is a nice symmetrical twin of what was observed in Dutch: in Dutch a restriction was introduced to exclude labials, in English and Romanian it is velars that are excluded by a structurally identical constraint.

King thus treats this particular change as a change in the rule component rather than an extension of a minor regularity to more and more environments. This latter possibility cannot be excluded, however, at least in a number of cases. Although it was mentioned in the preceding paragraph that this change is attested not only in words that have a neighbouring labial vowel, it can still be the case that indeed those were the first instances of the change, and later the rule extended its scope to all back vowels. In fact, the Dalmatian and Albanian data cited in 2.2.2 do show such a scenario. (This is a possible chronology for the English *-gh* words, too).

There is a final point to be considered. One important aspect of the various velar-labial interactions has been neglected so far: what impact all these changes have on the phoneme inventories of the respective languages. An important observation in connection with the Romanian, Dutch and English changes is that these phonotactic rule additions do not affect the phoneme inventory of the language, they only change some distributions in it. These rules do not delete a phoneme from the inventory or add a new phoneme to the system. In Romanian, /k/ can and does appear word-initially or intervocalically in words of Latin origin; similarly in English word-initial /x/ did not disappear but it gave /h/ as in *house*; also in Dutch, word-final /f/ is free to occur. Only, they are banned in some environments. However, in languages where labio-velars were affected before a vowel (Celtic and Greek), the (original) labio-velars did not survive, the inventory lost these phonemes altogether.

It is more than tempting to collapse this observation on phoneme inventories with the changes in the various prosodic environments. This gives a better and truly phonological typology of the velar-labial interactions across languages. In prevocalic position labio-velars undergo changes to plain labials (or velars) and this reduces the phoneme inventory. In preconsonantal and word-final positions, reductions of labials

to velars is only prosodically conditioned, the quality of preceding vowels is absolutely immaterial for the changes. On the other hand, changes of velars to labials only happen if there is a preceding conditioning labial vowel as well. In this environment, thus, the unmarked process is from labial to velar, the reverse process needs the conditioning of preceding labial vowels. Any of these changes in preconsonantal and word-final positions leaves the inventory intact.

These observations amount to saying that all labial-velar interactions are exclusively prosodically conditioned, the segmental environment only has a role in the marked process of velars turning to labials, and only the preceding labial vocalic environment matters even in those cases. The revised typology looks like this then:

- (18) The revised typology:
- A Phoneme inventory affected > changes in __V (2.1.1)
 - 1 Changes from labio-velars to plain labials and velars (2.1.2)
 - B Phoneme inventory not affected > changes in __C/#
 - 2 Reductions of labials to velars is only prosodically conditioned
 - 3 Velars turn into labials only when there is labial vowel preceding (2.2)

4. Conclusions

There are a number of important conclusions reached in the preceding discussion on the various interactions between labials and velars. Firstly, although there is no contextual phonological motivation for phenomena where a labio-velar turns into a plain labial (or velar), all such changes are phonologically conditioned since they occur prevocally. Plain labials show reductions to velars, and plain velars turn, under strict conditions, into labials in preconsonantal and word-final positions. In consequence, all velar-labial interactions are (prosodic) phonologically conditioned. Secondly, considerable support was found for the view that only complex labio-velars can turn into plain labials (or plain velars, of course). Plain velars and plain labials cannot undergo any comparable changes to labio-velars. Thirdly, there is no motivation for assuming an element in velars since it would not be used to account for any phonological phenomena. Fourthly, changes of velars to labials can only occur if there is a preceding labial vowel. Later in the history of a particular language, this environment can extend to cover more and more contexts. This step can be best captured by phonotactic rules. And finally, while these latter phonotactic rules do not affect the phoneme inventory of the language, only some distributions in it, changes of labio-velars typically reduce the inventory. These observations can be united in the revised typology in (18).

References:

- Ball, M. J. and J. Fife (eds.), 1993, *The Celtic Languages*. London: Routledge.
- Charette, M., 1992, «Mongolian and Polish meet Government Licensing», *SOAS Working Papers in Linguistics and Phonetics* 2:275-292. London.
- Cyran, E., 1997, *Resonance Elements in Phonology - A Study in Munster Irish*. PASE Studies and Monographs. Vol. 3. Lublin.

- Cyran, E. (ed.), 1998, *Structure and Interpretation - Studies in Phonology*. PASE Studies and Monographs. Vol. 4. Lublin.
- and M. Nilsson, 1998, «The Slavic [w → v] shift: a case for phonological strength». In: Cyran 1998: 89-100.
- Duanmu, S., 2002, *The Phonology of Standard Chinese*. Oxford: Oxford U. P.
- Durand, J., 1990, *Generative and Non-Linear Phonology*. Harlow: Longman.
- Eska, J. F. and D. E. Evans, 1993, *Continental Celtic*. In: Ball and Fife (eds.), pp. 26-63.
- Ferreiro, M., 1999 (4th ed.), *Gramática histórica galega. I. Fonética e Morfosintaxe*. Santiago de Compostela: Edicións Laiovento.
- Fodor, I. (ed.), 1999 (2000), *A világ nyelvei* [Languages of the world]. Budapest: Akadémiai Kiadó.
- Harris, J., 1997, «Licensing Inheritance: an integrated theory of neutralization», *Phonology* 14: 315-370. Cambridge: CUP.
- and G. Lindsey, 1995, «The elements of phonological representations». In: Durand and Katamba 1995: 34-79.
- Huber, D., 2002, *Velar phenomena: processes and their expression in phonology*. MA Thesis. Budapest: ELTE, ms.
- 2003a, «Velars and empty-headedness in Government Phonology», *Rivista di Grammatica Generativa* 28, 45-56. U. P.
- 2003b, «On some velar and palatal interactions». Presented at the Budapest-Vienna Phonology Roundtable 3, 5-6 April, 2003, Budapest.
- 2004a, «On the theoretical status of epenthesis». Presented at the 12th Manchester Phonology Meeting, 20-22 May, 2004, Manchester.
- 2004b, «Velar processes and their representation in phonology». In: *The Odd Yearbook 7* (2003), pp. 11-44. Budapest: ELTE.
- King, R. D., 1969, *Historical Linguistics and Generative Grammar*. New Jersey: Prentice-Hall.
- Lapesa, R., 1981 (9th ed.), *Historia de la lengua española*. Madrid: Editorial Gredos.
- Paradis, C. and J.-F. Prunet (eds.), 1991, *The Special Status of Coronals: external and internal evidence*. San Diego: Academic Press.
- Rix, H., 1976, *Historische Grammatik des Griechischen. Laut- und Formenlehre*. Darmstadt.
- Schmidt, K. H., 1993, «Insular Celtic: P and Q Celtic». In: Ball and Fife (eds.) pp. 64-98.
- Siptár, P. (ed.), 2001, *Szabálytalan fonológia* [Phonology without Rules]. Budapest: Tinta Könyvkiadó.
- Szigetvári, P., 2000, «Why CVCV». In: *The Even Yearbook 4*: 117-152. Budapest: ELTE.
- , 2001, «Szótagtalan fonológia [Phonology without syllables]». In: Siptár 2001: 37-76.
- Tamás, L., 1978 (3rd ed.), *Bevezetés az összehasonlító neolatin nyelvtudományba* [Introduction to Comparative Neo-Latin Linguistics]. Budapest: Tankönyvkiadó.
- van der Wal, M., 1992, *Geschiedenis van het Nederlands*. Utrecht: Het Spectrum B. V.

THE STRUCTURE OF PAIR-LIST ANSWERS

Aritz Irurtzun

EHU-U. Basque Country & HiTT

1. Introduction

This paper focuses on the structure of pair-list answers. These are the typical answers of multiple-Wh questions and questions with quantifiers like the ones in 1 and 2 respectively:

- (1) a. Who kissed whom?
b. [John] kissed [Mary]...
- (2) a. Who kissed everybody?
b. [John] kissed [Mary]...

Abstracting away from the patterns of the answers to questions with quantifiers, the goal of this paper is to analyze the following questions: What is the nature and discourse function of the elements in brackets in sentences like 1b? What is the grammatical encoding of the information-packaging of these constructions?

To start, compare the sentences in 3b and 4b, and the questions they answer (3a & 4a respectively):

- (3) a. Who bought beer?
b. [John] bought beer.
- (4) a. Who bought what?
b. [John] bought [beer]...

In the question-answer pair in 3, the question asks about the agent of the event of buying beer and the only element that is not given in the question that appears in the answer is the subject 'John', what is traditionally analyzed as being the focus of 3b (*cf.* Rooth (1985), Herburger (2000) and Krifka (2001) among many others). In 4, on the other hand, we have a multiple-Wh question in 4a and in its partial answer, two elements that are not expressed in the question; the subject 'John' and the object 'beer'. The question, as said, is what the nature and discourse function of these elements is.

In one of the most widely accepted analysis of the semantics of questions a question is taken to denote a set of propositions (*cf.* Hamblin 1973). For instance, the denotation of the question in 5a would be the set of propositions in 5b, where the Wh-phrase in the question has been replaced by different alternative values that are available in the context. Thus, an appropriate answer to the question in 5a will be one of the propositions in this set, for instance 5c:

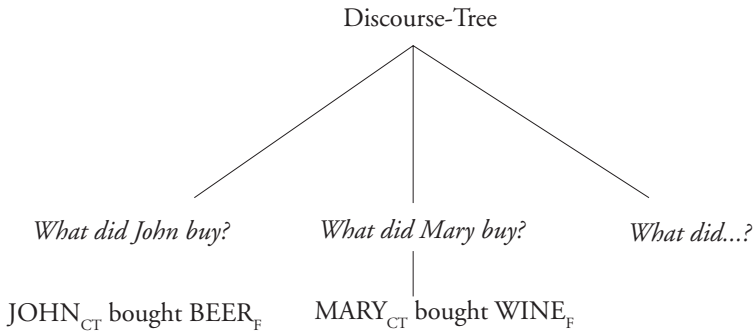
- (5) a. Who got the flu?
b. [[Who got the flu]]={[[Kepa got the flu]], [[Eider got the flu]], [[Adam got the flu]], [[Ibon got the flu]], ...}
c. Kepa got the flu.

According to this approach, then, a multiple Wh-question like 6a denotes a set of questions, that is, a set of sets of propositions like 6b. This question could be answered by the sentence in 6c:

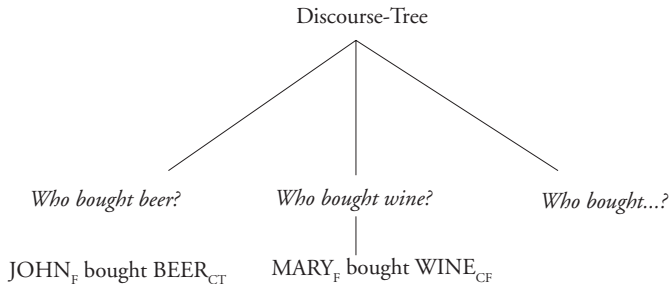
- (6) a. Who cooked what?
 b. $[[\text{Who cooked what}]] = \{ \{ [[\text{Adam cooked cod}]], [[\text{Adam cooked rice}]], [[\text{Adam cooked eggplants}]] \dots \}, \{ [[\text{Julen cooked rice}]], [[\text{Julen cooked pasta}]], [[\text{Julen cooked tuna}]] \dots \} \dots \}$
 c. Adam cooked eggplants and Julen cooked pasta.

This type of semantics approach to questions is adopted by Büring (2003) in his analysis of discourse structuration and answerhood, proposing that in an answer to a multiple-Wh question we have different possible answer strategies like those represented in the discourse trees (or D-Trees) in 7 and 8 (in this case, the choice of strategy would imply whether to start answering by the agents of the event of cooking or by its themes).

- (7) a. *Who bought what?*
 b. $[[\text{Who bought what?}]]^{\text{ct}} = \{ \{ x \text{ bought } y \mid y \in D_e \} \mid x \in D_e \}$



- (8) a. *Who bought what?*
 b. $[[\text{Who bought what?}]]^{\text{ct}} = \{ \{ x \text{ bought } y \mid x \in D_e \} \mid y \in D_e \}$



Thus, when answering a complex question like 7 or 8, a speaker can opt between whether to answer by 'buyers' or by 'buyees' and this, according to Büring, will determine the information-packaging nature of the elements not given in the question. Büring (2003) thus requires two independent discourse-configurational primitives: the 'contrastive topic' that would indicate the answer strategy to follow, and the 'focus'. Crucially, both information-packaging elements are analysed as having the very same semantic import: that is, rising alternative values *à la* Rooth (1985). Recall that according to Rooth's 'Alternative Semantics' approach, a sentence with focus would have two denotations: the 'Ordinary Semantic Value', that will be the proposition obtained compositionally by Montagovian function application (this proposition won't be affected by the focus), and the 'Focus Semantic Value', a set of propositions obtained by the substitution of the focused phrase with alternatives available in the discourse that match the focus in semantic type (*i.e.*, roughly, the semantic value of the question it answers in a Hamblin-type semantics of questions).

My concern here is that despite the representational interpretation in Büring (2003) captures in an elegant way the denotation of these constructions, the 'topicness' of the 'contrastive topics' proposed is not very well established; after all, both the 'focus' and the 'contrastive topic' are analyzed as having the very same semantic import. Furthermore, as Büring himself notes (Büring (2003: 512)), the so-called 'contrastive topic' doesn't behave in some relevant respects like other topics; for instance, its presence is mandatory and not optional (hence, they cannot be elided), and they answer (in part) the question instead of stating necessarily old/given information. Thus, I would want to suggest that we don't need the theoretical primitive of 'contrastive topic' in order to capture the semantics of these sentences. Therefore, the proposal to be developed in this paper is that in these constructions we have a *pair* of elements as the focus. For instance, in the case of the discourses of 7 and 8, the focal elements can be regarded as taking part in a relation denoted by the verb; the first element that stands for a Wh-word of the question sets the domain and the second one sets the range of the relation. Even more, as will be argued, with the adoption of the derivational analysis of the focus construction presented in section 2, the pairing semantics of these constructions will be derivative of their focal status in a straightforward way.

In a nutshell, then, in this paper I will be arguing that the semantic representation proposed by Büring (2003) is basically correct, but that we can dispense with the theoretical primitive of 'Contrastive Topic' for these constructions. Furthermore, I will argue that these elements should be better reanalyzed as being focal in nature. In order to do that, I will present in section 2 the derivational approach to the focus structure proposed in Irurtzun (2003b) and the neodavidsonian semantic representation for focus of Herburger (2000) as the theoretical framework in which I will base my analysis. Then, in section 3, I will present the derivation of split focus constructions and review some of the intonational, semantic and morphosyntactic properties of these sentences in different languages. I will argue that the behavior they display is to be expected, assuming the theory presented in this paper. A brief summarizing and concluding section follows.

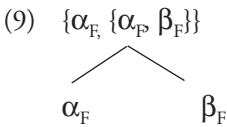
2. Focus structure and interpretation

In this section I will present the theoretical framework in which I will base my analysis: in 2.1. I present the derivational approach to the focus structure of Irurtzun

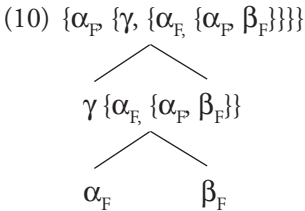
(2003b), and in 2.2. the Neodavidsonian semantics for focus of Herburger (2000). The conjunction of these two theories will set the basis of my analysis of the answers to multiple-Wh questions of section 3.

2.1. A derivational approach to the focus structure

According to the minimalist theory of focus structure construction proposed in Irurtzun (2003b), the [+F] feature is an optional formal feature and it is potentially assigned to several tokens of the numeration. Hence, the focus structure, instead of being ‘projected’ at PF from the element that got the nuclear stress, it is constructed derivationally by means of Merge in the narrow syntax, and nuclear stress is just assigned to it in PF. That is, technically the focus structure is built up as follows: when an element α and an element β undergo Merge both of them bearing the [+F] feature, a new syntactic object will be created that in “Bare Phrase Structure” terms (*cf.* Chomsky 1995a), will be a set-theoretic object containing only [+F] featured lexical items:



In that way, when a syntactic object/set of [+F] featured lexical items is merged with an element that does not itself bear the [+F] feature, the new syntactic/set-theoretic object will not be a set containing only [+F] featured lexical items, as the highest phrase in 10 shows:



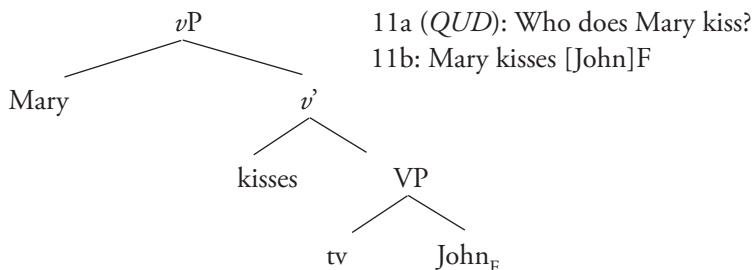
Although the head (and even the label) of the structure in 10 is marked as [+F], the whole structure won't be a set containing only [+F] featured lexical items, since the element γ (a member of $\{\gamma, \{\alpha_F, \{\alpha_P, \beta_F\}\}\}$) does not bear the [+F] feature itself. Thus, precisely because of the lack of the [+F] feature of γ , in this structure we will have just $\{\alpha_F, \{\alpha_P, \beta_F\}\}$ marked as focal. Assuming such a derivational construal, we keep a direct mapping between syntax and semantics and build semantic interpretation in a strict compositional way. Furthermore, with this derivational analysis, we observe one of the core minimalist assumptions; the ‘Inclusiveness Condition’ (*cf.* Chomsky 1995b: 228):

Any structure formed by the computation (in particular, π and λ) is constituted of elements already present in the lexical items selected for N; no new objects are added in the course of computation apart from rearrangements of lexical properties...

In order to show how the system works, let us say that we have the simplified numeration in 11, and that the *Question Under Discussion* is the one in 11a. When the

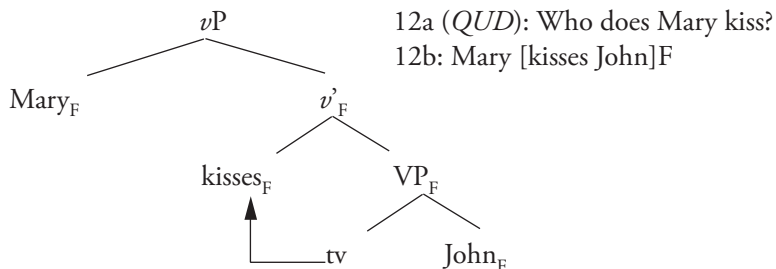
[+F] object (derived as in 10) is merged with the [+F] featureless verb, the new syntactic object (VP) won't be a set containing only [+F] featured lexical items. This will be so because the verb doesn't bear itself the [+F] feature. Such a configuration would end up in a sentence like 11b with $[John]_F$ as the only focal element:

(11): Lexical Array: $\{\{Mary\}, \{John_F\}, \{kiss\}, \{v\}\}$



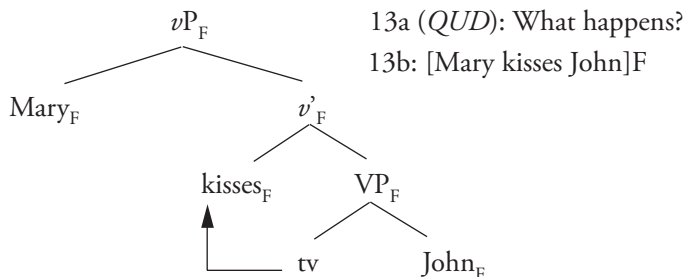
Right in the same way, if we have the numeration in 12, when the object bearing a [+F] feature is merged with the verb that itself bears the [+F] feature, the new object created (v') will be a set containing only [+F] featured lexical items, as in the sentence in 12b:

(12): Lexical Array: $\{\{Mary\}, \{John_F\}, \{kiss_F\}, \{v_F\}\}$



Instead, if we have the numeration under 13, when the object and the verb are merged, a new syntactic/set theoretic object is created made out of only elements that bear the [+F] feature. Once this object is merged with the light verb, and the new element is merged with the DP subject that itself bears the [+F] feature, we end up with a derivation that is a set containing only [+F] featured lexical items; that is an out-of-the-blue sentence (13b):

(13): Lexical Array: $\{\{Mary_F\}, \{John_F\}, \{kiss_F\}, \{v_F\}\}$



Therefore, recall that according to this proposal, for an element to bear the [+F] feature does not mean that it will be the *actual* focus of the sentence but just that it will take part in the composition of the focus structure, which will be composed out of all the [+F] marked material.

The system has some welcome predictions, among them, that it allows for the interface components to access the actual focus structure, since it is already set in the narrow syntax. Thus, for instance, the PF component will be sensitive to the already built F-Structure. As a brief example, many of the technical problems of a *Nuclear Stress Rule*-based theory of focus structure (*cf. i.a.* Cinque 1993, Neeleman & Reinhart 1998) are avoided if we allow the cinquean *Nuclear Stress Rule* (henceforth *NSR*) that assigns nuclear stress to the element with most grid marks (the most deeply embedded one) to apply just within the focus structure that we built up derivationally in narrow syntax. The definition of such a rule is in 14:

(14) *Nuclear Stress Rule*: Assign Nuclear Stress to the element with most grid marks within the focal structure.

This new *NSR*, will predict correctly and without any further stipulation the Nuclear Stress placement in different positions, given that different focus structures derive from different numerations (*cf. Irurtzun 2003b* for further discussion):

- (15) a. John boiled [WATER]_F d. [JOHN]_F boiled water
 b. John [boiled WATER]_F e. John [BOILED]_F water
 c. [John boiled WATER]_F

Recall, furthermore, that having severed the setting of the F-Structure from the nuclear stress placement weakens immediately the problematic nature of the so-called Schmerling' examples. These are marked cases of sentence focus with nuclear stress on the subject, like in 16c, a possible answer to the *out-of-the-blue* question in 16a, given an appropriate exclamative context:

- (16) a. What happened? b. [Truman DIED!]_F c. [JOHNSON died!]_F

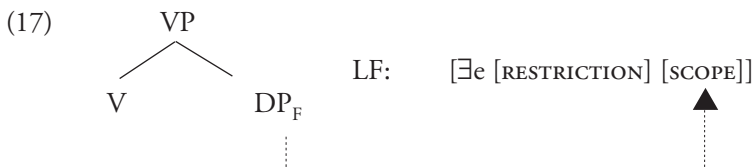
According to the literature, the most neutral type of answer to a question like 16a would be 16b, with nuclear stress on the verb. This is captured immediately by a *NSR*-based theory of the focus structure, since the verb is in a more embedded position than the subject, hence, it gets more metrical grid marks, and hence, it gets the nuclear stress. Thus, the embeddedness of the verb allows it to project its focal status higher up in the structure. However, in a context where it is a surprise that Johnson died, 16c is a natural *out-of-the-blue* sentence. And this is highly problematic for *NSR*-based approaches to the F-Structure, since according to these theories the F-Structure is set *via* the projection of the focal status of the item that gets the nuclear stress. Thus, in the case of 16c, it should be impossible for the nuclear stress on the subject to denote sentence-focus; an economy principle should ban it since nuclear stress on the verb (the option by default) provides that possibility (*cf. i.a.* Cinque 1993, Reinhart 2006 for discussion). Note that on the other hand, the focal status of these sentences is unproblematic for the approach defended here, since it is set inde-

pendently of the nuclear stress placement. Nuclear stress will be just a way to interpret in PF the focus structure. Thus, the marked stress placement could be explained as a marked stress shift from its assignment position (as said, the verb) due to the fact that the construction is an exclamation about that specific subject's death. Furthermore, if this is a matter of a PF stress shift (and hence, a local operation), we can understand the impossibility of having more material between the subject and the verb, as in 16c:

(16c) *[JOHNSON suddenly died!]_F

2.2. Focus semantics: a Neodavidsonian approach

On the other hand, in order to provide a semantic representation for focus constructions at *logical form*, I will adopt the proposal of Herburger (2000). Herburger frames her analysis within the Neodavidsonian tradition and proposes that, taking sentences to be descriptions of events, at *logical form* the focal material is mapped into the scope of a restricted existential quantification over events. As in 17:



For instance, the sentence in 18a as an answer to the question in 18b will have the Logical Form in 18c, where the non-focused chunk is the restrictor of the existential quantification (*i.e.*, the sentence's 'aboutness') and the focus is in the scope (*cf.* Herburger 2000):

- (18) a. Mary bought [BEER]_F b. What did Mary buy?
 c. $\exists e$ [Agent(e, mary) & Buy(e) & Past(e)] Theme(e, beer) & Agent(e, mary) & Buy(e) & Past(e)]

As said, the restriction will give the sentence's 'aboutness' information whereas the nuclear scope will give the focus (*cf.* von Heusinger (1999) for a similar analysis in *Discourse Representation Theory* terms). Thus, as argued earlier, marking an element as [+F] in the numeration doesn't mean that it will be the actual focus of the sentence but rather that it will take part in the syntactic derivation of the focus structure in narrow syntax, and that it will take part in the focus interpretation at *logical form*.

As presented in this section, the derivational analysis of focus structure construction proposed in Irurtzun (2003b) provides a narrow syntax setting of the actual focus structure and allows for its interpretation in both interface levels. At PF we just have to modify the mainstream *Nuclear Stress Rule* to make it focus-sensitive and we get immediately the correct nuclear stress placement in every focal structure. At *logical form*, and following Herburger (2000), I will assume that all the focal material is mapped into the scope of an existential quantification over events and that

the focus interpretation is obtained by the computation of all the [+F] featured material.

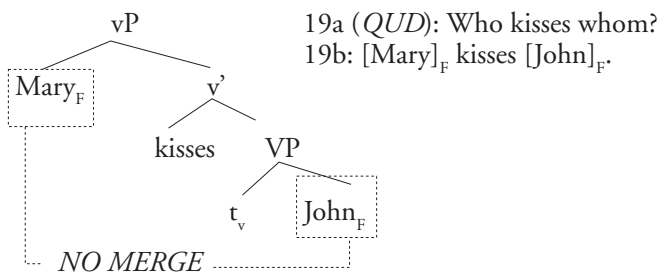
However, one of the predictions of such an approach is that, in principle, nothing should prevent the appearance of a split focus structure obtained by the assignment of [+F] features to lexical items that don't merge together. I will exploit this possibility in the next section arguing that such configurations give rise to the patterns of answers to multiple-Wh constructions.

3. Split Focus Structures

In this section I will analyze one of the possibilities that arise with the adoption of the derivational construal of the focus structure just proposed: the possibility of having syntactically split focus structures. Then, I will discuss some of the intonational, semantic and syntactic properties of these constructions and argue that in these instances of split foci, we have pairing answers to multiple-Wh questions like those represented with D-Trees in section 1.

As just presented in section 2, I am assuming that the focal structure is built up in the narrow syntax with the dynamics of the derivation: when two focal elements are merged together the new syntactic object created will also be focal. However, such a theory has an interesting prediction: whenever two elements enter the derivation bearing each of them a [+F] feature but they don't merge together, two isolated focus structures will arise. For instance, a common case could be when a DP subject and a DP object enter the derivation being [+F] marked but the verb doesn't bear it; something like 19:

(19): Lexical Array: {{Mary}_F} {John}_F {kiss}, {v}}



Thus, and following the type of semantic representation proposed by Herburger (2000), at the level of *logical form*, all the [+F] material will be mapped into the scope of a restricted quantification over events:

However, notwithstanding the idiosyncrasies of focus-marking tunes in different languages, there is some regularity in the tunes for ‘contrastive topics’ across languages: right as with the ‘B accents’ of English, in other languages like Basque or Serbo-Croatian the so-called ‘contrastive topics’ are characterized by a final pitch rise. For Central Basque, I have analyzed elsewhere these constructions as involving a tune composed by a H* pitch accent and a H- boundary tone (*cf.* Irurtzun 2003a). However, in this respect, the most interesting language that I am aware of is Serbo-Croatian as analyzed in Godjevac (2000). In this language, in an answer to a multiple-Wh question each of the elements bears a L*+H pitch accent; and, akin to English or Basque, the so-called ‘contrastive topic’ phrase ends in a H- phrase accent and the ‘focus’ in a L-. However, there is one additional tonal event involved in these constructions: an initial %H in the ‘focus’. This is shown in 22, as answering a question like ‘Who gave a lemon to whom?’:

- (22)
- | | | | | | | |
|-------------------------------|------|----|-------|------|----|-------|
| %L | L*+H | H- | %H | L*+H | L- | |
| | | | | | | |
| JE | LE | NA | je MA | RI | JI | dala. |
| ‘[JELENA] gave it [to MARY].’ | | | | | | |

Recall, that the %H boundary tone of 22 is not derived by the adjacent position of the H- phrase accent of ‘Jelena’, since, looking at 23 (where this adjacency does not hold), it seems that it is a categorical property of these constructions (since in normal/single focus utterances there is no %H at the left edge of the focus phrase):

- (23)
- | | | | | | | |
|---|------|----|---------------|----|------|-----|
| %L | L*+H | H- | | %H | L*+H | L- |
| | | | | | | |
| JE | LE | NA | je dala ravan | MA | RI | JI. |
| ‘[JELENA] gave the flat one to [to MARY]’ | | | | | | |

In my view this evidence shows that on the one hand, in answers to multiple-Wh questions both elements that stand for a Wh-phrase bear a pitch accent. On the other hand, that the tune differences between both elements are usually phrasal, and there is a striking regularity across languages in that the tunes associated to ‘contrastive topics’ end in a high tone. Furthermore, as observed in Serbo-Croatian, the so-called ‘foci’ of the answers to multiple-Wh questions are not the same elements as foci that answer single-Wh questions.

Thus, and following the ‘isolated focus-constructions’ proposal of 3.1, I would want to suggest that in these constructions we don’t have a ‘contrastive topic’ and a ‘focus’ (as proposed by Büring 2003), nor two independent foci (as answers to *Conjoined Questions*, see below) since the intonational patterns associated to them are not the same as those in sentences with a single focus. The conclusion would be that in these constructions, what we have is a single focus that is the *pair* of both elements, and the common high phrase accents could be analyzed as grammaticalized ‘continuation rise’ contours, something that would not be surprising under the analysis defended in this paper, whereby the focus structure is split among both elements bearing the [+F] features.

3.2. Semantic properties

As is widely acknowledged (*cf.* among others Bošković 2002, Büring 2003), in languages like English (24) or Basque (24) that show overt movement of (one of) the Wh words, sentences like 24b and 25b are *partial* answers of multiple Wh questions like 24a and 25a respectively:

- | | |
|--|--|
| (24) a. Who broke what? | b. John broke the door... (pair list) |
| (25) a. Zeinek erosi du zer?
which buy AUX what | b. Jonek atea hautsi du... (pair list)
Jon door break AUX |
| Who bought what? | John broke the door |

In fact, in English, a question like 24 in a scenario that demands a single-pair answer is incongruent. Scholars like Wachowicz (1974, 1975) or Bolinger (1978) make a distinction between two types of multiple-Wh questions: *Matching Questions* and *Conjoined Questions*. Matching questions are the real multiple-Wh questions, those questions like 21 that demand a *pair list* answer. The nature of this need for a multiple event is a mysterious and remarkable fact (*cf.* Bošković 2002 for a possible analysis). Thus, 24 which allows or rather demands a pair list answer is a good example of matching questions but examples like 26 and 27 (below) are not. Arguably, this is due to the impossibility of having several events of killing Robert Kennedy (26) or keeping one single dollar at the same time in various banks (27):

- (26) *Who killed Robert Kennedy when?
(27) *Who is keeping the silver dollar in which bank?

Comparing 26 and 27 to similar examples that allow the multiplicity of events like those in 28 (for 26) and 29 (for 27) make clear that the oddity of these sentences is strictly related to the necessity of having one single event and a pair-list answer:

- (28) a. Who saw Robert Kennedy when?
b. Who killed which Kennedy?
(29) Who kept the silver dollar in which bank?

On the other hand, we would have the conjoined questions; a conjunction of questions that demands for the independent identification of two variables. A case like these could be the one in 30, a grammatical variant of 26:

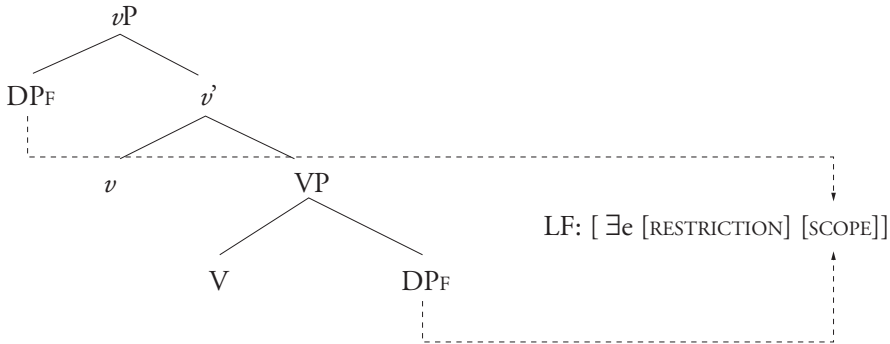
- (30) Who killed Robert Kennedy, and when did he do it?

Here the question is perfectly natural because it demands independently and in two single-Wh questions for two pieces of information. Thus, putting aside the conjoined questions, what I want to argue is that the pairing pattern of multiple-Wh questions is explained straightforwardly with the analysis of the derivation and interpretation of the focus structure presented in section 2. Following a line of thought developed in Chomsky (1973), Higginbotham & May (1981) and Gutiérrez-Rexach (1999) among others, I will assume that at LF, in a multiple-Wh question like 31a, an operator absorption takes place creating a compound polyadic operator that quantifies over pairs of variables. This is represented in 31b:

- (31) a. Who ate what?
 b. [WH x , WHY: person(x) & eatable thing(y)] x ate y

This LF representation for multiple-Wh questions is what will give us the *bijection* interpretation. Thus, the most natural assumption about the answers that these questions demand is to take both elements that stand for the pairs of variables in the question to be focal. The *uniqueness* of focus, the fact that each sentence has just one focus will be trivially obtained given the *logical form* representation assumed in section 2, whereby all the [+F] material will fall in the scope of an existential quantifier over events. In these cases the focus will be split, it will be the pair of elements being marked [+F], as they are mapped into the scope of the existential quantification over events. Basically, as depicted in 20, repeated here as 32 for convenience:

(32)



For instance, for the sentence in 33a (as a partial answer to 31a), we would have the *logical form* in 33b:

- (33) a. John ate pizza.
 b. \exists [Eating(e) & Past(e)] [Eating(e) & Past(e) & Agent(e , John) & Theme(e , pizza)]

The corollary of such a proposal is that there will be just one focus per sentence, even if it has the form of a pair. Instead of introducing this as a principle, this fact will be derivative of the nature of the *logical form* representation of sentences with focus, *i.e.* that the quantification over events just has one scope. I think this is a nice prediction, and one of the advantages of this proposal comparing to previous approaches.

3.3. Some morphosyntactic properties: the ‘contrast’ particles of Japanese and Korean

Finally, with the analysis just sketched, we can also account for the usage/lack of usage of *contrast* particles of Wh-in-situ languages like Japanese or Korean, where multiple-Wh questions can be answered with either a single-pair or pair-list answer

(*cf.* Hagstrom 1998, Bošković 2002). Bošković (2002) gives the following scenario for triggering single-pair answers: *John is in a store and in the distance sees somebody buying a piece of clothing, but does not see who it is and does not see what the person is buying.* With this scenario, in a ‘Wh-moving language’ like English, a question like 34 is incoherent (since, as said earlier, it is inherently a matching question) whereas its counterpart in a ‘Wh-in-situ language’ like Japanese in 35 is fine:

(34) Who bought what?

(35) Dare-ga nani-o katta no?
 who-nom what-acc bought Q
 ‘Who bought what?’

Whichever the explanation for the lack of single-pair reading in Wh-movement languages, the case is that this reading is available in Wh-in-situ languages. The striking fact here is that in this type of languages, an answer to a multiple-Wh question is different when it is a single-pair or a pair-list answer (an asymmetry that up to my knowledge wasn’t attested in the previous literature on the topic). In languages like Japanese or Korean that allow for the single-pair reading, the usage of some particles (‘-wa’ for Japanese, ‘-nun’ for Korean) varies with the type of answer; the appearance of those particles is mandatory in the first element when asked for a pair-list answer but, remarkably, in both languages, when the question demands a single pair, the answer cannot bear such a particle (*cf.* 36a-b for Japanese and 37a-b for Korean):

(36) a. Takako-wa wain-o kaimashita... (pair list)

Takako-WA wine-ACC bought
 ‘Takako bought wine...’

b. Takako-ga wain-o kaimashita (single pair)

Takako-GA wine-ACC bought
 ‘Takako bought wine...’

(37) a. Yenghui-nun wain-ul sassta.... (pair list)

Yenghui-NUN wine-ACC bought
 ‘Yenghui bought wine...’

b. Yenghui-ga wain-ul sassta. (single pair)

Yenghui-GA wine-ACC bought
 ‘Yenghui bought wine’

Again, despite these particles have been analyzed as conveying the discursive notion of ‘topic’, in these cases we cannot talk about a topic, since it answers partially the question and might not be mentioned in the previous discourse. Furthermore, as argued recently by some scholars (*cf.* Munakata 2002, Kuroda 2003, Maruyama 2003), they should be better reanalyzed as marking ‘contrast’, one of the core properties of focal elements. Hence, the appearance of these contrast particles in pair-list answers but not in single-pair ones would follow from the matching type of the former ones and the conjoined type of the latter ones.

4. Summary and Conclusions

In this paper, I have analyzed the properties of the answers of multiple-Wh questions. I have argued that in these constructions, we have a split focal structure and that at *logical form*, it leads towards having a pair of elements as being the actual focus. This analysis provides us with a natural understanding of the question-answer pairings since all the material that stands for a variable in the question is taken to be focal in nature. Thus, treating these answers as instances of split foci, we can dispense with the theoretical primitive of 'contrastive topic' and gain in understanding of the interface phenomena observed crosslinguistically.

References

- Aske, J., 1997, *Basque Word Order and Disorder*, PhD. Diss: U. Berkeley.
- Bolinger, D., 1958, «A Theory of Pitch Accent in English», *Word* 14. 109-149.
- , 1978, «Asking More Than One Thing at a Time». *Questions* ed. by H. Hiž, 107-130. Dordrecht: Reidel.
- Bošković, Ž., 2002, *On the interpretation of Multiple Questions. Celebration: An Electronic Festschrift in honor of Noam Chomsky's 70th birthday* ed. by J. Bresnan et alii: MIT Press. URL: <http://cognet.mit.edu/library/books/chomsky/celebration/>
- Büring, D., 1999, *Topic. Focus: Linguistic, Cognitive, and Computational Perspectives*, ed. by Peter Bosch & Rob van der Sandt. Cambridge (UK): Cambridge U. P.
- , 2003, «On D-Trees, Beans, and B-Accents», *Linguistics & Philosophy* 26:5. 511-545.
- Chomsky, N., 1973, «Conditions on Transformations». *A Festschrift for Morris Halle* ed. by S. Anderson & P. Kiparsky: Holt Rinehart & Winston. [Reprinted in Chomsky, N., 1977, *Essays on Form & Interpretation*. New York: Elsevier North Holland. 81-162].
- , 1995a, *Bare Phrase Structure. Government and Binding Theory and The Minimalist Program*. Ed. by Gert Webelhuth. Oxford, Blackwell.
- , 1995b, *The Minimalist Program*. Cambridge (MA): MIT Press.
- Elordieta, A., 2001, *Verb Movement & Constituent Permutations in Basque*. Leiden: LOT.
- Godjevać, S., 2000, *Intonation, Word Order, and Focus Projection in Serbo-Croatian*. PhD. Diss: The Ohio State University.
- Gutiérrez-Rexach, J., 1999, «Interrogatives and Polyadic Quantification. Proceedings of the International Conference» in *Questions*, Liverpool: University of Liverpool. 1-14.
- Hagstrom, P. A., 1998, *Decomposing Questions*. PhD. Diss: MIT.
- Herburger, E., 2000, *What Counts: Focus and Quantification*. Cambridge (Mass.): MIT Press.
- von Heusinger, K., 1999, «Intonation and Information Structure». Manuscript: U. Konstanz.
- Higginbotham, J. and R. May, 1981, «Questions, Quantifiers and Crossing», *The Linguistic Review* 1. 41-80.
- Irurtzun, A., 2003a, «The Intonational Phonology of Errenteria Basque». Paper presented at the 1st PaPI (Phonetics and Phonology in Iberia) Conference, Lisbon.
- , 2003b, «A Derivational Approach to the Focus Structure». Manuscript: EHU-UPV. Basque Country & HiTT.

- Jackendoff, R., 1972, *Semantic Interpretation in Generative Grammar*. Cambridge (Mass.): MIT Press.
- Krifka, M., 2001, «For a Structured Meaning Account of Questions and Answers». *Audiatu Vox Sapientiae: A Festschrift for Arnim von Stechow* ed. by C. Féry & W. Sternefeld. Berlin: Akademie Verlag. 287-319.
- , 1991, «A Compositional Semantics for Multiple Focus Constructions», *Proceedings from the First Conference on Semantics and Linguistic Theory (SALT 1)*. 127-158.
- Neeleman, A. and T. Reinhart, 1998, *Scrambling and the PF Interface. The Projection of Arguments; Lexical and Compositional Factors*, ed by M. Butt & W. Geuder (eds.). Stanford: CSLI Publications.
- Maruyama, A., 2003, «Japanese wa in conversational discourse: A Contrast Marker», *Studies in Language* 27-2. 245-285.
- Munakata, T., 2002, «Contrastive-Topic wa as Focus Interpretation Operator». Working papers in *Biolinguistics 1: Papers on Syntax and Semantics*, ed. by Y. Endo, R. Martin & H. Yamashita. Yokohama: BAY. 21-38.
- Rooth, M., 1985, *Association with Focus*. PhD. Diss: Umass.
- Steedman, M., 2000, *The Syntactic Process*, Cambridge (MA): MIT Press.
- Tancredi, C., 1992, *Deletion, Deaccenting and Presupposition*. PhD. Diss: MIT.
- Umbach, C., 2001, «Contrast & Contrastive Topic. Proceedings of ESSLLI 2001 Workshop» on *Information Structure, Discourse Structure and Discourse Semantics* ed. by I. Kruijff-Korbayová and M. Steedman: University of Helsinki. 175-188.
- Wachowicz, K., 1974, *On the Syntax & Semantics of Multiple Questions*. PhD. Diss: U. Texas, Austin.
- , 1975, «Multiple Questions», *Linguistica Silesiana* 1. 155-166.
- Williams, E., 1997, «Blocking and Anaphora», *LI* 28-4. 577-628.
- Wold, D. E., 1998, «How to interpret multiple foci without moving a focused constituent». *UMOP 21: Proceedings of the Workshop on Focus* ed. by Elena Benedicto, Maribel Romero and Satoshi Tomioka. Amherst (MA): GLSA. 277-289.

CONSEQUENCES OF PAIR-MERGE (AT THE INTERFACES)

Aritz Irurtzun

University of the Basque Country & HiTT

Ángel J. Gallego

Universitat Autònoma de Barcelona

1. Introduction

The goal of this paper is to explore the basic properties of adjuncts and some of the well-known puzzles these dependents pose for syntactic theorizing within the scenario provided by the Minimalist Program (*cf.* Chomsky 1995 through the present). In so doing, we will briefly discuss some controversial issues, like the argument-adjunct distinction, the status of the (still poorly understood, and worse formally classified) notion of ‘deviance’, and the semantic contribution of adjuncts, but the main focus of this paper will be the formal operation of *pair-Merge*, put forward by Chomsky (2000) and assumed to handle adjunction within the current framework. In league with Uriagereka (2003), we want to argue that adjuncts can give rise to two different readings, which we will call *Markovian* and *non-Markovian*. The first one is quite common in the literature, and plausibly instantiates the Davidsonian analysis, whereby adjuncts are mechanically concatenated predicates of the event. This reading is illustrated in 1, which can be LF-translated as in 2 (focus matters aside; *cf.* Herburger 2000, Irurtzun 2003, 2006, and section 5.1):

(1) Zidane plays soccer gracefully.

(2) $[\exists e: \text{play}(e) \ \& \ \text{Agent}(e, \text{Zidane}) \ \& \ \text{Theme}(e, \text{soccer}) \ \& \ \text{gracefully}(e)]$

Attention must be paid to the LF representation in 2, as there is no scope effect whatsoever, the adjunct just being conjoined to the preceding string. As for what we are referring to as *non-Markovian* reading (by and far, the interesting one), its more salient property is related to a scopal (that is, not merely concatenative) reading of a quantificational sort. This reading is illustrated in 3, which, as the reader may easily conclude, is, at first glance, identical to 1:

(3) Zidane plays soccer gracefully.

What is the difference between 1 and 3? We argue that the *non-Markovian* reading creates a quantificational interpretation, similar to the one explored in Hernanz’s (1993) analysis of free adjuncts. In particular, 3 can be roughly paraphrased as follows: “if/whenever Zidane plays soccer, he does so gracefully”. The intended reading is perhaps more salient in 4.

(4) Zidane plays soccer gracefully, but he plays basketball horribly.

In section 5 we propose an analysis for this variety of adjuncts, but before that, we need to spell-out what our assumptions on the issues to be discussed are, and what is at stake when we talk about adjuncts.

The paper is divided as follows: section 2 concentrates on the notion of deviance and some of its interpretations within the literature; in section 3 we turn our attention to the basic operation within minimalism (*i.e.*, *Merge*), and the particular variety which deals with adjunction, *pair-Merge*; in section 4 we present arguments in favor of the (neo-)Davidsonian treatment of adjuncts and briefly consider some accounts in which adjuncts are taken to be necessary for the sentence to be well-formed (in connection with what was said in section 2); section 5 is the empirical focus of this paper, for it considers the *non-Markovian* readings; section 6 summarizes the main conclusions.

2. Adjuncts and the Strong Minimalist Thesis

In this section we would like to consider the relation between Narrow Syntax and the so-called Interfaces, particularly so with respect to the connection between the SEM component (formerly, viewed as the LF level of representation) and the Intentional-Conceptual systems. We can formulate our aim here through the following question:

(5) What are the requirements imposed by the external (interpretive) systems?

The question in 5 is obviously related to what Chomsky (1995) calls *convergence* and *crash*: roughly put, a derivation converges at the interfaces if its (legibility) conditions are satisfied.¹ As Chomsky (1995, 2000) puts it, such notions must not be misunderstood: they are determined by formal inspection, hence not constituting an “obscure and intuition-bound notion” like those inherited from more traditional studies. At this point, it is useful to underscore that one of the most important achievements by generative grammar is to put forward an explicit system which leaves no room to intuitive approaches to grammatical phenomena —this was, as a matter of fact, one the reasons to eliminate the notion of “well-formedness”:

One of the “obscure and intuition-bound notions” that should be clarified or eliminated is set of well-formed (grammatical) expressions (E-language, in the terminology of my Knowledge of Language (1986), henceforth KOL). Though unproblematic (by stipulation) in the theory of formal languages, the notion remains obscure, perhaps lacking any empirical status, for natural language [...]. The issues are far from academic. It is well known that any 2-category partition of expressions will undercut much of the most significant linguistic work. The differential effects of ECP, subjacency, selectional constraints, etc., are far more revealing than any division into well- versus ill-formed, and bear directly on central principles of UG. In contrast, the point of a [\pm WF]-dichotomy remains obscure, even if it can be established in some nonarbitrary fashion. Suppose that Jones has

¹ Obviously, convergence is nothing but a new name to the Full Interpretation Principle (cf. Chomsky 1986a, 1995).

the I-Language L, some variety of English. As far as is known, it is meaningless to ask whether a weak wh-island violation or such an expression as “misery loves company” is, or is not, a member of the E-language weakly generated by L; and nothing would follow from a discovery (or stipulation) one way or another. These expressions have their status, determined by L; they are parsable, appropriate in certain situations, have a definite meaning, etc. (Chomsky 1990: 143, 145)

Chomsky’s (1990) point is easy to spell-out: expressions must reach the interpretive systems in such a way that they can receive an interpretation, that being all SEM can plausibly manipulate. Needless to say, interpretations may have a vast (and often unexpected) range of possibilities, some of them plausibly considered as deviant, but this must not mislead us, for deviance is not, unless we define it in a fine-grained way (and we lack any method for that, as far as we know), a criterion to rule out expressions. Let us suppose, following Chomsky (1990, 2000, 2005), the next working hypothesis:

(6) SEM assigns linguistic expressions an interpretation

The statement in 6 is not to be understood as an output filter (or an internal principle), it is just a consequence of the idea that syntax is an optimal solution to interface matters —Chomsky’s (2000) *Strong Minimalist Thesis*. In the case at issue, by proposing 6 we are assuming that SEM cares about one thing, to be able to assign an interpretation to the expressions the computational system generates. But even if we assume so, it is important to be clear about what determines if expressions receive ‘an interpretation’. Feature checking, in the sense pursued since Chomsky (1981), is a familiar candidate to test this. Consider the data in 7:

- | | |
|-----------------------------------|-------------------------------------|
| (7) a. *María cantamos. (Spanish) | b. *Me se entregó. (Spanish) |
| María sing-3.PL | CL-to-me CL-to-him/her gave-up-3.SG |
| ‘María sing’ | ‘He gave me to him/her’ |

Note that the problem in 7 is not only that those sentences are unintelligible: they are uninterpretable. Thus, the problem is not semantic in any obvious way (the theta affairs of those expressions make perfect sense), but rather syntactic: in both cases, agreement fails, which can be formally expressed by arguing that Chomsky’s (2000; 2001) *Agree* cannot value the (uninterpretable) ϕ -features of the functional categories T and v^* , hence causing a crash. Consider next the sentences in 8, which are interpretable, however odd they may sound:

- | |
|--|
| (8) a. María es lingüista amablemente. (Spanish) |
| María be-3.SG linguist gently |
| ‘María is a linguist gently’ |
| b. María se leyó el libro durante dos horas. (Spanish) |
| María CL read-3.SG the book during two hours |
| ‘María read the book during two hours’ |

Like 7, the sentences in 8 are usually ruled out as ungrammatical by most scholars and grammar books, so are the ones in 9, 10, and 11, for reasons we return to:

- (9) a. Love fears the chair. b. Golf plays John.
 c. Colorless green ideas sleep furiously.
- (10) a. Elena drank a beer during 5 minutes.
 b. Elena saddled the horse during 1 hour.
- (11) a. Mary arrived the book in the table.
 b. Juan went the car. c. Juan broke.

It is common practice in the literature to sanction these sentences as ill-formed by appealing to semantic notions, like selectional and subcategorization features (in Chomsky's (1965) sense) or *Aktionsart*. As we just said, it may perfectly be the case that the expressions in 8 sound deviant, but this must not lead us to conclude that they must undergo syntactic licensing before reaching SEM, let alone that they cannot be generated, since actually they can. Let us be a little bit more specific: what is the problem with these sentences? Under fairly standard accounts, 8a and 8b are said to be out because adjuncts must meet aspectual requirements, but this is odd to say the least, because adjuncts do not participate in *Aktionsart* business —only objects do (see Harley 2003). In sum, arguing that adjuncts must undergo syntactic licensing is not only empirically wrong, but conceptually problematic as well.²

Our reasoning is a *non*-standard one, undoubtedly, and we could be urged to answer whether this is not too generous an account, one opening *Pandora's Box*. We believe it is not. This observation could be made, for instance, by approaches like Bosque's (1989), which could be taken as a proposal in which adjuncts must be licensed:

We might recall that the so-called adjuncts are not freely added to any given predicate, for it is obvious that not all of them denote actions or proceses which take place at a time and are carried out in a certain manner, with a certain goal, and in a certain place. If the sentence Juan bought a yacht allows manner adjuncts, while Juan has a yacht does not, it is because inserting adjuncts is not as free as is normally thought. [from Bosque (1989: 137) —our translation, AI & AG]

Once again, notice how Bosque's (1989) perspective raises conceptual and empirical questions. On conceptual grounds, it seems rather sensible to us that sentences like 8, 9, 10, and 11 have no problems upon reaching SEM. Empirically, the issue

² This is also clear from what can be gathered from the GB literature, where (to the best of our knowledge) there was no Adjunct Criterion oposed to the Theta Criterion (cf. Chomsky 1981). Likewise, the presence of modifiers in traditional transformational grammars of the Standard Theory and Extended Standard Theory is regarded as optional, not having any phrase structure rule specific for them.

arises as to what mechanisms are to be invoked to license adjuncts. In principle, one could postulate different devices to rule in adjuncts (*e.g.*, SPEC-head configurations, agreement processes, etc), and further proceed to find out whether it is empirically satisfactory, but this is not the issue: the issue is whether there is *bona fide* evidence that there must be something like such devices (apart from our intuition, of course).

Consider, in this respect, the examples 12:

- (12) a. *Furiously sleep ideas green colorless.
 b. *What_i did Mary believe [the idea that John bought t_i]?
 c. *Where_i have you bought that book [because John was t_i]?

There is a very telling difference between 8-9-10-and-11 *vis-à-vis* 12: the latter cannot (and are not) generated. For us, all the cases in 8, 9, 10, and 11 are semantically odd, causing a post-SEM problem, about which we do not have anything useful to say.³ What is going on in 12 is somewhat (and crucially) different: we are before syntactic violations of both context-free 12a and context-sensitive processes 12b,c.

In what follows we will be assuming this stiff viewpoint, and will be concerned with properties that adjuncts display on purely computational grounds, hence avoiding any account in which adjuncts must be licensed. In order to restrict the range of data, we will limit ourselves to the study of what Ernst (1998) calls participant or semi-argumental adjuncts, which could be said to receive theta-roles like /Goal/, /Instrument/, /Benefactive/, /Locative/, /Source/, and /Manner/; as a matter of fact, here we will be assuming that adjuncts relate to the main (verbal) predicate through a variety of theta-roles, in the sense of what Larson & Segal (1995) discuss.

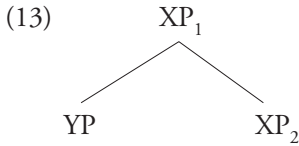
3. Adjuncts and pair-Merge

Before delving into the different readings adjuncts may give rise to, we must assess the technical details of these dependents: their formal nature, the SEM and PHON puzzles they give rise to, and the operations that have been put forward to handle them.

Within minimalism, there is one basic structure-building operation, *Merge*, which inherits the role of X-bar algorithms of previous models. Since Chomsky (2000), *Merge* is assumed to come into two flavors: *set-Merge* and *pair-Merge*. The former is the minimalist alias for *substitution* and is of little interest in what we have to say here (it creates regular phrase-structure configurations); *pair-Merge*, obviously, replaces *adjunction* and constitutes the focus of this paper.

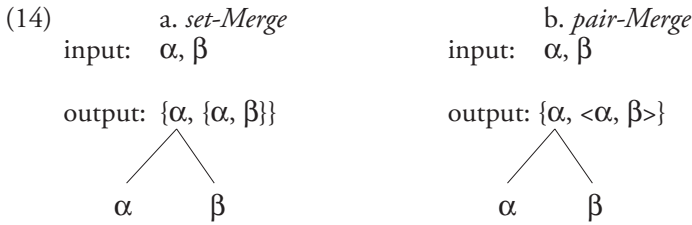
The most standard account of adjunction goes back to May's (1985) and Chomsky's (1986) analyses, whereby a category is split into (two) segments, as depicted in 13:

³ Furthermore, the oddity of this type of sentences can be exploited to gain a poetic function; thus, it is not uncommon to see sentences akin to those in 8-10 in poems (in fact, even the famous 9c has been used in such a way by authors like John Hollander and Clive James). Sentences like those in 12 on the other hand are restricted to examples of agrammaticality in theoretical linguistics bibliography



The goal of 13, or any notational variant of it, is to render YP (*i.e.*, the adjunct) structurally out-of-sight so that all formal structural dependencies one can think of cannot apply (*e.g.*, dominance, c-command, sisterhood, etc.). This is intended to capture the inherent asymmetry Chomsky (2004) thinks adjunction involves, with adjuncts playing no syntactic role. The main idea is, details aside, the same that Chomsky (1995) put forward when he cornered adjuncts out of Narrow Syntax given that they participate in no computational business (*e.g.*, they form no thematic configuration, they enter into no Case assignment mechanism, etc.), they are just ‘there’, as Boeckx (2003) puts it, being interface-driven creatures.

Within minimalism, the formal distinction between set-Merge and pair-Merge is expressed as in 14:



14b differs from Chomsky’s (1995) ‘Bare Phrase Structure’ formulation in that the asymmetry is captured in the operation rather than in the label. Here we will adopt the specifics of Chomsky’s (2004) analysis of adjuncts, and his idea that these syntactic objects are placed in a separate plane. This view is consistent with Lebeaux’s (1991) findings about anti-reconstruction effects without violating cyclicity, and dispensing with late-insertion analyses (*cf.* Lebeaux 1991 and Stepanov 2001):

- (15) a. $[_{CP} [_{DP} \text{Which picture [that John}_z \text{ took]}]_i \text{ did he}_z \text{ like the most } t_i \text{]?}$
 b. $*[_{CP} [_{DP} \text{Which claim [that John}_z \text{ did not like Mary]}]_i \text{ did he}_z \text{ made } t_i \text{]?}$

In 15a there is no Condition-C effect because the relative clause, being an adjunct, does not need to reconstruct in the first-Merge position of the object DP. Chomsky (2004), nevertheless, does not preclude the possibility of having reconstruction effects when the structure is shipped to the interfaces. Hence, he proposes that the complex (recall, in parallel) structure is undone when *Transfer* applies by means of an operation called *Simplification* (SIMPL) —actually, an optional part of *Transfer*. The ‘optional’ part is relevant, for it correctly predicts both 15a and 16, as noted in Gallego (2006), where the adjunct does reconstruct in the first-Merge position of the object DP.

- (16) $[_{CP} [_{DP} \text{Which papers [that he}_z \text{ wrote]}]_i \text{ did every linguist}_z \text{ publish } t_i \text{]?}$

In 16 the QP *every linguist* binds the pronoun *he*, indicating that SIMPL affects the lowest position of the chain, not the one that gets transferred to the PHON component. This is not an original datum, all in all, for it was already noted by Lebeau (1991), who used examples like 17 to support a late-insertion analysis of adjuncts:

- (17) [_{CP} [Which paper that he_j gave to Bresnan_z]_i did every student_j think that she_z would like t_i]]?

In 17, the relative clauses must be simplified in the SPEC-C position of the embedded verb (*like*), the position where the relevant binding effect applies.

Together with these binding data, there are further grounds to think that adjuncts manifest a paratactic nature (or *Markovian*, to use Uriagereka's 2004 terminology). If so, we expect for adjuncts to trigger weak semantic effects, along with null context-sensitive consequences. We believe this is true, and from this it follows the concatenative (scopeless) nature of adjuncts we mentioned at the outset. Things being so, we do not expect the sentences in 18 to be different, truth-conditions-wise.

- (18) a. John kissed Mary passionately in the park.
b. John kissed Mary in the park passionately.

Plausibly, the semantic interpretation of the examples in 18 is a direct consequence of adjuncts not creating regular phrase structure configurations, but a more limited, *Markovian*-like, array. More data can be used to make the same point. Consider the examples in 19, adapted from Uriagereka (2003): only the adjunct in 19a licenses the NPI, a fact we take to suggest that adjuncts-to-the-left give rise to *bona fide* phrase structure (in particular, we want to argue that such adjuncts are merged as SPECS).

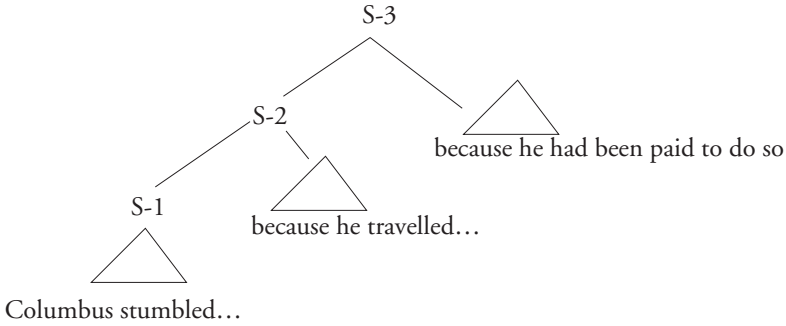
- (19) a. Bajo ninguna circunstancia puede ningún chico hacer eso. (Spanish)
under no circumstance can-3.SG any boy do-INF that
'Under no circumstance can any boy to-do that'
b. *Puede ningún chico hacer eso bajo ninguna circunstancia. (Spanish)
can-3.SG any boy do-INF that under no circumstance
'Any boy can do that under no circumstance'

Ellipsis processes also confirm Chomsky's (2004) pair-Merge analysis. Lasnik & Uriagereka (2005) note that in a sentence like 20 the gapping chunk *do so* can be interpreted either as 21a or as 21b, the reading in 21c being impossible.

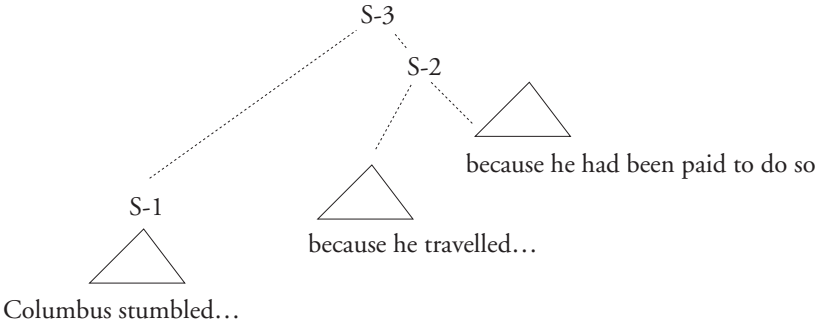
- (20) Columbus stumbled onto Santo Domingo because he travelled parallel...
... to the Equator from the Canary Islands because he had been paid to *do so*.
(21) a. do so: Travel parallel to the Equator from the Canary Islands
b. do so: Stumbled onto Santo Domingo
c. do so: *Stumbled onto Santo Domingo because he travelled parallel...
... to the Equator from the Canary Islands

This fact argues, yet again, for the *Markovian* syntax of 24, for neither 22 nor 23 would do: the former would predict 21c is possible, while 22 would rule out the interpretation in 21b.

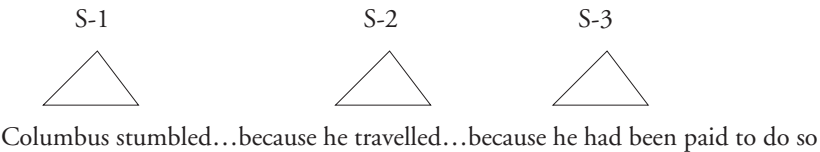
(22)



(23)



(24)



But adjuncts not only posit semantic puzzles, they do linearization ones as well: on the one hand, these dependents seem to combine in a regular fashion with the VP, scoping over whatever they adjoin to, but they normally appear to the right, which, under any version of Kayne's (1994) LCA, predicts that they should be to the left. Martin & Uriagereka (2000) consider three possible solutions to this paradox:⁴

- (25) a. Kayne's (1994) LCA is wrong.
- b. Final linear order obtains by means of massive realigning movements.
- c. No command relations can be established in adjunction.

⁴ An alternative route is taken by Larson (2004), who, in order to capture the NPI, binding, linear order and focus properties of adjuncts, assumes that adjuncts Merge with the verb before arguments do.

Martin & Uriagereka (2000) dismiss 25a and 25b, endorsing 25c. As these authors note, 25c may well be the solution Chomsky (1995: 339) hints in passing, when he defines *c-command* as in 26:

(26) *C-command*

X *c-commands* Y if (a) every Z that dominates X dominates Y and X and Y are disconnected.

Here, the ‘disconnected’ part of the definition in 25 is the key, for it is concerned with the segment/category distinction we considered above (*cf.* May 1985 and Chomsky 1986). In this respect, Chomsky (1995: 340) notes that “[i]f ‘disconnected’ in [26] requires [strong] dissociation of X, Y —say, that neither is a segment of a category that contains the other— then no ordering is determined for [X, YP₁] by the LCA.” We will assume that this is precisely what pair-Merge buys us, a weak form of dissociation in which no label is created (*cf.* Chametzky 2000, Hornstein *et al.* 2005, Moro 2000, and Uriagereka 2003). This solves the technical problem about the LCA, but we are not done yet: even if adjunction does not create a canonical phrase structure dependency, we do not want adjuncts to be completely dissociated from the VP they modify, so we must find an alternative, *c-command*-less, modification dependency which captures the scope effects and linear order we have seen in section 1. By exploiting the notion of *Numeration* (NUM; *cf.* Chomsky 1995), Uriagereka (2003) suggests a dynamic activation procedure that derives both linear order and scope effects. In this paper, we will assume this proposal, whose details are as follows:

(27) *Syntactic Activation*

A syntactic object SO is activated when it leaves a Numeration NUM and enters a derivational workspace D_{WS}

We also assume 28:

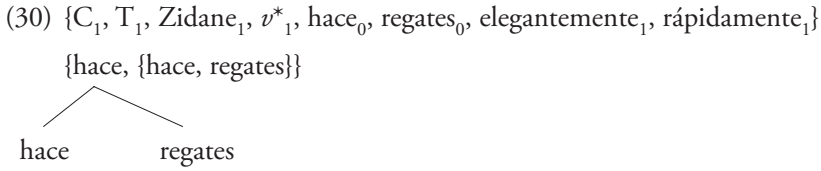
(28) *Consequences of Syntactic Activation*

A modifier’s derivational activation directly determines its linear order and scope.

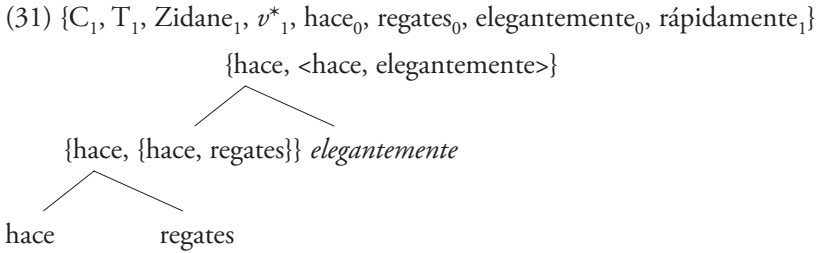
27 and 28 are not principles, but rather particular assumptions about the nature of derivational dynamics. 27 is relevant in that it capitalizes on ordering, but, crucially, we want it to apply to both arguments and adjuncts. However, when applied to the former, regular *c-command* relations emerge, and 28 becomes useless: linear order and scope are not decided during the computation, but at the interfaces. We want to argue that 27, together with the particular effects of pair-Merge, is what allows us to seriously entertain 28: since *c-command* is unavailable, the system resorts to an internal device in order to yield the desired output. Let us test this process with the example in 29b, with the NUM in 29a:

- (29) a. {C₁, T₁, Zidane₁, v*₁, hace₁, regates₁, elegantemente₁, rápidamente₁}
 b. Zidane hace regates elegantemente rápidamente. (Spanish)
 Zidane do-3.SG dribblings gracefully quickly
 ‘Zidane makes dribblings gracefully quickly’

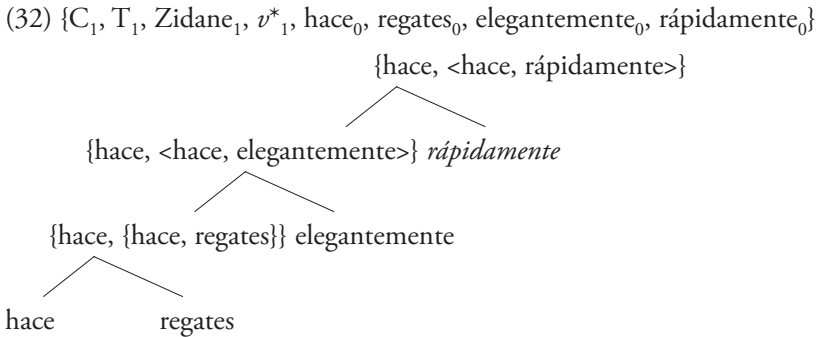
The first derivational step involves the creation of a verb-complement dependency by activating *hace* and *regates*. Crucially, set-Merge suffices to do that:



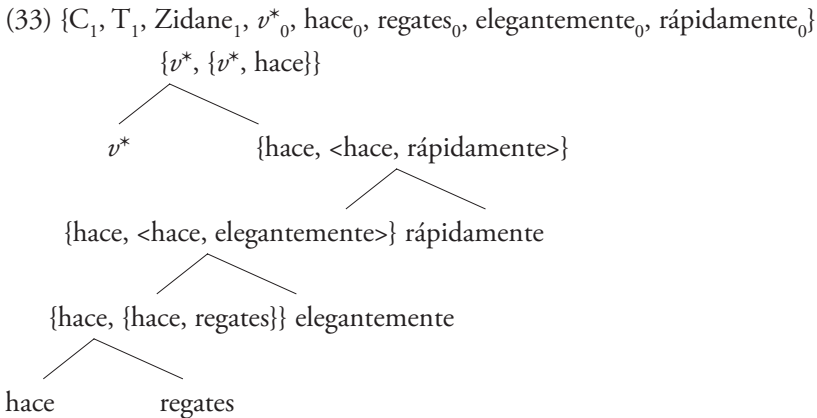
In the next step, the adjunct *elegantemente* is activated. Since pair-Merge creates no new label, 28 overrides the interface role of Kayne's (1994) LCA.

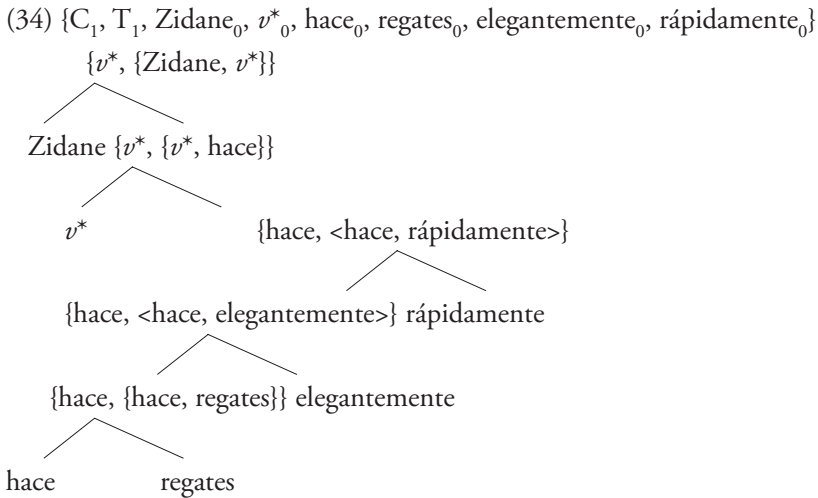


The adjunct *rápidamente* is activated next:

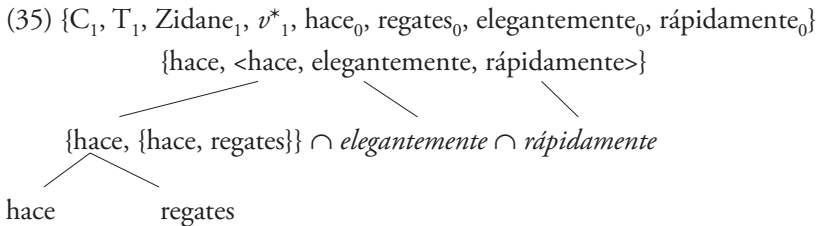


The following operations involve set-Merge of v^* and the external argument, the DP Zidane.





But we are cheating. The derivation from 30 to 34 is actually the one we will assign to *non-Markovian* readings of adjuncts. The syntax of purely *Markovian* readings would still stick to Chomsky’s (2004) pair-Merge, the activation procedure applying in a different mode: all adjuncts would be activated ‘at the same time’ (in the same ‘dimension’, to use Uriagereka’s 2003 terms),⁵ as shown in 35:



In 35, the adjuncts *elegantemente* and *rápidamente* stand in a structural relation that has no scope import whatsoever, thus accounting for the scopeless, list-like, interpretation of adjuncts.

This is enough for this section. In the preceding lines, we have seen some of the most important properties which make adjuncts special creatures. These concern (but are not restricted to) linearization, binding, and semantic effects, which makes us believe it is reasonable for them to be introduced by an additional mechanism: Chomsky’s (2004) pair-Merge. Before focusing on the two readings adjuncts can trigger, in the next section we will explore the basic semantic import of adjunct elements, and will discuss their impossibility of having them in verb-meanings. The corollary of this discussion will be that semantically adjuncts are just like they are syntactically; *i.e.* just adjuncts.

⁵ In (34), we assume that the computational system can operate with more than one syntactic object at once, so we cannot call this instance of Merge ‘pair’-Merge.

4. Semantic import of adjuncts

Having stated that adjuncts are merged in a separate plane *via* the structure-building operation of pair-Merge, in this section we are going to analyze the semantic import of adjuncts and the way they contribute the meaning of an utterance. Basically, the question we want to analyze is the following one:

(36) Do adjuncts show the same semantic import as arguments?

First we should make clear that, as already stated, in this work we are just going to observe a small subset of adjuncts, namely, those adjuncts that appear at the right edge of the clause. Unfortunately, we don't have much to say about the semantics of more problematic left peripheric scopal adjuncts (speaker oriented adverbs and so (*cf.* Fodor 1972 for discussion on these issues)). To begin with, we will review very briefly a possible analysis of adjuncts as being selected by the predicate, that is, as saturating a verbal function. Then, we will show that this type of analysis is mislead and will propose that, just like in syntax, in semantics adjuncts are just that; adjuncts (*i.e.*, they are not selected by the verb (not even the so-called *necessary* ones)). The argumentation will be based in the potential unboundedness of adjunct clustering and the analytic relations between sentences with and without adjuncts. In order to capture these properties, we propose an eventish semantics whereby adjuncts are (just) event predicates introduced *via* predicate-concatenation.

4.1. Adjuncts in verb-meanings

One of the most widely held analysis of verb-meanings takes verbs to denote open functions that get saturated with the arguments they take. Thus, the lexical entry of the verb 'stabbed' in 37 would be something along the lines in 38:

(37) Brutus stabbed Caesar. (38) $\lambda y.\{\lambda x. \text{true iff } x \text{ stabs } y\}$

That is, the transitive verb 'stab₂' denotes a dyadic function that will give the value *true* in case 'x stabs y'.⁶ Thus, an analysis of adjuncts as being directly selected by the verb would require, in a functionalist approach to verb meanings like the one in 38, a richer lexical entry for the verb in 39 in order to accommodate the insertion of the adjunct 'in the ides of March'. It could be represented as in 40, where *t* denotes a temporal variable.

(39) Brutus stabbed Cesar in the ides of March.

(40) $\lambda y.\{\lambda x.\{\lambda t. \text{true iff } x \text{ stabs } y \text{ in } t\}$

One could wonder for the necessity of including the adjunct in the verb-meaning, after all, it seems to be just optional. However, this claim is controversial, and some constructions where adjuncts are apparently needed have been claimed to show the need for the inclusion of these elements in verb-meanings. In section 2 we advanced some cases, here we will be more explicit in order to show the problems that such a position entangles. For instance, Grimshaw (1990) observes that in some construc-

⁶ We will use subscript numbers to signal the number of arguments that a predicate takes.

tions like the passive in 40, the appearance of an adjunct seems to be mandatory, thus, 41a, with the agent expressed, is completely grammatical whereas 41b is ungrammatical (according to the judgments in Grimshaw 1990):

- (41) a. The city was destroyed *by the enemy*. b. *The city was destroyed [Ø].

Furthermore, the data in 42 reported by Grimshaw & Vikner (1993) would show that it is not strictly the agent what is lacking in a passive, but just any type of adjunct:

- (42) a. *This house was built. c. This house was built yesterday.
b. This house was built by a French architect.

Similarly, another type of research that would show the need of the insertion of adjunct-like information in verb-meanings would be the contextualist approach of ‘Unarticulated Constituents’ (*cf. i.a.*, Recanati 2002). According to this trend of research, a predicate like ‘to rain’ metaphysically demands a place to apply, and so, if not explicitly provided in the syntax, this information has to be provided by the context, to give a value to a covert variable (_C):

- (43) It is raining_C.

These data, if consistent, could be taken to force the introduction of adjuncts (or a variable for content of an adjunct in the case of 43) in the lexical entries of predicates. However, as we argue in the next section, we believe that this analysis is misguided.

4.2. Problems for this analysis

We think that there are two main sets of interrelated problems for an analysis that purports the lexical requirement of adjuncts:⁷ (i) the fact that adjunction can take place unboundedly, and (ii) the mysterious analyticity between sentences with adjuncts and sentences without them.

(i) *Adjunction without limits*: it is a truism that a sentence can potentially have an unbounded number of adjuncts. As an example, observe sentence 44, taken from Bresnan (1982):

- (44) Fred deftly handed the toy to the baby by reaching behind his back over lunch at noon in a restaurant last Sunday in Back Bay without interrupting the discussion.

Here, notwithstanding the issue about the argumentality of the dative phrase ‘to the baby’, eight adjuncts can be clearly identified: ‘by reaching’, ‘behind his back’, ‘over lunch’, ‘at midnight’, ‘in a restaurant’, ‘last Sunday’, ‘in Back Bay’ and ‘without interrupting the discussion’). The example stops there but we could add as many spatio-temporal or purpose clauses as we can imagine. Then, the question is that postulating a lexical requirement for adjuncts would require a very complex entry for the predicate ‘to hand’ in this case, and, basically, different and *ad hoc* lexical entries for

⁷ Here we will provide counterarguments only to the idea that adjuncts are required syntactically, and we won't talk about the contextual variable approach. See Cappelen & Lepore (2005) for a critique of such a view.

each time a predicate appears with an adjunct. This argument brings us to the second problem that we observe with a proposal that postulates the lexical requirement of adjuncts: the analyticity problem.

(ii) *Inferences that should not exist: analytic relations*: Postulating a lexical requirement for adjuncts implies, for instance, postulating different lexical entries for the predicate ‘to stab’ in 45 and in 46, the first one would be a ternary predicate (‘to stab₃’) that requires the syntactic presence of two participants and a time, and the second one a binary predicate that just requires two participants (‘to stab₂’):

(45) Brutus stabbed₃ Cesar in the ides of March.

(46) Brutus stabbed₂ Cesar.

Such is the logic implied the argument: there we have two different and independent lexical entries; ‘to stab₃’ and ‘to stab₂’ (in principle, as independent as ‘elbow’ and ‘paraphrase’), but so it happens that they denote the same type of event and they have the very same phonological matrix, roughly /stæb/. Obviously, the problem lies in the purported independence between the predicates in 45 and 46, given that English speakers know analytically, *i.e.* independently of facts, that whenever the content of 45 is true the content of 46 is also true (the same as with the causative-inchoative alternation (*cf.* Pietroski 2003)). Again, the remarkable fact about this is that the relation between those sentences is analytical; our very knowledge of English suffices to provide this information. Thus, the only way to explain the analyticity of the inference having two independent predicates would be *via* the introduction of a meaning postulate relating both predicates; something like 47:⁸

(47) $\text{Stab}_3(x, y, t) \leftrightarrow \text{Stab}_2(x, y) \ \& \ \text{Time}(t)$

And, obviously, the problem is that if adjunction can apply without limits, there should be as well an unlimited number of meaning postulates relating predicates with adjuncts and predicates without them. Thus, we don’t believe that an analysis based on the lexical necessity of adjuncts is of the right track.

In the next section we propose to analyze adjuncts in a different manner; as predicates of the event introduced by conjunctions.

4.3. Proposal: adjuncts and predicate conjunctions

Having stated the problems of an analysis of adjuncts as lexically required by the verb, our goal in this section is to analyze the semantic import of adjuncts. Our analysis will be based in the Davidsonian tradition (*cf.* Davidson 1967a, Taylor 1985 among others). According to this trend of analysis adjuncts add a predicate to the event denoted by the verb. Thus, for instance, sentence 46, repeated here as 48a for convenience, has the LF in 48b whereas sentence 45, 49a here, corresponds with the logical form in 49b:⁹

⁸ In order to be consistent with the proposal we are discussing we are using functionalist representations for the predicate.

⁹ We abstract away from the representation of tense for simplicity.

- (48) a. Brutus stabbed Cesar.
 b. $\exists e$ [Stab(e) & Agent(e, Brutus) & Theme(e, Caesar)]
- (49) a. Brutus stabbed Cesar in the ides of March.
 b. $\exists e$ [Stab(e) & Agent(e, Brutus) & Theme(e, Caesar) & Temporal Location (e, ides of March)]

Following Davidson (1967a), this type of representation allows us to account for the paradigm of entailments of 50:

- (50) a. Brutus stabbed Cesar in the back with a knife.
 b. Brutus stabbed Cesar in the back.
 c. Brutus stabbed Cesar with a knife.
 d. Brutus stabbed Cesar.

Having these four sentences, we observe that the proposition expressed by sentence *a* entails all *b*, *c*, *d*, as well as the conjunction of *b* and *c*, that *b* entails *d* and, likewise, that *c* entails *d*. Obviously, neither *b*, nor *c*, nor *d* entails *a*, but the crucial fact is that the conjunction of *b* and *c* doesn't entail *a*. As pointed out by Davidson, we can account for these patterns of entailment if we take the modifiers in 50 to be pure adjuncts, predicates of the event denoted by the verb. Thus, in a neo-Davidsonian fashion, the LF representations of the sentences in 50 would be those in 51, with the correlation $x \rightarrow x'$:

- (51) a'. $\exists e$ [Stab(e) & Agent(e, Brutus) & Theme(e, Cesar) & Spatial-Location(e, back) & Instrument(e, knife)]
 b'. $\exists e$ [Stab(e) & Agent(e, Brutus) & Theme(e, Cesar) & Spatial-Location(e, back)]
 c'. $\exists e$ [Stab(e) & Agent(e, Brutus) & Theme(e, Cesar) & Instrument(e, knife)]
 d'. $\exists e$ [Stab(e) & Agent(e, Brutus) & Theme(e, Cesar)]

With this type of representation, the entailment pattern is naturally explained as entailments between the LFs of the propositions expressed by the sentences.

Furthermore, the possibility of adjunction without limits doesn't create any problem since adjuncts are introduced as such, as adjuncts to the event by predicate conjunction.

In the next section we will explore the nature of adjunct clustering departing from this analysis of adjunct semantics.

5. Adjunct clusters and the interpretation of non-Markovian adjuncts

In this final section we would like to propose a syntax from which *non-Markovian* readings can follow. As advance in the outset, Uriagereka (2003) analyzes clustered adjunction as giving rise to two types of modification patterns: a *Markovian* one and a *non-Markovian* one. The first one arises when adjuncts show up in a paratactic fashion, creating no scope/framing effect among them (this is, in short, the reading corresponding with the Davidsonian treatment of adverbs); as for the *non-Markovian* one, it involves a framing effect. Martin & Uriagereka (2000) discuss left-to-right scopal effects as we saw in the introduction, but we can also find the opposite pattern whereby the outermost adjunct is somehow interpreted within the syntactic projection of the innermost one. Consider, to see this, 52, which can display both interpretations:

- (52) Juan se cayó por borracho por idiota. (Spanish)
 Juan CL fell-3.SG because drunk because idiot
 ‘Juan fell because (he was) drunk because (he was) idiot’

→ *Markovian* (scopeless):

‘The reasons why Juan fell are that he was drunk, idiot, etc.’

→ *Non-Markovian* (scopal):

‘The reason why Juan fell is that he was drunk, which, in turn, happened because we was an idiot’

Compare next 52 with 53, where the adjunct ordering is reversed:

- (53) Juan se cayó por idiota por borracho.
 Juan CL fell-3.SG because idiot because drunk
 ‘Juan fell because (he was) idiot because (he was) drunk’

→ *Markovian* (scopeless):

‘The reasons why Juan fell are that he was drunk, idiot, etc.’

→ *Non-Markovian* (scopal):

‘The reason why Juan fell is that he was an idiot, which, in turn, happened because we was drunk’.

Note that the *Markovian* reading is semantically identical in both 52 and 53, regardless of linear order, for adjuncts show no scope effects. Things are different when it comes to the *non-Markovian* reading: in those cases, adjuncts involve a different (*non*-paratactic or scopal) syntax with *non*-trivial consequences for the semantics. The reality of the scopal facts can be clearly illustrated as in 54a (in Spanish, and without any list intonation), where the cause of the event of Juan’s scare is the scar, which was caused by the accident; but, crucially, the intentional cause of Juan’s scare cannot be the accident, nor the cumulating force of the scare and the accident. Thus, as we will argue, the bare concatenative LF of 54b will not be accurate to represent the *non-Markovian* meaning of 54a (*pace* Parsons 1990):

- (54a) Juan se asustó por la cicatriz por el accidente. (Spanish)
 Juan CL scared-3.SG because the scar because the accident
 ‘Juan scared because of the scar because of the accident’
 (54b) $\exists e$ [scare(e) & Experiencer(e, Juan) & Cause(e, scar) & Cause(e, accident)]

Furthermore, an LF along the lines of 54b would entail that of 54c, quite inconveniently, since, remarkably, 54a does not entail 54d under the relevant reading (the accident can be completely unknown to Juan):

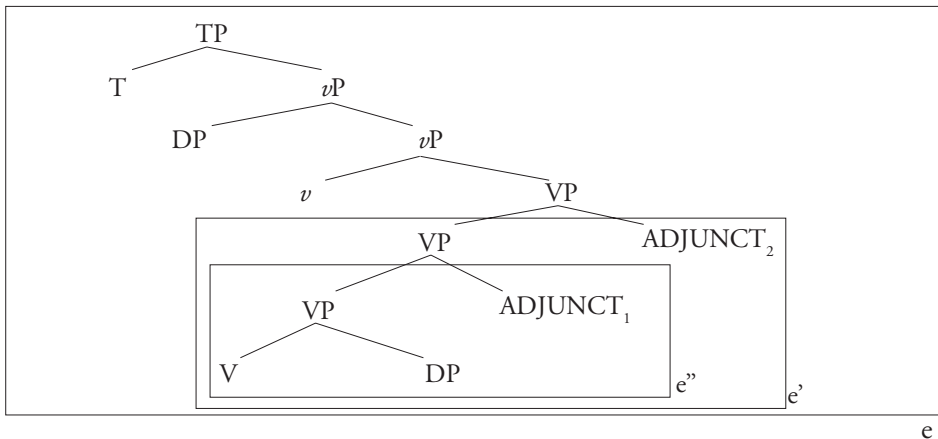
- (54c) $\exists e$ [scare(e) & Experiencer(e, Juan) & Cause(e, accident)]
 (54d) Juan se asustó por el accidente. (Spanish)
 Juan CL scared-3.SG because the accident
 ‘Juan scared because of the accident’

Interestingly enough, this pattern does not appear to be restricted to one type of adjunct. That is to say, regardless of the particular semantics (*i.e.*, cause, location, condition, etc.), different adjunct clusters behave as just indicated, hence potentially displaying two readings. This holds in the case of conditional adjuncts:

- (55) a. Uno de los dos se va a tener que ir si me pegas...
 one of the two CL go-3.SG to have-INF that go-INF if CL-me hit-2.SG
 ... si digo lo que pienso. (Spanish)
 if say-1.SG the that think-1.SG
 ‘One of the two of us will have to leave if you hit me if I say what I think’
- b. Uno de los dos se va a tener que ir si digo...
 one of the two CL go-3.SG to have-INF that go-INF if say-1.SG
 ... lo que pienso si me pegas. (Spanish)
 the that think-1.SG if CL-me hit-2.SG
 ‘One of the two of us will have to leave if I say what I think if you hit me’

So, how could we capture the semantics of these structures? Recall that in these cases each of the adjuncts denotes an event. This is patently true for the cases of 55, but we believe that it is also true for the cases of 52-53 and 54, where each of the adjuncts denotes a cause relating two events (*cf.* Pietroski 2000). Thus, we propose that in order to capture the *scopal reading*, we would need to enrich the LFs of these sentences by postulating the introduction of a new *subevent* in each syntactic subcycle. This is depicted in 56:

(56)



This type of syntactic structuration provides us immediately with the ‘framing’ semantics we observe in these constructions. Then, for instance, the LF representation we propose for the *non-Markovian* reading of 52 is the following one:

- (57) $\exists e$ [Theme(e , Juan) & Falling(e) & $\exists e''$ [Cause(e , e'') & Experiencer(e'' , Juan) & Being-drunk(e'') & $\exists e'$ [Cause(e'' , e') & Experiencer(e' , Juan) & Being-idiot(e')]]]

This type of representation captures the scopal character of these adjunct clusters where each adjunct takes scope over (or frames in a new subevent) the previous chunk of structure. The question, now, is how to get both representations from the very same syntactic structure. Our solution to this issue is to deny it; that is, to deny that both readings have the same underlying syntactic structure. Recall that in our discussion of the *Markovian* reading we underlined that the order of the adjuncts doesn't matter, as there is no scopal effects among them. Thus, a natural way to capture this fact would be to posit that these adjunct clusters are created in a separate plane as such, and then introduced as a cluster to the rest of the derivational spine (*cf.* 35). It would be then this latter pair-merge what would give the adjuncts their modification pattern. In the case of *non-Markovian* clusters like these, as we said before (*cf.* 34), each of the adjuncts would be introduced separately, triggering in each instance of pair-merge a 'framing' effect with the introduction of a new eventuality.

5.1. Two types of non-Markovian adjunct clusters

We just saw that adjunct clusters give rise to two types of readings: a *Markovian* one where there is no scope among the adjuncts and a *non-Markovian* one where there are clear scopal effects. Observe now the data in 29, modified here as 58:

- (58) Zidane hace regates elegantemente fácilmente. (Spanish)
 Zidane do-3.SG dribblings gracefully easily
 'Zidane makes dribblings gracefully easily'

There the *Markovian* reading would be like those observed for previous clusters, *i.e.*, that of a purely conjunctivist-concatenative adjunct cluster (*cf.* Davidson 1967b, Taylor 1985). Hence, we would want to propose that the LF of these readings is something along the lines in 59:

- (59) a. $\exists e$ [Agent(e, Zidane) & do(e) & Theme(e, dribblings) & Manner(e, easy) & Manner(e, elegant)] (=59b)
 b. $\exists e$ [Agent(e, Zidane) & do(e) & Theme(e, dribblings) & Manner(e, elegant) & Manner(e, easy)] (=59a)

However, note that, as we said in the introduction, the scopal reading of this sentence is the reverse of the ones in 52-53 and 54: here, the innermost adjunct seems to take scope over the outermost one. Similar observations are made by Ernst (2000; 2002), who discusses sentences like those in 60, where the outermost adjunct does not create a particular cycle within the innermost one, but rather the other way around.

- (60) a. They run fast awkwardly, but run slowly smoothly.
 b. They play soft well enough, but play loudly pretty poorly.

The question that arises at this point is this: what are the semantics of this type of expressions? Note here that these adjuncts, contrary to those involved in 52-54, do not denote eventualities, but rather, they are plane predicates of the event denoted by the verb, pure modifiers, then. Roughly, the first part of 58a is interpreted as follows: 'whenever they run fast, they run fast awkwardly'. In other words, the seman-

tics of these *non-Markovian* adjuncts is akin that of conditionals of the sort ‘P in any event in which Q’, as analyzed by Lycan (2001), of the form ‘P if Q: (e)(In(e,Q) \supset In(e,P)). In this vein, in order to capture the assertive nature of these constructions we would like to suggest that these *non-Markovian* (scopal) constructions are to receive the LF of 61, involving a universal quantification:

(61) $\forall e$ [run(e) & Agent(e, they) & Manner(e, fast)] & Manner(e, awkwardly)

It is worth pointing out that, although ultimately departing from the type of framing effect previously discussed, 61 is nonetheless similar to it in that the canonical paratactic reading vanishes. However, the modification pattern is the opposite one. The crucial point, we believe, is the focus-background partition of the clause. Recall that the LF representation in 61, with a restricted quantification over events, shows the same structure as the LFs of focus-affected readings (*cf.* Herburger 2000). Thus, we believe that this is what could be at hand in the scopal effects observed in such constructions: a pure effect of focus (hence, the likeliness of having them in a contrastive environment like in 60). In fact, note that this type of manner adjunct clusters that create a framing effect cannot be uttered in *out-of-the-blue* environments, and they have to be accompanied by a clear topic —comment intonation contour.¹⁰ Thus, the conclusion is that the scopal effect of these adjunct clusters would not derive from a framing-activation syntax like the one represented in 56, but from independent sources; the focus-background partition of the clause:¹¹

(62) [They run fast]_{Background} [AWkwardly]_{Focus}.

In fact, from a functionalist point of view, it shouldn’t be surprising that adjuncts get focused. As we saw in sections 3 and 4, they are not selected by the verb and, if they appear in the derivation, it is likely that they have such an interpretation. Then, if both manner adjuncts in the cluster are to be focused we get the *Markovian* reading, with no scope among them. If only one of them is focused, we get the *non-Markovian* reading of 61, but as we said, this reading is triggered just by having focus on one of the adjuncts. If it is the outermost one, the background-focus partition of the clause might be masked by the fact that in *out-of-the-blue* sentences nuclear stress also falls in the rightmost position.¹² If it is the innermost adjunct that is focused, its focal status is clearer, given that the intonational structure of the clause is also affected.

The corollary of this discussion would be that event-denoting adjuncts can create framing effects on their own, just by being activated in separate planes. Then, the outermost adjuncts (or events) frame the innermost ones. The case of *non-eventive* adjuncts is different since the framing effects they lead to are the opposite as those of the eventive ones. Furthermore, additional evidence supports this thesis. The data we have in mind is provided by Hernanz’s (1993) free adjuncts, which display the same eventive reading we are interested in:

¹⁰ Contrary to the event-denoting adjunct clusters that are naturally uttered in *out-of-the-blue* environments.

¹¹ Capital letters indicate the nuclear stress placement.

¹² However, as we said, the topic intonation of the pre-focal chunk makes clear the intended reading.

- (63) En París, María estudia.
 In Paris, María study-3.SG
 ‘If/When she is in Paris, María studies’

The interesting thing to note about 63 is that, just like the data in 52 through 55, the adjunct is somehow interpreted as denoting an event (of course, using the term ‘event’ in a wide sense and covering all sorts of eventualities, including ‘states’). Quite crucially for our purposes, manner adverbs, like the ones in 58 and 60, cannot trigger this ‘eventive’ reading:

- (64) Tranquilamente, María estudia.
 Peacefully María study-3.SG
 ‘María studies peacefully’

In plain terms: 64 does not mean ‘if she is quite, María studies’. This contrast suggests, once again, that there is something deep that teases event-denoting and *non*-event-denoting adjuncts apart: plausibly, as we have suggested, this follows from the latter class of modifier not being able to license an eventive subcycle. Although promising and certainly accurate (at least in descriptive terms), we cannot fail to mention that more needs to be said about these facts: how does this asymmetry affect semantic ontologies?, does it have a syntactic reflex (in phrase structure terms)?, etc. Interesting issues that we leave open for future research.

6. Summary and Conclusions

Here we have analyzed the basic nature of adjuncts as optional and *non*-selected elements (not even in the cases of the so-called ‘obligatory adjuncts’). We showed that an argumental view of adjuncts (*cf.* Grimshaw 1990, Grimshaw & Vikner 1993) is either *ad hoc* as to which adjuncts are argumental or would require the insertion of an infinite number of meaning postulates relating adjunct taking and not-taking verbs.

Summarizing, our analysis of adjuncts explores the possibility that the operation of pair-Merge has both Narrow Syntactic and Interface-like properties of its own. First, it creates ordered pairs, not sets, which can be assumed to follow from adjuncts living in a “separate plane”. As for the interfaces, pair-Merge has a special status as well: on the phonologic side, it poses an *a priori* knock down problem for Kayne’s (1994) LCA (trivially solved, if pair-Merge involves no c-command paths before Spell-out); on the semantic side, it creates “predicate composition”, together with extremely interesting interpretive effects. Following Martin & Uriagereka (2000) and Uriagereka (2003) we adopt the idea that adjuncts display two types of readings, a *Markovian* and a *non-Markovian* one: under the first one, adjuncts are interpreted as independent predicates of the event (the traditional approach stemming from Davidson 1967a), whereas under the second one, adjuncts create a framing (scopal) effect which blocks the expected entailment patterns. The two readings, we argued, derive from the different syntactic composition of the adjunct clusters.

References

- Boeckx, C., 2003, «Review of T. Ernst, 2002», *The syntax of adjuncts*, Cambridge U. P., «Studies in the Linguistic Sciences» 32. 95-101.
- Bosque, I., 1989, *Las Categorías Gramaticales*. Madrid: Cátedra.
- and P. J. Masullo, 1998, *On Verbal Quantification in Spanish, Studies on the Syntax of Central Romance Languages* ed. by O. Fullana & F. Roca, Girona: U. Girona, 9-64.
- Bresnan, J., 1982, «Polyadicity», *The Mental Representation of Grammatical Relations* ed. J. Bresnan, Cambridge (MA): MIT Press, 149-172.
- Cappelen, H. and E. Lepore, 2005, *Insensitive Semantics: A Defense of Semantic Minimalism and Speech Act Pluralism*, Oxford: Blackwell.
- Chametzky, R. A., 2000, *Phrase Structure: from GB to Minimalism*, Malden: Blackwell.
- Chomsky, N., 1965, *Aspects of the Theory of Syntax*. Cambridge: MIT.
- , 1981, *Lectures on Government and Binding: The Pisa Lectures*. Dordrecht: Foris Publications.
- , 1990, «On Formalization and Formal Linguistics», *NLLT* 8, 143-147.
- , 1986, *Knowledge of Language: Its Nature, Its Origins, Its Use*, New York: Praeger.
- , 1995, *The Minimalist Program*. Cambridge: MIT Press.
- , 2000, *Minimalist Inquiries: The Framework. Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik* ed. by D. Michaels, R. Martin & J. Uriagereka, Cambridge (MA): MIT Press.
- 2001, *Derivation by Phase. Ken Hale: A Life in Language*, ed by M. Kenstowicz. Cambridge: MIT Press. 1-52.
- 2004, «Beyond Explanatory Adequacy, Structures and Beyond». *The Cartography of Syntactic Structures*, Volume 3 ed. by A. Belletti, Oxford/New York: Oxford U. P., 104-131.
- 2005, *On Phases*. manuscript: MIT.
- Davidson, D. D., 1967a, *The Logical Form of Action Sentences, The Logic of Decision and Action* ed. by N. Rescher, Pittsburgh: University of Pittsburgh Press [Reprinted in D. Davidson. *Essays on Actions and Events* (2. ed.). Oxford/New York: Oxford U. P., 105-122].
- 1967b, *Reply to Martin, The Logical Form of Action Sentences. The Logic of Decision and Action* ed. by N. Rescher. Pittsburgh: University of Pittsburgh Press [Reprinted in D. Davidson. *Essays on Actions and Events* (2. ed.), Oxford/New York: Oxford U. P., 129-148].
- Ernst, T., 1998, «The Scopal Basis of Adverb Licensing», *Proceedings of NELS* 28, 127-142.
- 2002, *The Syntax of Adjuncts*. Cambridge: Cambridge U. P.
- Fodor, J. A., 1972, *Troubles About Actions. Semantics of Natural Language* ed. by D. Davidson and G. Harman. Dordrecht: Reidel. 48-69.
- Gallego, Á. J., 2006, «Connectivity in Markovian Dependencies». Manuscript: UAB.
- Grimshaw, J., 1990, *Argument Structure*, Cambridge (MA): MIT Press.
- and S. Vikner, 1993, *Obligatory Adjuncts and the Structure of Events, Knowledge and Language: Vol. II Lexical and Conceptual Structure* ed. by E. Reuland & W. Abraham, Kluwer, Dordrecht, 145-159.
- Harley, H., 2003, «How do verbs get their names? Denominal verbs, Manner Incorporation and the ontology of verb roots in English». Manuscript: U.Arizona.
- Herburger, E., 2000, *What Counts: Focus and Quantification*. Cambridge (Mass.): The MIT Press.

- Hernanz, M. Ll., 1993, «A propósito de los adjuntos libres». *Sintaxi. Teoria i Perspectives*, ed by in A. Viana. 125-173.
- Hornstein, N., J. Nunes and P. Pietroski, 2005, *Some Thoughts on Adjunction*, manuscript. UMD-USP.
- Irurtzun, A., 2006, «Focus & Clause Structuration in the Minimalist Program». *Minimalist Theorizing* ed. by C. Boeckx. Berlin: John Benjamins. 68-96.
- 2003, *The Locus of [+F]ocus*, ms: EHU-UPV.
- and Á. J. Gallego, 2004, *Adjuntos concatenados y (sub)cyclos sintácticos*. Paper presented at the XXXIV Simposio de la Sociedad Española de Lingüística, Madrid.
- Kayne, R., 1994, *The Antisymmetry of Syntax*. Cambridge: MIT Press.
- Larson, R., K. and G. Segal, 1995, *Knowledge of Meaning*. Cambridge: MIT Press.
- Lasnik, H. and J. Uriagereka, 2005, *A Course in Minimalist Syntax*. Oxford: Blackwell.
- Lebeaux, D., 1991, *Relative clauses, licensing and the nature of the derivation*. *Syntax and Semantics 25: Perspectives on phrase structure*, ed. by S. Rothstein. New York: Academic Press. 209-239.
- Lycan, W. G., 2001, *Real Conditionals*. Oxford: Oxford U. P.
- Martin, R. and J. Uriagereka, 2000, *Collapsed Waves*. Manuscript: U. Tsukuba & UMD.
- May, R., 1985, *Logical Form*. Cambridge: MIT.
- Moro, A., 2000, *Dynamic Antisymmetry*. Cambridge: MIT Press.
- Parsons, T., 1990, *Events in the Semantics of English*. Cambridge: MIT Press.
- Pietroski, P., 2003, *Small Verbs, Complex Events: Analyticity Without Sinonimy*. *Chomsky and his Critics* ed. by N. Hornstein and L. Antony. Oxford: Blackwell, 179-214.
- 2000, *Causing Actions*. Oxford: Oxford U. P.
- Recanati, F., 2002, «Unarticulated Constituents», *Linguistics and Philosophy* 25. 299-345.
- Stepanov, A., 2001, «The Timing of Adjunction», *Proceedings of NELS* 30. Amherst: GLSA. 597-611.
- Taylor, B., 1985, *Modes of Occurrence: Verbs, Adverbs & Events*. Oxford & NY: Blackwell.
- Uriagereka, J., 2003, *Pure Adjuncts*. Manuscript, UMD.
- 2004, *Where do Adjuncts Live?* Manuscript: UMD & UPV-EHU.
- Vendler, Z., 1967, *Linguistics in Philosophy*. Ithaca: Cornell U. P.

CASE MARKING AND PREPOSITIONAL MARKING. SOME REMARKS CONCERNING *DE*-PHRASES IN ROMANIAN

Alexandru Mardale

Université Paris

Universitatea din București, Facultatea de Litere

Abstract

In this paper, we analyze two types of nominal constructions in Romanian: DPs morphologically marked for Genitive and complex *DE*-phrases. The two types of construction are alike insofar as they involve a relation (which may either pertain to the lexical meaning of the head N or else be contextually triggered by the presence of the second argument), but they differ regarding the nature of the second argument. A strong correlation can be shown to exist between syntactic categories (DPs vs. NPs), syntactic functions (arguments vs. modifiers), Case marking (synthetic vs. analytic) and semantic type (type <e> vs. type <e, t>).

0. A few remarks about case in Romanian

Romanian is a Romance language which partially inherits from Latin morphological case, namely the Dative case and the Genitive case which are homonymous. Dative is assigned in verbal constructions (1) while Genitive is assigned in nominal constructions (2):¹

- | | |
|---|--|
| (1) am dat (cărți) regel ui (Dative) | (2) cărțile regel ui (Genitive) |
| have-1 given (books) king-the-D | books-the king-the-G |
| 'I gave (books) to the king' | 'the books of the king' |

The difference is visible when substituting by a possessive pronoun:

- | |
|--|
| (3) i-am dat (cărți) lui / * sale (Dative) |
| him-CL-D have given (books) him-D / his-G |
| 'I gave (books) to him' |
| (4) cărțile lui / sale (Genitive) |
| books-the him-G / his-G |
| 'his books' |

¹ Abbreviations used in glosses and diagrams: DE = Romanian Preposition *de*, G = (morphological) Genitive Case, D = (morphological) Dative Case, 1, 2, 3 = 1st, 2nd, 3rd person, CL = clitic, AUX = auxiliary, ACC = Accusative.

1. Introduction

Romanian displays a remarkable alternation between DPs morphologically marked for Genitive case and PPs headed by the preposition *DE*;² this alternation appears with several types of nouns: relational nouns (5a), deverbal nouns (5b), picture nouns (5c), object-denoting nouns (5d):

- | | | | |
|--------|---------------------------|-----|------------------------|
| (5) a. | fiul regelui | vs. | fiul de rege |
| | son-the king-the-G | | son-the DE king |
| b. | construirea caselor | vs. | construirea de case |
| | building-the houses-the-G | | building-the DE houses |
| c. | fotografia grupului | vs. | fotografia de grup |
| | picture-the group-the-G | | picture-the DE group |
| d. | ușa bisericii | vs. | ușa de biserică |
| | door-the church-the-G | | door-the DE church |

This phenomenon is not a characteristic of Romanian, but it appears in other languages too (see, *inter alia*, Munn 1998, Corblin 2001 and Dobrovie-Sorin 2001a, for an analysis of English or French equivalents):

- | | | | | |
|--------|---------------------------------|-----|-----------------------|-----------|
| (6) a. | <i>the room of the men</i> | vs. | <i>the men's room</i> | (English) |
| b. | <i>le fils du (de + le) roi</i> | vs. | <i>le fils de roi</i> | (French) |
| | the son of the king | | the son DE king | |

This paper is organized as follows: in sections 2 and 3 we present previous analyses and we give arguments against a unitary treatment of the two constructions; in section 4, we discuss the conditions under which the two types of constructions are used in Romanian; in section 5, we propose an analysis for each of these constructions.

2. Previous analyses

Traditional grammars (see, for example, *GLR* 1966) as well as handbooks analyze these two types of constructions as respectively synthetic (i.e. morphological) vs. analytic Genitives. The arguments in favour of such an analysis are the following:

- (i) The possibility to substitute the DPs marked with morphological case by *DE*-phrases in which *DE* would have take the functions of casual inflection (cf. *supra* (5) and *infra* (7));
- (ii) Both constructions express similar semantic values : alienable possession (7a), inalienable possession (7b), human relationship (7c), goal (7d), content (7e), location (7f), time (7g), quality (7h) etc. :

² There are other constructions with prepositions which can alternate with morphological Genitive constructions (e.g. *cartea copiilor* 'book-the children-the-G' vs. *cartea a trei copii* 'book-the A three children'). We will not discuss this type here.

(7) a.	<i>curtea de împărat</i>	/	<i>curtea împăratului</i>
	court-the DE emperor		court-the emperor-the-G
b.	<i>gulerul de cămașă</i>	/	<i>gulerul cămășii</i>
	collar-the DE shirt		collar-the shirt-the-G
c.	<i>nepotul de unchi</i>	/	<i>nepotul unchiului</i>
	nephew-the DE uncle		nephew-the uncle-the-G
d.	<i>camera de oaspeți</i>	/	<i>camera oaspeților</i>
	room-the DE guests		room-the guests-the-G
e.	<i>ostrovul de flori</i>	/	<i>ostrovul florilor</i>
	isle-the DE flowers		isle-the flowers-the-G
f.	<i>aerul de munte</i>	/	<i>aerul muntelui</i>
	air-the DE mountain		air-the mountain-the-G
g.	<i>căldura de vară</i>	/	<i>căldura verii</i>
	heat-the DE summer		heat-the summer-the-G
h.	<i>omul de datorie</i>	/	<i>omul datoriei</i>
	man-the DE honour		man-the honour-the-G

3. Limitations of the classical analysis

On the one hand, formal alternation as well as similarity of semantic values do not necessarily imply identical structures.

On the other hand, classical analysis ignores the categorial status of the adnominal constituent: DP with Genitive case vs. *DE*-NP. Both are treated the same way with respect to the distinction between DP and NP.

4. Conditions of use

As we will see in the next subsections, there are several diagnostic tests which help in distinguishing between the two types of constructions.

4.1. Formal constraints

The constructions with morphological case are necessarily nominal phrases governed by a determiner, either definite or indefinite (8):

(8) a.	<i>fiul regelui</i>	/	<i>fiul unui rege</i>
	son-the king-the-G		son-the a-G king
b.	<i>*fiul rege</i>		
	son-the king		

In contrast, the complement of *DE* cannot be headed by a determiner, regardless of its nature 9a, but can have (adjectival or prepositional) modifiers 9b:

(9) a.	<i>*fiul de rege</i>	/	<i>*fiul de un rege</i>
	son-the DE king-the		son-the DE a king
b.	<i>fiul de rege african</i>	/	<i>construirea de case din lemn</i>
	son-the DE king African		building-the DE houses of wood

4.2. Distribution in predicate position³

DPs marked with morphological case cannot appear after the copula (10a); in order for them to appear after the copula, we need to insert the so-called genitive article *al, a, ai, ale*⁴ in front of the Genitive DP (10b):

- (10) a. *fiul este regelui; *ușa este bisericii
 son-the is king-the-G; door-the is church-the-G
 b. fiul este al regelui; ușa este a bisericii
 son-the is A-the king-the-G; door-the is A church-the-G

In contrast, prepositional constructions can appear after the copula (11):

- (11) a. fiul este de rege (nu de sclav)
 son-the is DE king (not DE slave)
 b. ușa este de biserică (nu de casă)
 door-the is DE church (not DE house)

4.3. Distribution in preverbal subject position

DPs marked with morphological case are frequent in preverbal subject position whether or not they are anaphorically related to another DP (12):

- (12) *Fiul regelui nu a venit la întrunirea Curții.*
 son-the king-the-G not has-AUX come at meeting-the Court-the-G
 ‘The son of the king has not come at the Court’s meeting’

The so-called analytic Genitive (i.e. *DE*-phrases) are less natural in these position especially when the head noun takes the definite determiner and the construction is not anaphorically related to another DP (13):

- (13) ??*Fiul de rege nu a venit la întrunirea Curții.*
 son-the DE king not has-AUX come at meeting-the Court-the-G
 ‘The king’s son has not come at the Court’s meeting’

4.4. The *a avea* ‘to have’ paraphrase

DPs marked with morphological case, except the ones in which the head is a deverbal noun (see 5b above), can be paraphrased by *a avea* ‘to have’ (14):

- (14) *soția avocatului* → *avocatul are soție*
 wife-the lawyer-the-G lawyer-the has wife

This is not the case for prepositional constructions (15):

- (15) *soția de avocat* → ???
 wife-the DE lawyer

³ By predicate position we understand post-copular position (cf. Milner 1982).

⁴ This article is made up of the preposition *a* followed by the definite article.

4.5. Special cases

There are exceptions to the free substitution between the constructions with morphological Genitive and the constructions with the preposition *DE*, namely compounds. On the one hand, there are constructions taking only the synthetic form:

- (16) a. *floarea soarelui* vs. **floarea de soare*
 flower-the sun-the-G flower-the DE sun
 'sunflower'
 b. *regina nopții* vs. **regina de noapte*
 queen-the night-the-G queen-the DE night
 'night flower'
 c. *iarba dracului* vs. **iarba de drac*
 grass-the devil-the-G grass-the DE devil
 'weeds'

On the other hand, there are constructions taking only the analytic form:

- (17) a. *floarea de colț* vs. **floarea colțului*
 flower-the DE corner flower-the corner-the-G
 'edelweiss'
 b. *laptele de pasăre* vs. **laptele păsării*
 milk-the DE bird milk-the bird-the-G
 'dessert'
 c. *dintele de lapte* vs. **dintele laptelui*
 tooth-the DE milk tooth-the milk-the-G
 'milk tooth'

4.6. Interim conclusion

Once again, free substitution as well as similarity of semantic values of the two constructions are not reason enough for them to be analysed the same way. As a consequence, the Genitive analysis is not appropriate for both nominal types presented above.

5. An alternative analysis

The differences observed in 4. can be accounted for by a different analysis:

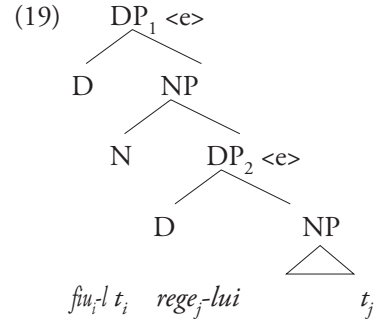
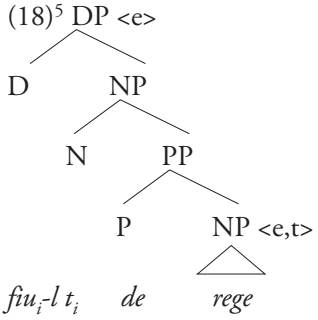
5.1. Morphosyntax

Generalizations

In Romanian, Genitive case can only be marked on the determiner (only the determiner can carry case markings) => *The constructions with morphological case are projections of D(eterminer) (i.e. DPs) taking argument positions.*

Those projections of N that do not have a determiner (i.e. NPs) cannot mark the case morphologically, hence the insertion of the preposition *DE* => *The constructions with DE are NPs taking modifier positions.*

Structures



5.2. Semantics

While in the constructions with morphological case (e.g. 19) the head N denotes a relation between two individuals (the one denoted by DP₁ and the one denoted by DP₂) (see Beyssade & Dobrovie-Sorin 2005), in the prepositional constructions (e.g. 18), the head N denotes a relation between an individual (denoted by DP₁) and a property (denoted by NP₂) (see Kolliakou 1999).

This explains several phenomena. First, why certain prepositional constructions may alternate with an AP (20):

- (20) a. *fii de rege* → *fii regal*
 son-the DE king son-the royal
- b. *ușa de biserică* → *ușa bisericescă*
 door-the DE church door-the church-ADJ

Second, this explains why DPs marked with Genitive case may alternate with personal pronouns (also marked with Genitive case) (21):

- (21) a. *fii regelui* → *fii lui*
 son-the king-the-G son-the him-G
- b. *ușa bisericii* → *ușa ei*
 door-the church-the-G door-the her-G

⁵ The structure proposed in (18) may be conceived of differently with respect to the nature of DE (see Mardale 2005), i.e. the last is not a preposition, but the spell-out of the functional category Mod(ifier) (see Rubin 2002). The arguments in favour of this analysis are the following: (a) DE can not alternate with another preposition (cf. i); (b) DE is excluded when it combines with an argumental PP (cf. ii); (c) DE is obligatory when it combines with an adjoined PP (cf. iii):

- (i) *aerul de munte* vs. **aerul la munte* (ii) **Ion a mers de la munte.*
 air-the DE mountain air-the at mountain John has-AUX walked DE at mountain
 ‘the mountain’s air’ John breathed the mountain’s air’
- (iii) *Ion a respirat aerul de la munte.*
 John has-AUX breath air-the DE at mountain
 ‘John breathed the mountain’s air’

Third, this explains why the complement of *DE* cannot serve as anaphoric antecedents for another DP (22a), while the adnominal constituent marked with morphological case may do so (23):

- (22) a. **El este fiul de [rege]_i pe care_i tînăra speră să îl_i întâlnească.*
 he is son-the DE [king]_i ACC-which_i young-the hopes that him_i meet
 b. *El este [fiul de rege]_i pe care_i tînăra speră să îl_i întâlnească.*
 he is [son-the DE king]_i ACC-which_i young-the hopes that him_i meet
 'He is the king's son that the youngwoman hopes to meet'
- (23) a. *El este fiul regelui pe care tînăra speră să îl întâlnească.* (ambiguous)
 he is son-the king-the ACC-which young-the hopes that him meet
 'He is the son of the king that the youngwoman hopes to meet'
 b. *El este fiul [regelui]_i pe care_i tînăra speră să îl_i întâlnească.*
 he is son-the [king-the-G]_i ACC-which_i young-the hopes that him_i meet
 c. *El este [fiul regelui]_j pe care_i tînăra speră să îl_j întâlnească.*
 he is [son-the king-the-G]_j ACC-which_j young-the hopes that him_j meet

More precisely, the noun *rege* 'king' in (22a) cannot serve as antecedent for the anaphorical pronoun *îl* 'him' because the former is non referential (i.e. it denotes a property). In contrast, the whole DP *fiul de rege* 'the king's son' in (22b) can be the antecedent of the pronoun, because the head *fiul* 'the son' is referential (i.e. it denotes an individual which has a certain property). As for the example in (23a), it is ambiguous. The nouns *fiul* 'the son' and *regelui* 'the king-G' can serve as antecedent for the anaphoric *îl* 'him' because they are both referential (i.e. they denote individuals). As a result, we can obtain two types of readings: (i) the one in (23b) with *regelui* 'king-the-G' being the antecedent of *îl* 'him' and (ii) the one in (23c) with *fiul* 'the son' being the antecedent of *îl* 'him'.

5.3. What about special cases?

Compounds which only allow the synthetic form denote unique entities (such as the sun, the night, the devil etc.), i.e. individuals, hence the Genitive construction (see 16 above).

Others refer to non unique entities (such as corners / mountains, birds, milk etc.), hence the prepositional construction (see 17 above).

6. Conclusion

The two constructions analyzed here are alike insofar as they involve a relation (which may either pertain to the lexical meaning of the head N or else be contextually triggered by the presence of the second argument), but they differ regarding the nature of the second argument: a strong correlation can be shown to exist between syntactic categories (DPs vs. NPs), Case marking (morphological vs. prepositional) and semantic type (type <e> vs. type <e, t>) (see also Dobrovie-Sorin 2001a).

7. References

- Abney, S., 1987, *The English noun phrase in its sentential aspect*, PH dissertation, MIT.
- Bartning, I., 1993, «La préposition *de* et les interprétations possibles des syntagmes nominaux complexes. Essai d'approche cognitive». *Lexique (Les prépositions. Méthodes d'analyse)* 11.163-192.
- Beyssade, C. and C. Dobrovie-Sorin, 2005, «A syntax based analysis of predication». Paper presented at Journées de Sémantique et Modélisation 3, Paris, 17-18 march.
- Corblin, F., 2001, *Défini et génitif: le cas des définis défectifs*. Jean-Marie, Marandin (ed.). Cahier Jean-Claude Milner. 19-54. Verdier, Paris.
- Cornilescu, A., 1993, «Notes on the Structure of Romanian DP and the Assignment of the Genitive Case». Paper presented at the University of Venice Working Papers in Linguistics, Venice.
- Dobrovie-Sorin, C., 2000, «(In)definiteness Spread: from Romanian Genitives to Hebrew Construct State Nominals». V., Motapanyane (ed.). *Comparative Studies in Romanian Syntax*, Benjamins.
- 2001, *De la syntaxe à l'interprétation*, de Milner (1982) à Milner (1995): le génitif. J. M., Marandin (ed.), Cahier Jean-Claude Milner, Verdier, Paris.
- (GLR) *Gramatica limbii române*, 1966, vol. I-II, ediția a doua, București, Editura Academiei.
- Grosu, A., 1998, «On the Distribution of Genitive Phrases in Romanian», *Linguistics* 26.
- Kolliakou, D., 1999, «DE-phrase Extractability and Individual / Property Denotation», *NLLT* 17.713-781.
- Mardale, A., 2005b, «Pourquoi les génitifs dits “analytiques” du roumain ne sont pas de vrais génitifs?». Paper presented at Rencontre des Jeunes Chercheurs, Université Paris 3, 21 May 2005.
- Milner, J.-C., 1982, *Les génitifs adnominaux en français. Ordres et raisons de langues*. 69-140. Paris. Editions du Seuil.
- Munn, A., 1998, «The Possessor that Stayed Close to Home». *Proceedings of WECOL*.
- Rubin, E. J., 2002, «The Structure of Modifiers», electronic ms. www.hum.utah.edu/linguistics/Faculty/rubin.htm

ADJECTIVAL PASSIVES AND ADJECTIVAL DECAUSATIVES IN HEBREW

Aya Meltzer
Tel-Aviv University

Abstract

The distinction between adjectival passives and verbal passives is a very well known one. In this paper, I try to define the operation that forms adjectival passives in Hebrew. I claim that a close look at Hebrew adjectival passives reveals that they do not form a homogenous group, but rather two groups, which behave differently with regard to their interpretation. Adjectives of the first group behave like verbal passives in that they have an implicit Agent in their interpretation; adjectives of the second group behave like unaccusative verbs, in that the external argument of the transitive verb is no longer a part of their semantics. Based on this parallelism, I label the first type of adjectives ('true') *adjectival passives* and the second —*adjectival decausatives*. Having established that there are two types of adjectival passives, I claim that they are derived by the same operations which derive the corresponding verb types. Therefore, no additional operations need to be stipulated in order to account for adjectival passive formation.

1. Introduction

There is, in generative studies, a well-known distinction between adjectival and verbal passives (see, for example, Wasow 1977). Many studies have tried to define the operations that form the two types of passives; but while verbal passive formation seems to be quite understood, there is still debate on the nature of the operation that forms adjectival passives (for a very influential analysis see Levin and Rappaport 1986). In this paper I will try to define this operation for Hebrew. I will first show that there are two classes of adjectival passives in Hebrew; one class behaves on a par with verbal passives, while the other behaves on a par with unaccusative verbs. I will therefore label the two types of adjectives *adjectival passives* and *adjectival decausatives*. I will then argue that the two types of adjectives are formed through the same operations that form the corresponding types of verbs.

The paper is organized as follows: in chapter 2, I will present the main empirical facts concerning the morphology of adjectival passives in Hebrew. In chapters 3 and 4, I show some evidence that there are, in Hebrew, two different types of adjectival passives. I will then discuss the parallelism which I believe exists between the adjective system and the verb system. I will argue that the two classes of adjectival passives correspond to two types of verbs: passives and unaccusatives, and are derived by

the same operations which form these two types of verbs. In chapter 5, I will make a small digression and discuss the verbal system. In particular, I will discuss the operations which, I believe, generate passive and unaccusative verbs. In chapter 6, I will present some data that reinforces the proposal that adjectival passives and adjectival decausatives are derived by the same operations which derive passive and unaccusative verbs, respectively. In chapter 7 I will discuss apparent counter examples to my analysis. Chapter 8 presents a cross-linguistic discussion regarding the phenomenon of adjectival passives and adjectival decausatives.

2. The morphology of adjectival passives in Hebrew

Adjectival passives in Hebrew appear in one of four templates, presented in 1-4:

- (1) *muCCaC*. This template is related to the active template *hiCCiC*. Examples: *mumca* ('invented'), *munax* ('placed, laid'), *mud'ag* ('worried'), *mugaz* ('carbonated'), *muxan* ('prepared, ready'), *mukpa* ('frozen').
- (2) *meCuCaC*. This template is related to the active template *CiCeC*. Examples: *megulgal* ('rolled'), *mevulbal* ('confused'), *mesusal* ('curly'), *meluxlax* ('dirty'), *megulaf* ('engraved, carved'), *mecuyar* ('drawn, sketched, illustrated').
- (3) *niCCaC*. This template is related to the active template *CaCaC*, and is comparatively rare for adjectives. Examples: *nistar* ('hidden, concealed, invisible'), *nirgaz* ('annoyed, angry, furious').
- (4) *CaCuC*. This template is also related to the active template *CaCaC*. Examples: *hafux* ('reversed, inverted, upside down'), *kafu* ('frozen'), *sagur* ('closed'), *katuv* ('written'), *patu'ax* ('open'), *kavuy* ('extinguished'), *afuy* ('baked').

It is important to notice that the first three templates above are also used to derive verbal passives in the present tense. Thus, most of the forms in 1-3 are ambiguous, though the adjectival reading is more accessible. The fourth template, on the other hand, creates only adjectives. This can be seen when inserting the various forms into contexts that clearly demand a verb or an adjective. Such contexts can serve as tests to determine whether a given form is a verb or an adjective (see appendix).

3. The non-uniform behavior of adjectival passives in Hebrew

In this section I will show that adjectival passives in Hebrew do not behave uniformly with regard to the accessibility of the external argument of the transitive verb alternate.

It is well known that passive verbs consistently pass tests which show that their external argument, though not realized in the syntax, is still present in their interpretation. Adjectival passives, on the other hand, behave non-uniformly with respect to such tests: some of them pass the tests, which means that an Agent is present in their semantics, while others fail them, thus lacking an Agent altogether. I will discuss here three such tests: realization of an Instrument θ -role, addition of Agent-oriented adverbs, and cancellation of the Agent entailment.

3.1. Realization of an Instrument θ -role

The first test that detects the existence of an implicit Agent is suggested the Instrument Generalization (Reinhart and Siloni to appear). This generalization states that an argument bearing the Instrument θ -role can only be realized when an Agent is present in the sentence explicitly (mapped to the syntax) or implicitly (inferred).

(5a) is grammatical because there is an Agent realized in the sentence, while (5b) is ungrammatical because there is no Agent, explicit or implicit, in the sentence.

- (5) a. Max ate the soup with a spoon. b. *Max hated the soup with a spoon.

Verbal passives consistently allow the realization of the Instrument θ -role, as can be seen in (6):

- (6) a. The soup was eaten with a spoon.
b. The window was broken with a stone.

Adjectival passives, on the other hand, behave non-uniformly with respect to this test: some of them allow the realization of the instrument θ -role (7), while others disallow it (8):

- (7) a. *ha-mixtav katuv be-et.*
the-letter written in-pen
'The letter is written with a pen.' (adjectival reading)
b. *ha-kelev kašur be-recu'a.*
the-dog tied in-leash
'The dog is tied with a leash.' (adjectival reading)
c. *ha-bayit nā'ul be-mafte'ax.*
the-house locked in-key
'The house is locked with a key.' (adjectival reading)
d. *Max natan li kufsa mudbeket be-devek plasti.*
Max gave to+me box glued in-glue plastic
'Max gave me a box which is glued with plastic glue.'
- (8) a. **ha-kise šavur be-patiš.* c. **ha-yeled xavut be-maklot.*
the-chair broken in-hammer the-child beaten in-sticks
b. **ha-bayit patuax be-mafte'ax.* d. **ha-kufsa dvuka be-devek plasti.*
the-house open in-key the-box glued in-glue plastic

3.2. Use of Agent-oriented adverbs

The second test that detects an implicit Agent has to do with the use of Agent-oriented adverbs: only an Agent, explicit or implicit, can license an Agent-oriented adverb.

(9a) is grammatical because an Agent is realized in the sentence. (9b) is ungrammatical because the Agent role is neither realized, nor inferred:

- (9) a. Max ate the soup on purpose.
b. *The wind opened the door on purpose.

As with the previous test, verbal passives consistently behave as if an external argument is inferred, present in the interpretation:

- (10) a. The soup was eaten on purpose.
 b. The window was broken on purpose.

But, in this case as well, adjectival passives behave non-uniformly. Some license an Agent oriented adverb (11), others do not (12):

- (11) a. *ha-sefer katuv be-kišaron.*
 the-book written in-talent
 ‘The book is written with talent.’
 b. *ha-xulca ha-zot tšura be-xoser mikco’iyut.*
 the-shirt the-this sewn in-lack (of) professionalism
 ‘This shirt is sewn unprofessionally.’
 c. *al ha-kir haya poster mudbak be-rašlanut.*
 on-the-wall there+was poster glued in-carelessness
 ‘There was on the wall a poster which was glued carelessly.’
- (12) a. **ha-bakbuk sagur be-zadon.* b. **ha-poster davuk be-rašlanut.*
 the-bottle closed maliciously the-poster glued in-carelessness

3.3. Cancellation of the entailment of an Agent

Another way to tell whether there is an inferred Agent in a sentence is by a denial of the existence of an Agent. If this denial creates a contradiction, it means that there is, in fact, an inferred Agent in the sentence.

Again, verbal passives behave as if they have an Agent in their interpretation. Trying to deny its existence renders the sentence a contradiction (13).

- (13) *ha-ma’im hupe’u, lamrot še-afexad lo hikpi otam.* (contradiction)
 the water were frozen (verbal reading), though no one froze it

Adjectival passives, on the other hand, behave non-uniformly here as well. With some, the denial of the existence of an Agent creates a contradiction (14), with others—it doesn’t (15).

- (14) *ha-mixtav katuv, lamrot še-afexad lo katav oto.* (contradiction)
 the-letter written, though no one wrote it
- (15) *ha-kufsa ptuxa, lamrot še-afexad lo patax ota.*
 the-box open, though no one opened it

To conclude this section: I have shown that unlike verbal passives, which systematically behave as if they have an implicit Agent, adjectival passives behave non-uniformly. Some of them show the existence of an external argument in their interpretation, while others do not.

4. Definition of the two types of adjectival passives

In the previous section I have shown that adjectival passives behave non-uniformly with regard to the existence of an external argument in their interpretation. Unlike verbal passives, that consistently show the existence of an implicit Agent, the si-

tuation with adjectival passives is more complex. Some of them show the existence of an external argument in the interpretation, while others do not.

It is well known that, in contrast with verbal passives, unaccusative verbs consistently fail tests that detect the existence of an external argument, as can be seen in (16):

- (16) a. *The window broke with a stone.
 b. *The window broke on purpose.
 c. The ship sank, though no one sank it. (not contradictory)

This contrastive behavior of passive and unaccusative verbs parallels the contrastive behavior of the two groups of adjectival passives observed above. This suggests that in fact, the two types of adjectival passives correspond to the two types of verbs: passives and unaccusatives.

If this is indeed the case, one may wonder why it is that both types of adjectives have passive morphology. To answer this question, we can take a look at the Hebrew verbal system. In this system, the correlation between the morphology of a verb and its type (passive, unaccusative, reflexive, etc.) is not completely systematic. For example, the *hitXaXeX* template is used to generate unaccusative, reflexive, reciprocal and even some passive verbs. The *niXXaX* template is used to derive passive, unaccusative, reflexive and reciprocal verbs as well. Therefore, in order to decide whether a verb is passive or unaccusative, we cannot rely on its morphology alone. Rather, we have to determine if it has an external argument in its interpretation or not. If the external argument is still present in the interpretation, the verb is passive. If the external argument is missing, the verb is unaccusative.

I suggest that the same holds for adjectives: what has been taken to be typical passive morphology for adjectives are in fact morphological forms that are not exclusive to passive. The fact that an adjective bears such morphology cannot on its own indicate that it is passive. The decision whether an adjective is passive or not should be based on whether or not it has an external argument in its interpretation. Adjectives that pass tests for the existence of an external argument are “true” *adjectival passives*. These adjectives parallel in their behavior verbal passives. To the adjectives that do not pass these tests, meaning, do not have an external argument at all, I will refer as *adjectival decausatives*. The behavior of these adjectives parallels that of unaccusative verbs.

From now on, I use the term *adjectival passives* in its narrow meaning, that is —adjectives which have an implicit external argument, and not just any adjective that has the so-called passive morphology.

The parallelism between the verbal and the adjectival system can offer a straightforward answer to the question of “adjectival passive formation”: I suggest that adjectival passives and adjectival decausatives are formed through the same operations that form passive and unaccusative verbs, respectively (plus, of course, a category-changing operation). I will now discuss these verb-forming operations briefly.

5. Operations in the verbal system - passive and unaccusative verbs

Both passive verbs and unaccusatives are intransitive verbs, which do not assign Accusative Case and do not realize their external θ -role in its canonical position. The difference between passives and unaccusatives lies in the status of the unrealized ex-

ternal argument. As was shown above, the external argument of passives is accessible, present in their interpretation, while the external argument of unaccusatives is missing altogether. This difference must be accounted for by the difference in the operations that form the two types of verbs.

5.1. Verbal passive formation - Saturation

Verbal passivization in Hebrew takes as input transitive verbs whose external θ -role is Agent or Cause (and perhaps some verbs whose external theta role is Experiencer.)

Passivization does the following: syntactically, it prevents the external argument from being mapped to the subject position, and cancels the verb's ability to assign Accusative Case. Semantically, it performs an existential closure on the external argument (Chierchia 1995, Reinhart 2000, 2002 among others). I will refer to this operation as Saturation: the external argument is saturated. An example is given in (17):

- (17) a. The gangster was murdered.
 b. interpretation: $\exists e \exists x (\text{Murder}(e) \wedge \text{Agent}(e, x) \wedge \text{Theme}(e, \text{the gangster}))$

As can be seen in (17), there is an Agent present in the interpretation of a passive sentence. Therefore, passive verbs allow the realization of the Instrument θ -role and the addition of Agent-oriented adverbs, and the denial of the Agent creates a contradiction.

5.2 Unaccusative verb formation - De-causatization

Unaccusative verbs, like passive verbs, are derived from their transitive alternates (Chierchia 1989, Levin & Rappaport 1994, Reinhart 2000, 2002). I will assume here the operation presented in Reinhart (2000, 2002), which I will refer to as De-causatization.

De-causatization takes as input transitive verbs whose external θ -role is Cause. These are verbs like *break*, *open*, etc., whose external role can be realized either as an Agent or as an inanimate Cause. What de-causatization does is to reduce this role: $V(\theta_1(\text{cause}), \theta_2) \rightarrow V(\theta_2)$. Here, the θ -role is not merely saturated, but totally reduced. Therefore, such verbs will not allow the realization of the Instrument θ -role, the addition of Agent-oriented adverbs, etc.

6. Reinforcement of the analysis

In chapter 2, I suggested that adjectival passives and adjectival decausatives are derived by the same operations that derive the corresponding verbs, namely, Saturation and De-causatization, respectively. This analysis has a strong prediction regarding the existence/non-existence of certain adjectival forms, which is borne out. This fact reinforces the analysis suggested here.

According to the input that Saturation and De-causatization take, it is obvious that verbs whose external θ -role is Agent will undergo passivization, but will not undergo De-causatization, and therefore will have a verbal passive alternate, but not an unaccusative one. Verbs whose external θ -role is Cause will undergo both opera-

tions and have both corresponding verb types. This is indeed the case in the verbal system, as shown in (18) and (19):

- (18) a. Max/ *The paint painted the picture.
 b. The picture was painted. c. *The picture painted.
- (19) a. Max / A gust of wind opened the door.
 b. The door was opened. c. The door opened.

If indeed adjectival passives and adjectival decausatives are derived by the same operations, the prediction is that the situation should be the same in the adjectival system. I will now show that this prediction is borne out, by showing the following:

a. Transitive verbs whose external θ -role is Agent have adjectival passive alternates, but no adjectival decausative alternates.

b. Transitive verbs whose external θ -role is Cause have both adjectival passive and adjectival decausative alternates.

6.1. Adjectival forms of transitive verbs whose external θ -role is Agent

Verbs like *katav* ('write'), *kašar* ('tie'), *šamar* ('guard'), *nigev* ('wipe dry'), *hidpis* ('type'), *talaš* ('tear off, tear out'), *cilem* ('photograph'), etc., whose external θ -role is Agent, are predicted to undergo Saturation and have an adjectival passive alternate, but to not have an adjectival decausative alternate.

The prediction is borne out: the adjectives derived from these verbs show the existence of an external argument (with some exceptions that will be dealt with in chapter 5):

- (20) a. *hamixtav katuv be-et / be-kišaron.*
 the-letter written in-pen / in-talent
 'The letter is written with a pen / with talent.'
- b. *ha-kelev kašur be-recu'a.* c. *ha-ictadion šamur bi-kfida.*
 the-dog tied in-leash the-stadium guarded impeccably
 'The dog is tied with a leash.' 'The stadium is carefully guarded.'
- d. *Max natan li daf mudpas be-rašlanut / be-mexonat ktiva.*
 Max gave to+me paper typed in-carelessness / in-typewriter
 'Max gave me a paper which is typed carelessly / with a typewriter.'
- e. *ha-mixtav katuv, lamrot še-af exad lo katav oto.* (contradiction)
 the-letter written, though no one wrote it

Therefore, the adjectival forms of such verbs are passive. In addition, these verbs do not have another adjectival counterpart which is decausative.

6.2. Adjectival forms of transitive verbs whose external θ -role is Cause

Verbs like *hikpi* ('freeze'), *nipe'ax* ('inflate, blow up'), *sibex* ('complicate'), *pizer* ('scatter'), *kicer* ('shorten'), *ximem* ('heat'), *saraf* ('burn'), *šavar* ('break'), etc., whose external θ -role is Cause, are predicted to undergo both Saturation and De-causativization, and have both an adjectival passive and an adjectival decausative alternate.

I believe that this prediction is borne out as well, in one of four ways (i)-(iv):

(i) Some verbs have two morphologically distinct adjectival alternates —one passive, the other decausative. Some examples are given in (21):

(21)	<i>verb</i> ($\theta 1 =$ Cause)	<i>adjectival passive</i>	<i>adjectival decausative</i>
	<i>hikpi</i> 'freeze'	<i>mukpa</i> 'frozen'	<i>kafu</i> 'frozen'
	<i>nipe'ax</i> 'inflate, blow up'	<i>menupax</i> 'inflated, blown up'	<i>nafu'ax</i> 'swollen, inflated'
	<i>pina</i> 'clear off, vacate'	<i>mefune</i> 'vacated, evacuated'	<i>panuy</i> 'vacant, empty'
	<i>hidbik</i> 'glue, attach'	<i>mudbak</i> 'glued, attached'	<i>davuk</i> 'attached'
	<i>hevix</i> 'embarrass'	<i>muvox</i> 'embarrassed'	<i>navox</i> 'embarrassed'

The adjectives in the second column show accessibility of the external argument. The ones in the third one do not:

- (22) a. **ha-kufsa dvuka be-devek plasti.*
the-box glued in-gluce plastic
b. *Max natan li kufsa mudbeket be-devek plasti.*
Max gave to+me box glued in-gluce plastic
'Max gave me a box which is glued with plastic glue.'
- (23) a. **ha-rikma kfu'a be-xankan nozli.*
the-tissue frozen in-nitrogen liquid
b. *bet ha-xolim kibel mislo'ax šel rekamot mukpa'ot be-xankan nozli.*
the hospital received shipment of tissues frozen in-nitrogen liquid
- (24) a. **ha-poster davuk be-rašlanut.*
the-poster glued in-carelessness
b. *yeš al ha-kir poster mudbak be-rašlanut.*
there is on the wall poster glued in-carelessness
- (25) a. **kibalti kadur nafu'ax be-maš'evat gumi.*
I-received ball inflated in-pump rubber
b. ?*kibalti kadur menupax be-maš'evat gumi.*
I-received ball inflated in-pump rubber
'I received a ball which was inflated with a rubber pump.'
- (26) a. *ha-giv'a ha-zo pnuya, lamrot še-afexad / šum davar lo pina ota.*
the-hill the-this vacant, though that-no one / nothing evacuated it
b. *ha-giv'a ha-zo mefuna, lamrot še-afexad / šum davar lo pina ota* (contradiction)
the-hill the-this vacated, though that-no one /nothing evacuated it
- (27) a. *Max navox, lamrot še-šum davar / afexad lo hevix oto.*
Max embarrassed, though that-nothing / no one embarrassed him
b. *Max muvox, lamrot še-šum davar / afexad lo hevix oto.* (contradiction)
Max embarrassed, though that-nothing / no one embarrassed him

This shows that the adjectives in the second column are passive, the ones in the third column are decausative.

(ii) Some verbs have two adjectival alternates —one decausative, and the other ambiguous between passive and decausative. Some examples are given in (28):

(28)	<i>transitive verb</i>	<i>ambiguous form</i>	<i>decausative form</i>
	<i>sibex</i> ‘complicate’	<i>mesubax</i> ‘complicated’	<i>savux</i> ‘complicated’
	<i>pizer</i> ‘scatter’	<i>mefuzar</i> ‘scattered’	<i>pazur</i> ‘scattered’
	<i>ikem</i> ‘bend, twist’	<i>me’ukam</i> ‘bent, twisted’	<i>akum</i> ‘crooked, twisted, bent’

The adjectives in the second column show accessibility of the external argument, while those in the third do not:

- (29) a. *?ha-sukar al ha-uga haya mefuzar be-nedivut.*
the-sugar on the-cake was scattered in-generosity
b. **ha-sukar al ha-uga haya pazur be-nedivut.*
the-sugar on the-cake was scattered in-generosity
- (30) a. *mot ha-barzel nir’e me’ukam be-ko’ax.*
pole the-iron seems bent in-power
‘The iron pole seems forcefully bent.’
b. **mot ha-barzel nir’e akum be-ko’ax.*
pole the-iron seems bent in-power

But, both forms do not entail the existence of an Agent:

- (31) *ha-alim mefuzarim / pzurim po, lamrot še-af exad / šum davar lo pizer otam.*
the-leaves scattered here, although that-no one / nothing scattered them
- (32) *ha-anaf ha-ze me’ukam / akum, lamrot še-af exad / šum davar lo ikem oto.*
the-branch the-this bent, although that-no one / nothing bent it

So, the forms of the second column can behave either as passives (showing accessibility of the external argument) or as decausatives (not entailing the existence of an Agent). Therefore I suggest that they are ambiguous. The forms in the third column are unambiguously decausative.

(iii) Some verbs have two adjectival alternates —one passive, with so-called passive morphology, the other decausative, without such morphology. Some examples are given in (33):

(33)	<i>transitive verb</i>	<i>adjectival passive</i>	<i>adjectival decausative</i>
	<i>kicer</i> ‘shorten’	<i>mekucar</i> ‘shortened’	<i>kacar</i> ‘short’
	<i>ximem</i> ‘heat’	<i>mexumam</i> ‘heated’	<i>xam</i> ‘hot’
	<i>kerer</i> ‘cool’	<i>mekurar</i> ‘cooled’	<i>kar</i> ‘cold’

The adjectives in the third column, though not bearing the so-called passive morphology, share the other properties with the adjectival decausatives discussed so far: they have a transitive alternate whose external θ -role is Cause, but this θ -role seems to have been totally eliminated during the derivation. The adjectives in the second column are passive —they have an external argument in the semantics.

(iv) Some verbs, like *saraf* (‘burn’), *šavar* (‘break’), *sagar* (‘close’), *patax* (‘open’), *gilgel* (‘roll’), *lixlex* (‘dirty, sully’), *kilkel* (‘damage, spoil’), *nipec* (‘smash’) have only one corresponding adjectival form. This form seems at first sight to behave like a decausative: in its most natural interpretation it does not entail the existence of an

Agent (34), and it does not readily allow the realization of an Instrument role, or an Agent-oriented adverb (35):

- (34) a. *ha-kufsa sgura, lamrot še- af exad lo sagar ota.* (not contradictory)
 the-box closed, though no one closed it
 b. *ha-tanur mekulkal, lamrot af exad lo kilkel oto.* (not contradictory)
 the-over broken (out of order), though no one damaged it
- (35) a. **ha-delet sgura be-mafte'ax.* b. **ha-kise šavur be-ko'ax.*
 the-door closed in-key the-chair broken in-strength

But there are some examples which seem to show that even in this case, the external argument can be traced:

- (36) *ha-xalonot sgurim be-rašlanut.*
 the-windows closed in-carelessness
- (37) *Max me'ašen sigaria megulgelet be-meyumanut.*
 max is smoking a cigarette rolled in-skill

Theoretically, there are two possible ways to analyze this case: either, for some reason, these verbs only have an adjectival decausative alternate, and not a passive one; or —these adjectival forms are ambiguous between a passive and a decausative reading, and for some reason do not pass the tests detecting the existence of an external argument.

The second analysis is much more appealing, since it maintains uniformity in the group of verbs whose external θ -role is Cause (namely, that all of them can undergo both Saturation and De-causativization). Notice that in the verbal system as well some of these verbs have one morphological form which is ambiguous between a passive and an unaccusative reading (*nišbar*, 'was broken, broke'; *nisgar*, 'was closed, closed'). This analysis is also tenable because there is an independent explanation for the ungrammaticality of many of the sentences such as (35), in which these adjectives seem not to allow addition of an Instrument θ -role, or Agent-oriented adverbs (see chapter 7).

A very good argument in favor of these forms being ambiguous would be if there was no other option —if there were morphological reasons why there can't be two different forms. I believe that this is the case here. From the last sections we can draw some conclusions about the morphology of the adjectives I am discussing: an adjectival passive of a verb is in the passive template related to the active verb's template. An adjectival decausative is generally in the *CaCuC* template (or in non-passive morphology). Now let us look at the verbs listed in the beginning of this section. Some of them are in the *CaCaC* template. There are two passive templates that correlate to this template: *niCCaC* and *CaCuC*. I mentioned before already that for some reason, the *niCCaC* template is in general very rare for adjectives. Therefore, verbs in the *CaCaC* template are predicted to have an adjectival passive alternate in the remaining related template: *CaCuC*. But since this is also the general template for adjectival decausatives, such forms will be ambiguous between passive and decausative.

The rest of the verbs mentioned in the beginning of the section (with one exception - *nipec* 'smash') are verbs with four consonants in the root. Their verbal passive alternate will be in the predicted form, in the passive template related to the template

in which they appear (*CiCeC*). But their decausative alternate cannot be in the predicted *CaCuC* template, because the paradigm of this template cannot “host” four-consonantal roots. So, the passive form is used to express the decausative meaning as well. I still cannot explain why the decausative reading of the adjectives of group (iv) is so strong that it almost ‘overrides’ the passive one.

The conclusion of the last section is that verbs of type (a) above ($\theta 1 = \text{Agent}$) have only an adjectival passive alternate, while the verbs of type (b) ($\theta 1 = \text{Cause}$) seem to have two adjectival alternates: one passive and one decausative. These facts reinforce the analysis that the two types of adjectives are derived through the same operations that derive verbs. They show that the adjectival system parallels the verbal system with regard to the input and the output of the operations.

7. Explanation of the Counter Examples

One prediction that seems to have many counter examples is that every verb whose external theta-role is Agent or Cause will be able to undergo Saturation, and therefore that the resulting adjective will behave as if it has an external argument in its interpretation. Consider for example (38):

- (38) a. **ha-kise* *šavur* *be-patiš* / *be-ko'ax*.
 the-chair broken in-hammer / in-force
 b. **ha-yeled* *muke* *be-maklot*. c. **ha-delet* *sgura* *be-zadon*.
 the-child beaten in-sticks the-door closed in-evil

If the adjectives in (38) are ‘true’ adjectival passives, with an implicit Agent, why are the sentences ungrammatical?

When we modify a verb with an Instrument argument, or with an adverb, we modify the event. But adjectives do not describe events, they describe states, and they lack an event variable. Therefore, an Instrument role or an adverb that we add must relate also to the state, and not only to the event that led to it. So the Instrument or the adverbial description must still be relevant, in a way ‘visible’, in the state. Consider (39) and (40):

- (39) *ha-kelev kašur* *be-recu'a*. (40) **ha-yeled* *muke* *be-maklot*.
 the-dog tied in-leash The-child bitten in-sticks

When we see a tied dog, we also see what it is tied with. On the other hand, if we see a boy which was hit, we can perhaps only guess what he was hit with, but the Instrument is no longer ‘visible’ and it is not a part of the state. Consider next (41) and (42):

- (41) **ha-mixtav* *katuv* *be-et* *yafa*.
 the-letter written in-pen beautiful
 ‘The letter is written with a beautiful pen.’
 (42) *ha-mixtav* *katuv* *be-et* *šxora*.
 the-letter written in-pen black
 ‘The letter is written with a black pen.’ (Julia Horvath p.c.)

(41) is ungrammatical because the pen being beautiful cannot be detected from looking at the written letter. (42), on the other hand, is grammatical, but we interpret it in a very specific way: the sentence claims that the ink in the pen is black, not that the pen itself is black. The reason is the same as in the previous examples: the pen itself being black is not detectable from the resulting state. But, the ink in the pen being black is detectable from the written letter, and therefore the addition of an Instrument role is grammatical, and this is the interpretation that we assign to the sentence. The same is true for Agent-oriented adverbs:

- (43) *ha-poster mudbak be-rašlanut.* (44) **ha-delet sgura be-zadon.*
 the-poster glued in-carelessness the-door closed in-evil

(43) is fine, because the adverb is still relevant to the state. By looking at a glued poster we can tell if it has been glued carelessly, maybe because it is glued unevenly, has loose ends, etc. On the other hand, when we look at a closed door, we cannot tell if it was closed with good or bad intentions.

To conclude this section:

- An argument bearing the Instrument θ -role can only be realized when an Agent is present in the sentence, explicitly or implicitly, and when the instrument is detectable from the state.
- An Agent-oriented adverb can only be realized when an Agent is present in the sentence, explicitly or implicitly, and when the adverbial description is detectable from the state.

8. Adjectival passives and adjectival decausatives - a cross-linguistic perspective

Having established the fact that there are two distinct types of adjectival passives in Hebrew, a natural question arises: is this phenomenon unique to Hebrew, or does it exist in other languages as well? Theoretically, there is no a priori reason why these two types of adjectives should not exist in other languages. Given the analysis presented here, the two types of adjectives are derived through Saturation and Decausativization: the operations that form passive and unaccusative verbs. It is very well known that passive and unaccusative verbs exist in many languages, meaning that these two operations are operative in the verbal system of many languages. Unless there is some feature of the adjectival system which prevents these operations (or one of them) from applying in it, the prediction is that Saturation and Decausativization will derive adjectives as well.

8.1. Hungarian

As was shown in chapter 6, the distinction between adjectival passives and adjectival decausatives in Hebrew is very clear in some cases, since they are realized through two morphologically distinct forms. Another language which marks morphologically the two types of adjectives is Hungarian. Some examples are given in (45) (Horvath and Siloni to appear).

(45) <i>Transitive Verb</i>	<i>Adjectival Passive</i>	<i>Adjectival Decausative</i>
<i>olvaszt</i> 'melt'	<i>olvaszt-ott</i> 'melted'	<i>olvad-t</i> 'melted'
<i>kinyit</i> 'open up'	<i>kinyit-ott</i> 'opened up'	<i>kinyil-t</i> 'opened up'
<i>fagyaszt</i> 'freeze'	<i>fagyasztott-ott</i> 'frozen'	<i>fagy-ott</i> 'frozen'
<i>megrongál</i> 'damage'	<i>megrongál-t</i> 'damaged'	<i>megrongálód-ott</i> 'damaged'

As can be seen from the following noun phrases, the forms of the second column allow addition of Agent-oriented adverbs and Instruments, while those in the third do not:

- (46) a. *a szándékosan befagyasztott tó*
 the intentionally in-freeze-caus.-adj.part. pond
 'the intentionally frozen pond'
- b. *a (*szándékosan) befagyott tó*
 the intentionally in-freeze-adj.part. pond
- (47) a. *a késsel megrongált asztal*
 the knife-with perf.-damage.trans.-adj.part. table
 'the damaged with a knife table'
- b. *a (*késsel) megrongálódott asztal*
 the knife-with perf.-damage-unacc.-adj.part

The Hungarian data is easily predicted and explained by the analysis presented here. Notice that all the verbs in (45) have as their external θ -role the Cause role, and are therefore predicted to have two corresponding adjectival forms. The data in fact reinforces the proposed analysis: the forms which I labeled adjectival decausatives are very similar to the forms of the corresponding unaccusative verbs, both containing identical morphemes; for example, compare the forms *olvad* 'melt (unaccusative)', and *olvadt* 'melted (adjectival decausative)'. The shared morphemes may indicate that both forms shared some operation in their derivation, namely De-causatization.

Hungarian, then, systematically derives both adjectival passives and adjectival decausatives using different morphology. I have shown that in Hebrew the situation is more complex: sometimes there are two different forms for the two types of adjectives, and sometimes one form is ambiguous between the two readings. This indicates a theoretical option for morphologically poor languages: both adjectival passives and adjectival decausatives exist in such languages, but both types of adjectives have an identical form. What I would like to show now is that this is the case with English.

8.2. English

Embick (2004) presents evidence that in English there are two types of adjectival passives, which he labels 'statives' and 'resultatives'. In many cases, the two types are identical in form; this is the case with *closed*, *broken* and *bent*, for example. In other cases, the two types have different forms; examples are *open* (stative) - *opened* (resultative), *rotten* - *rotted*, *shaven* - *shaved* and more. Embick uses several tests that distinguish between the two types

of adjectives. The one relevant to the current discussion has to do with adverbial modification —resultatives, but not statives, allow modification by manner (and other) adverbials:

- (48) a. The package remained carefully opened.
 b. *The package remained carefully open.

Notice that in both cases the form in question appears as a complement of *remained*, which is a context that allows only adjectives. Therefore, both forms are adjectival.

My suggestion is that the adjectives which Embick labels statives are adjectival decausatives, and those he labels resultatives are adjectival passives. This is a natural conclusion from the diagnostics presented in (48), which Embick uses to distinguish between the two types of adjectives. It is identical to the test presented in chapter 2 to detect the presence of an implicit Agent.

So, English data suggest that in English as well there are two types of adjectival passives: ‘true’ adjectival passives and adjectival decausatives. The fact that the two types of adjectives often have the same morphology can obscure the distinction, but a close look at the behavior and interpretation of these adjectives reveals it.

9. Conclusion

I began by showing that the group of Hebrew adjectives which is usually referred to as adjectival passives actually consists of two groups: one type of adjectives behaves as if they lack an external argument altogether; the other type behaves as if an external argument is present in their interpretation.

Based on a comparison with the verbal system, I called the first type *adjectival decausatives*, and the second one —*adjectival passives*.

I proposed that the operations that form these adjectives are the same as the operations that form unaccusative and passive verbs, but also involve category change, from verb to adjective. Thus, decausative adjectives are formed through De-causativization: total reduction of the external argument of the transitive verb. Passive adjectives are formed through Saturation: an existential closure upon the external argument of the transitive verb.

I believe that this analysis is better than former attempts to define adjectival passive formation because of two reasons: first, it explains and predicts more empirical data, especially concerning the non-uniform behavior of these adjectives with regard to the presence of an external argument. Second, the analysis makes use of known and established operations to explain a new set of data, without stipulating new processes. In fact, given that we accept the difference between passive and unaccusative verbs, and the need for two distinct operations to derive these two types of verbs, an additional stipulation would be required to prevent both operations from operating in the adjectival system as well.

Appendix - the distinction between verbal and adjectival passives in Hebrew

- (1) Contexts which allow verbs and do not allow adjectives:
 a. Simple inversion (predicate-subject order: possible with some verbs, not possible at all with adjectives).
 b. Modification by an event modifier.

- (2) Contexts which allow adjectives and do not allow verbs:
 a. Post nominal position.
 b. Following the copula in the future tense.

Sentences (3-4) show that the form *mumca* ('invented') is ambiguous between a verb and an adjective —it can appear in both types of contexts:

- (3) a. *mumca'im xamiša patentin be-yom ba-maxon ha-ze.*
 (are) invented five patents in-day in-the-institution the-this
 'Five patents are invented each day in this institution.'
 b. *sisma'ot xadašot mumca'ot pa'amayim be-šavu'a.*
 passwords new (are) invented twice in-week
 'New passwords are invented twice a week.'
- (4) a. *ha-iton ha-ze lo mefarsem uvdot mumca'ot.*
 the-paper the-this not publish facts invented
 'This paper doesn't publish invented (made-up) facts.'
 b. *yeš li hargāša še-hateruc šelo yihye mumca.*
 there is to+me feeling that-the-excuse his will+be invented
 'I have a feeling that his excuse will be a fabrication.'

The sentences in (5) show that *hafux* ('inside-out, inverted') is an adjective:

- (5) a. **hafuxot xameš xulcot ba-megera ha-zot.*
 inverted (inside-out) five shirts in-the-drawer the-this
 b. **ha-xulcot ha-ele hafuxot pa'amayim be-šavu'a.*
 the-shirts the-these inverted (inside-out) twice in-week
 c. *Max tamid holex im xulca hafuxa.*
 Max always walks with shirt inside-out
 d. *maxar ha-xulca šel Max tihiye hafuxa.*
 tomorrow the-shirt of Max will+be inside-out

References

- Burzio, L., 1986, *Italian Syntax. A Government-Binding Approach*. Reidel, Dordrecht.
- Chierchia, G., 1989, «A Semantics for Unaccusatives and its Syntactic Consequences», in A. Alexiadou, E. Anagnostopoulou and M. Everaert (eds.), *The Unaccusativity Puzzle: Studies on the Syntax-Lexicon Interface*, 60-84. Oxford U. P., 2004.
- Chierchia, G., 1995, «The Variability of Impersonal Subjects», in E. Bach, E. Jelinek, K. Kratzer and B. Partee (eds.), *Quantification in Natural Language*, 107-143. Kluwer, Dordrecht.
- Cinque, G., 1990, «Ergative Adjectives and the Lexicalist Hypothesis», *NLLT* 8. 1-39.
- Doron, E., 2000, «Habeynoni Hasavil» («The Passive Participle»), *Balshanut Ivrit* ('Hebrew Linguistics'), 47. 39-62.
- 2003, «Agency and Voice: the Semantics of the Semitic Templates», *NLLT* 11. 1-67.
- Dubinsky, S. and R. Simango, 1996, «Passive and Stative in Chichewa: Evidence for Modular Distinctions in Grammar», *Lg*, 72. 749-781.
- Embick, D., 2004, «On the Structure of Resultative Participles in English», *LI* 35:3. 355-392.

- Fox, D. and Y. Grodzinsky, 1998, «Children's Passive: a View from the *By*-phrase», *LI* 29:2. 311-332.
- Friedemann, M. A., 2000, «Worry is not Preoccupier». Handout, Utrecht Discussion Group.
- Grimshaw, J., 1990, *Argument Structure*, MIT Press, Cambridge, Mass.
- Horvath, J. and T. Siloni, 2003, «Against the Little-*v* Hypothesis», *Proceedings of IATL* 19.
- and T. Siloni, to appear, «Adjectival Passives: Active Lexicon».
- Kratzer, A., 2000, «Building Statives», *Berkeley Linguistic Society* 26.
- Landau, I., 2002, «A Typology of Psych Passives». In Hirotani, M. (ed.), *Proceedings of the 32nd Conference of the North Eastern Linguistic Society*, GLSA, UMASS, Amherst. 271-286.
- Levin, B. and M. Rappaport, 1986, «The Formation of Adjectival Passives», *LI* 17. 623-661.
- Marelj, M., 2004, *Middles*, Ph.D. diss., Utrecht Institute of Linguistics, OTS.
- Pesetsky, D. M., 1995, *Zero Syntax: Experiencer and Cascades*, MIT Press, Cambridge Mass.
- Pustejovsky, J., 1991, «The Syntax of Event Structure», *Cognition* 41, 47-81.
- Reinhart, T., 1996, «Syntactic Effects of Lexical Operations: Reflexives and Unaccusatives», *OTS Working Papers*.
- 2000, «The Theta System: Syntactic Realization of Verbal Concepts», *OTS Working Papers in Linguistics*, 00,01/TL.
- 2002, «The Theta System —an Overview», *Theoretical Linguistics* 28:3, 229-290.
- and T. Siloni, 2005, «The Lexicon-Syntax Parameter: Reflexivization and Other Arity Operations», *LI* 36:3, 389-436.
- Siloni, T., 2002, «Active Lexicon», *Theoretical Linguistics* 28:3, 383-400.
- to appear, «Al Binyan Hitpa'el» (In Hebrew).
- Wasow, T., 1977, «Transformations and the lexicon». In P. Culicover, T. Wasow and A. Ak-majian (eds.). *Formal syntax*, 327-360. New York: Academic Press.

ARGUMENT STRUCTURE AND MORPHOLOGY: THE CASE OF *EN*- PREFIXATION REVISITED*

Susanna Padrosa Trias

Universitat Autònoma de Barcelona

Abstract

In this paper I argue that *en*-prefixed words in Catalan and English (e.g. *amor*_N ‘love’ > [[*en+amor*]_{V+ar}]_V ‘to make someone fall in love’; *noble*_A > [*en+noble*]_V) are not exceptions to the Right-hand Head Rule (RHR; Williams 1981a). I argue that a \emptyset -suffix, and not the prefix *en-*, is responsible for the conversion of adjectives and nouns to verbs (Neeleman & Schipper 1992). The θ -grid of N/A-to-V prefixations provides the empirical evidence in favour of the conversion-suffix. The \emptyset -suffix will be responsible for the presence of a [+c] role and the prefix will account for the [-c-m] features sometimes present in denominal verbs. I will also show that an unaccusative approach (Grimshaw 1990, Sportiche 1998) to reflexives (in Romance) can deal with the data more satisfactorily than an unergative one (Reinhart & Siloni 1999). Finally, a syntactic theory of argument structure (cf. Hale & Keyser 1993, 1998, 2002) will prove not to be sufficient to account for the data.

1. Introduction¹

The present study deals with the derivation of words and the consequences that word formation processes have on the argument structure of the base. A current topic in generative grammar is whether word-structure is built by the laws of the syntax (cf. Baker 1988, Marantz 1997, 2001, Hale & Keyser (henceforth HK) 1993, 1998, 2002, Mateu 2001a/b, 2002, 2005) or by the laws of the morphological component (cf. Williams 1981a, 1981b, Selkirk 1982, Di Sciullo & Williams 1987, Di Sciullo 1997, Kiparsky 1997, Varela & Haouet 2001, Williams 2004, Lieber 2004). Here I will adopt a modular theory of grammar that brings together the two different views. I will assume that morphology constitutes a component on its own that interacts with the other

¹ I am indebted to Anna Bartra, Montse Capdevila, Anna Espunya, Mireia Llinàs, Jaume Mateu, Ad Neeleman, and Maria Josep Solé for their comments, and valuable suggestions on earlier versions of this paper. I am also grateful to the audience at the 2004 CamLing Second Postgraduate Conference and 2005 BIDE Student Conference in Linguistics for their questions and suggestions. Needless to say, all remaining errors are my own. Research for this paper has been partially funded by the Generalitat de Catalunya (2003BBR-00003) and by the MEC and FEDER (BFF2003-08364-C02-01).

This article is a development of Padrosa's (2005a) study of Catalan *en-* prefixation and draws most of its content from an unpublished MA dissertation (Padrosa 2005b).

components of grammar, i.e. syntax, semantics and phonology (cf. Jackendoff 1997, Ackema & Neeleman 2004) and that there is some regularity behind the lexical items in any language, a belief that goes back to Gruber (1965), Chomsky (1970), Halle (1973), Jackendoff (1975), among others.

Headedness in morphology is regular. For instance, affixation processes in English (e.g. Williams 1981a) and Catalan (e.g. Mascaró 1986) are typically right-headed. Since these will be the languages under analysis, Williams' (1981a) Right-hand Head Rule (RHR) becomes relevant. The RHR² states that the head of a morphologically complex word is rightmost. The head will assign its category to the entire word by means of a mechanism referred to as percolation (see section 2.3 for discussion of such a mechanism). A direct result of the RHR is that suffixes, but not prefixes, are expected to determine the category of the word they attach to, since the head determines the properties of the whole. The following examples show that suffixes (1a, 2a) are typically category-changing and prefixes (1b, 2b) category-neutral. (1) and (2) illustrate the point for English and Catalan respectively.

- (1) a. mad_A+ness_N = madness_N b. re+write_V = rewrite_V
 character_N+ize_V = characterize_V im+polite_A = impolite_A
- (2) a. groc_A 'yellow' + or_N = grogor_N 'yellowness/having the quality of yellow'
 industrial_A 'industrial' + itzar_V = industrialitzar_V 'industrialize'
 b. a + dormir_V 'to sleep' = adormir_V 'to make somebody fall asleep'
 anti + higiènic_A 'hygienic' = antihigiènic_A 'antihygienic'

Although the RHR seems to apply quite consistently, there are some exceptions to the claim that the head in morphological constructions is on the right and these need to be accounted for. For example, Williams (1981a) observes that the English prefix *en-* systematically converts N(ouns) and A(djectives) into V(erbs), thus displaying the behaviour of a head:

- (3) rage_N > [en+rage]_V dear_A > [en+dear]_V
 case_N > [en+case]_V noble_A > [en+noble]_V

A similar scenario exists in Catalan. The prefix *en-* also seems to convert Ns and As into Vs in a productive way:³

- (4) amor_N 'love' > [[en+amor]_V+ar]_V 'to make someone fall in love'
 caixa_N 'box' > [[en+caixa]_V+ar]_V 'to put (something) in boxes'
 car_A 'expensive' > [[en+car]_V+ir]_V 'to raise the price (of something)'
 cendrós_A 'ashy' > [[en+cendrós]_V+ar]_V 'to cover something with ashes'

² Selkirk (1982) points out that the RHR is not universal and notes (citing from Lieber 1980) that left-headed types predominate in Vietnamese, for example. The RHR must therefore be stated as part of the grammar of Catalan and English, a parameter set for those languages with right-headed morphology.

³ Of the alleged category-changing prefixes in Catalan (cf. Cabré & Rigau 1986, Cabré 1988, 1994: *a-*, *en-* (*em-*), *re-*, *des-* (*es-*)) and English (cf. Siegel 1979, Williams 1981a, Selkirk 1982: *a-*, *be-*, *de-*, *en-* (*em-*)), *en-* has been chosen to be the most productive one in the two languages.

In front of these counterexamples to the RHR, one is faced with different alternatives to explain them.⁴ The first one is to say that these words have no head. However, this is not a very attractive option since all complex words seem to have a head. The notion of head, which plays an important role in syntax, can also be applied to the internal structure of words. Work on heads in morphology has been well-established for a long time (cf. Williams 1981a, Selkirk 1982, Scalise 1984, 1988a/b, Di Sciullo & Williams 1987).

A second option is to ascribe the prefix *en-* the attribute of a head and assign it to the category V. For instance, Williams (1981a) provides two arguments to support this view for English, the first of which is that it accounts for the systematic assignment of *en-X* words to the category V. The second argument is that *en-* potentiates the affix *-ment*, as seen in *ennoblement*, *enragement*, and *endearment*. As is usually observed in morphology studies, the potentiation of affix_X by affix_Y indicates that the latter must be in the head position. In this sense, it seems plausible to say that *en-X* words have leftmost heads.

A third alternative to deal with the counterexamples to the RHR is not to treat them as exceptions, which is the view defended by Neeleman & Schipper (1992) when dealing with apparent category-changing verbal prefixation in Dutch. The authors argue that prior to prefixation there is a conversion process of As and Ns to Vs, by means of a zero-affix. Some evidence for this conversion-analysis comes from the argument structure of Vs, assuming that the Θ -grid of a complex word is derived from the thematic information of its morphemes via Θ -role percolation. The Dutch prefix *ver-* provides a Theme when it attaches to a V. That becomes clear if the V *dobbelen* (5a), which takes an Agent, is contrasted with the prefixed version of the same V (5b), which takes an Agent and a Theme. However, when *ver-* is attached to a N/A, there is a Theme (which in this case originates in the A due to the Rel(ativized) RHR), and an optional Agent which cannot have originated in the prefix (see (6)), assuming that the prefix *ver-* provides a stable Θ -role. In (5) it was established that the prefix provides a Theme, although its features are sometimes not visible, i.e. when the base on its right has the same features, as seen in (6). Another source for the Agent has to be found. Hence, the postulation of the conversion suffix.

- (5) a. *dobbelen*_V 'to gamble' Agent
 b. *verdobbelen*_V 'to gamble away' Agent Theme
- (6) a. *nieuw*_A 'new' Theme
 b. *vernieuwen*_V 'to renew' Agent Theme

⁴ I discard the possibility that in Catalan the final suffix is responsible for the category change, since this suffix is part of the inflectional paradigm and inflectional elements do not change category. This option is considered and rejected in Padrosa (2005b), who gives an overview of the different analyses proposed in the literature to account for parasynthetic constructions (e.g. *ennegrir* 'to blacken', *embolden*). The overview includes the three alternatives proposed in the present study plus others, leaving the zero-conversion suffix as the only possible option. In addition, the same zero-suffix can explain the many cases of conversion from a N or A to a V without a prefix in Catalan (i) and English (ii).

(i) a. *salN* 'salt' - *salarV* 'to salt'
 b. *arrelN* 'root' - *arrelarV* 'to root'

(ii) a. *saltN* - to *saltV*
 b. *rootN* - to *rootV*

Reinhart's (2000, 2001) assumptions go well with the modular approach to grammar adopted here, and by adopting her theta system and a Θ -role percolation approach to the inheritance of thematic information (Gràcia 1992, 1995, Neeleman and Schipper 1992), I will try to find out which of the two last alternatives (i.e. *en*-prefixations having leftmost heads and having a zero-suffix) is the most adequate one, thus addressing the question of whether the complex words derived by *en*-prefixation in both English and Catalan (like those in (3) and (4)) are really exceptions to the RHR or not.

To carry out this task, I will focus on the argument structure of derived Vs and investigate the possible source of Θ -roles, which in turn will allow me to address the issue of whether the prefix contributes to the Θ -grid of the derived word. If the prefix does indeed contribute to the Θ -grid of the resulting word, I will corroborate a Θ -role percolation approach to the inheritance of thematic information (cf. Boij 1988, Levin & Rappaport 1988, Gràcia 1992, 1995 and Neeleman & Schipper 1992 and Mateu's 2001a, 2002) view of complex denominal Vs. Mateu argues that the preverb of complex denominal Vs in Germanic languages (such as the German word *ver+gärtnern* 'to away-garden') is part of the main thematic structure, thus also contributing to the resulting Θ -grid of the predicate.

Reinhart's theta system (2000, 2001) represents one of the different reinterpretations of the 'Theta' theory in Chomsky's Principles and Parameters approach which have been proposed recently. Another reinterpretation is embodied in HK (1993, 1998, 2002). While Reinhart's proposal relies on Θ -roles, HK's is based on direct interpretation of the structure. According to the latter, the position of an argument in their lexical-syntactic structures equals its thematic role. For instance, the object is not assigned the role Theme, because it is already a Theme as a result of its being in a specific structural position which has this particular semantics. Therefore, the source of Θ -roles will be crucial to determine which approach is superior. If thematic roles always originate in the same position, then HK's approach should be favoured for economy reasons, i.e. the semantics can be read off from the structure and there is no need for a linking system between Θ -roles and syntactic positions. If Θ -roles do not always come from the same structural position, then Reinhart's framework should be adopted. The two different views of Theta theory will be compared, although my study will be, as already noted, framed within Reinhart's theta system. HK's (1993, 1998, 2002) proposal will be briefly discussed to see how their analysis can explain the data presented in section 3. If their account can deal with the data satisfactorily, that will mean that my analysis should be revised and modified accordingly.

Given that the Catalan data will involve many reflexive Vs, a position as to how to consider them will be taken. That is, my study will provide an answer to the question of whether reflexive Vs should be treated as either unaccusative (Grimshaw 1990, Sportiche 1998) or unergative (Reinhart & Siloni 1999) entries.

This paper is organized as follows. Section 2 contains some theoretical background to understand Reinhart's theta system (2.1), a brief explanation of the different approaches to reflexives (2.2), and some discussion about Θ -percolation and inheritance (2.3). In section 3 the results of the data are presented and discussed. Finally section 4 provides the present study with some conclusions and questions for further research.

2. Theoretical background

This section provides the basics of Reinhart's (2000, 2001) theta system (including her linking system), some discussion about the different analyses of reflexives, and a brief explanation of how Θ -percolation and inheritance work.

2.1. Reinhart's theta system

Reinhart's theta system (2000, 2001) represents a formal definition of thematic roles. By proposing two binary features: [+/-*c*] and [+/-*m*] (which result in eight feature clusters; see below), Reinhart derives the Θ -roles of the 'Theta theory' found in the Principles and Parameters framework (Chomsky 1995). Seeing that causality is crucial in thematic structures and observing that there is an overlap between the Cause and Agent roles: 'if an argument is an agent of some change of state, it is also a cause for this change' (Reinhart 2000: 25), Reinhart labels the property they share [*c*], 'cause change'. Then, she notes that agency, unlike causality, involves volition and intention, and she labels this feature [*m*], 'mental state of the participant'. By assuming two features and two possible values for each, the system generates eight feature-combinations or, in Reinhart's terms, eight feature bundles. Although some of them, namely the mixed-value clusters ([+*c*-*m*]) and the unary clusters ([-*c*]), are more varied in their role interpretation than fully specified clusters with a [+] value for each feature (e.g. [+*c*+*m*]), there is still a (strong) correspondence between the clusters and the Θ -roles. Here I reproduce the correlations (Reinhart 2001: 3):

- (7) [+*c*+*m*] agent [+*c*-*m*] instrument
 [-*c*-*m*] theme/patient [-*c*+*m*] experiencer
 [+*c*] cause (unspecified for /*m*; consistent with agent and instrument)
 [+*m*] (unspecified for /*c*) with verbs such as *love*, *know*, *believe* (externally generated); *laugh*, *cry*, *sleep* (requiring an animate argument)
 [-*m*] (unspecified for /*c*) usually expressing subject matter/locative source
 [-*c*] (unspecified for /*m*) usually expressing internal roles like goal, benefactor (typically dative or PP)

Any linking theory about Θ -roles has to map the thematic specification (irrespective of its representation by means of Θ -role labels, feature clusters, etc.) of a lexical entry onto syntactic positions. That is, there must be rules or some mapping connecting the notion agent or the cluster [+*c*+*m*] to notions like external and to a specific position in the sentence. (See Williams 1981b, Carrier-Duncan 1985, Baker 1988, Grimshaw 1990, Neeleman & Schipper 1992, Samek-Lodovici 2003, for some linking suggestions).⁵ Reinhart (2001) proposes that there is a lexical operation which assigns indices to the roles on the V's Θ -grid: 1 marks the external role and 2 marks the internal role. These marking procedures only apply to verbal entries with at least

⁵ For example, Williams (1981a) distinguishes the external Θ -role by underlining it or in Grimshaw's (1990) thematic hierarchy, the external Θ -role corresponds to the least embedded one. These are just some of the conventions to relate roles to syntactic positions. That is, the underlined Θ -role in Williams or the least embedded Θ -role in Grimshaw's system is merged externally.

two arguments, by assigning index 2 to a [-] cluster ([-c-m], [-c], [-m]) and index 1 to a [+] cluster ([+c+m], [+c], [+m]). The result is that a cluster marked 2 must merge internally and a cluster marked 1 must merge externally.⁶ Only mixed clusters ([+c-m], [-c+m]), which are not marked, can merge in either position, subject to other requirements (e.g. Merge externally whenever possible for economy reasons, since the external position must always be filled eventually).

Given Reinhart's assumption that each V is associated with only one thematic structure and that all Vs are underlyingly transitive,⁷ she derives reflexives, unaccusatives and unergatives by means of a lexical operation called reduction, which reduces the V's arity by one. If the internal argument is reduced (i.e. if the operation Reinhart calls reflexivization applies), a reflexive entry is derived. If the external argument (necessarily specified as [+c]) is reduced (i.e. expletivization has applied, in Reinhart's terms), the result can either be an unaccusative or an unergative alternate, a result which will depend on the feature specification of the remaining argument. To see how the marking procedures work and how the mapping is established, we will consider the basic verb entry of *break* (*{John/The storm/The stone} broke the window*) and its unaccusative variant (*The window broke*).

- (8) a. Base entry: *break* ([+c], [-c-m])
 b. Marking: *break* ([+c]1, [-c-m]2)
 c. Reduction_{Expletiv.} (*break*) ([-c-m]2)

(8a) indicates that the V *break* is transitive and thus takes two feature clusters (two arguments). The marking system establishes that the [+c] (cause) cluster is marked 1 and that the [-c-m] (theme/patient) is marked 2. The mapping instructions will then determine that the [+c] and [-c-m] arguments will merge externally and internally on the transitive variant respectively. Although the [+c] argument will not be present if expletivization takes place (8c), such process does not directly affect the remaining argument because it is still marked 2. It cannot merge externally, although it can move to the external position later to satisfy the Extended Projection Principle (EPP). Given that the remaining argument is [-c-m], an unaccusative verb is derived. If the remaining feature cluster had been [-c+m] (*{The man/The storm/The box} worried Mary-Mary worried*), the argument bearing such specification would have been able to merge externally, since the cluster, being mixed, would not have been given an index. The requirement of external merger whenever possible would have had its effects and an unergative would have been derived.

Finally, to exemplify reflexivization, consider (9).

- (9) a. John dressed the baby. ([+c+m]1, [-c-m]2)
 b. John dressed. ([+c+m]1)

⁶ *External merging* refers to that role merged outside the maximal projection of its predicate, and *internal merging* refers to those roles merged within the maximal projection of their predicate.

⁷ The relationship between the causative and inchoative forms of a V is still an open issue in generative grammar. It is generally assumed that one form derives from the other in the lexicon, but it is not clear which form is the basic one. For example, Reinhart (2000, 2001) believes that the inchoative form derives from the causative, whereas HK (1998) claims the opposite. See Gràcia (1995) for a different view, according to which the two forms share the same base, but neither of them is derived from the other.

The internal argument marked 2 in (9a) has been reduced in (9b). This has no effect on the merging of the remaining argument, since it is still marked 1. It will merge externally, as in (9a).

Because Reinhart's marking procedures apply only for the arguments on a V's Θ -grid, it will only be at the verbal node, and not before, that arguments will get index 1 ([+] cluster), 2 ([-] cluster) or no marking at all (mixed cluster). That means that the Θ -roles of As and Ns cannot follow Reinhart's marking system, since this is not applicable to them. With respect to the relationship between the Θ -roles of As and Ns and their syntactic position (whether they are external or internal arguments), I follow the regularities already established in other work (for example, Williams 1981b). That is, a [-c-m] role on an A will be external. The same role on a N will be internal and the R-role,⁸ which is associated with Ns, will be external. Such approach seems problematic at first sight, because the external argument of an A (*This apple is edible*) is internalized when it is on the V's node (*I ate the apple*). However, this apparent internalization is explained if we adopt the view according to which only Θ -roles percolate, and the notions external or internal are determined by the category the Θ -roles are associated with (see Neeleman & Schipper 1992 for a similar view). In other words, the A will force an argument specified as [-c-m] to be external, whereas the same role on a V will be given index 2, which will determine internal merging.

2.2. Two analyses of reflexives: unergative vs. unaccusative

As for the treatment of reflexives (quite abundant in my Catalan survey, cf. Appendix B), we have just seen that, according to Reinhart, they are unergative entries, which have been derived by reducing the internal argument of a transitive V (see Reinhart & Sioni 1999 and Reinhart 2000, 2001 for details). However, Reinhart is somehow forced to stipulate that reflexivization is the result of reducing the internal argument, because she already has an external reduction operation for expletivization (recall that this is how she derives unaccusative and unergative entries). Similarly, one could also stipulate that reflexivization is the outcome of reducing the external Θ -role, and that *se* is the obligatory marker (in Romance languages) that results from the reduction operation. In fact, this is roughly the unaccusative approach to reflexives, which has also been defended (see Grimshaw 1990, Sportiche 1998, for instance). According to this approach, the subject of reflexives, like the subject of unaccusative verbs, is the underlying object. Within the unaccusative analysis of reflexives, there are two different variants: the lexical and the syntactic. While the former assumes that the external argument is lexically absorbed, the latter assumes

⁸ The source of the R-role is to be found in Williams (1981b), who notes that Ns also have external Θ -roles. In sentence (i),

(i) I consider that [destruction of a city by evil forces]

the predicative NP *destruction* has two internal arguments: the Theme *a city* and the Agent *evil forces*, but it also has an external argument which has no counterpart in the verbal system, i.e. *that*, which he gives the label R. That is, *destruction of a city by evil forces* is predicated of *that*. "The label R is meant to suggest 'referential', since it is this argument position R that is involved in referential uses of NPs as well" (p. 86)

that the clitic *se* is the external argument present in syntax. On theoretical grounds, there is no reason for choosing one approach (reflexives as unergatives vs. reflexives as unaccusatives) over the other. The data of my Catalan study will, however, suggest that the unaccusative approach to reflexives is the one which seems to be on the right track.

2.3. Θ -percolation and inheritance

As far as the Θ -percolation approach is concerned, the basic idea is that the thematic information of a complex word is derived from the different elements that form the word, irrespective of whether they are prefixes or suffixes.⁹ This view of Θ -percolation is in conflict with the RHR, which states that only the head is able to transfer its features. The data analysed in my study will show that the strict RHR (Williams 1981a) has to be abandoned, in favour of the Rel. RHR (Di Sciullo & Williams 1987: 25-28), according to which the head for a specific feature is the rightmost element that contains the feature in question. To illustrate this, consider the Latin word in (10), which according to the Rel. RHR will have two heads, given that both *bi* and *tur* are the rightmost elements with respect to the features they are marked; i.e. the former is specified as [+future] and the latter is marked with the feature [+passive].

- (10) ama bi tur
 [+future] [+passive]

Regarding inheritance, it refers to the relationship between the argument structure of a derived word and its input elements. A complex word inherits an argument from the base when the argument may be represented as an argument of the derived word either syntactically (sometimes referred to as external or syntactic inheritance) or internally to the complex word (sometimes called internal or morphological inheritance). To see the effects of inheritance, consider (11).

- (11) a. Manchester is industrial.
 b. The government industrialized Manchester.

The fact that the suffix *-ize* forms agentive Vs from As can be explained under the assumption that *-ize* provides an Agent role. The immediate consequence will be that the A's external Theme will be inherited as the V's internal Theme. Inheritance then accounts for the shared thematic structure between (11a) and (11b) (cf. e.g. Booij 1988, Levin & Rappaport 1988, Neeleman & Schipper 1992, Gràcia 1992, 1995, Gràcia et al. 2000, Williams 2004) (see e.g. Hoekstra & van der Putten 1988 for a different view).

⁹ Other early statements of feature percolation can be found in Selkirk (1982), Fabb (1984), Scalise (1984), and Lieber (1989). For a modern version of a mechanism similar to percolation, see Neeleman & van de Koot (2002) who use upward copying of functions introduced by terminal nodes.

3. Prefixed verbs

This section provides the results of the Catalan (Padrosa 2005a) and English data. More specifically, this section analyses how Catalan and English *en*-prefixation works with respect to underived Vs (sections 3.1.1 and 3.2.1), deadjectival (sections 3.1.2 and 3.2.2) and denominal (sections 3.1.3 and 3.2.3) Vs. The study of “all” *en*-prefixation Vs has been carried out by means of dictionaries.¹⁰

3.1. Catalan data

The Catalan classification of prefixed Vs presented in this section largely agrees with that of Gràcia et al. (2000). We both have reached the conclusion that there is no regularity in prefixed Vs whose source is a V, and that deadjectival Vs have the meaning ‘to make A’ when used transitively and ‘to become A’ when used intransitively. The only difference has to do with denominal Vs, which Gràcia et al. have classified into four categories, while I have classified them into three, namely location Vs, locatum Vs, and Vs of creation. Their fourth group includes Vs like *engelosir* ‘to make somebody jealous’ and *embasardir* ‘to frighten’ which in my classification have been included in the locatum group. Although they are not typical locatum Vs with a physical object being placed somewhere, they still show the same behaviour and semantic paraphrase. For instance, if you *frighten* somebody, you ‘put fear into that person’ somehow. Other Vs which I have included in the locatum group are *encoratjar* ‘to encourage’ and *enrabiar* ‘to enrage’.

3.1.1. *V-to-V* prefixation

This study has focused on the Catalan prefixed Vs which maintain a semantic relation with their bases and speakers are aware of the connection. For instance, pairs of Vs like *cantar* ‘to sing’ and *encantar* ‘to cast a spell on somebody’ have not been included because the relation between them is lost, i.e. the prefix has become lexicalized and is not seen as a prefix any more (see details in the introduction to Appendix B).

Although the remaining Vs (seven on my list) should be relevant to find out how the argument structure of the prefixed V differs with respect to its base, no conclusions can be drawn (maybe due to its reduced number).

- (12) a. Una barca va travessar l’Atlàntic.
‘A boat crossed the Atlantic’
- (13) a. Van entravessar un tronc al mig del carrer.
‘They laid a trunk across the street’
b. Se m’ha entravessat un osset a la gola.
‘A little bone got caught in my throat’

¹⁰ Concerning the Catalan data, the *Gran diccionari de la llengua catalana* (GDLC) (1998) and the *Diccionari de la llengua catalana* (DLC) (1995) have been the main tools, whereas the *Diccionari general de la llengua catalana* (DGLC) (1932) has been used for clarification and further reference when necessary. As for the English data, the *Collins Cobuild English Dictionary* (CCED) (1995) has been used in conjunction with the Concise Oxford Dictionary (COD) (2001) and the Merriam-Webster Online Dictionary (MWOD).

Based on the examples given in (12, 13), one could suggest that the prefix *en-* gives a causative meaning to the V. That is, if people laid a tree trunk across the street, they caused the trunk to be somewhere. However, if we look at the base V to which the prefix attaches, we can also have a causative interpretation. Although such reading is not available in (12a), the same V can be used with a clear causative interpretation, as the following sentence shows:

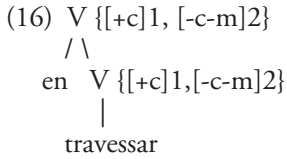
- (12) b. Li vaig travessar el pit amb l'espasa.
 I thrust the sword through his chest.
 'I caused the sword to go through his chest'

Although the prefix is not the source for the causative reading, because such meaning is available without the prefix, one could entertain the idea that the prefix contributes to the Θ -grid of the prefixed V. It could seem that the *en-* prefix adds a locative role (*al mig del carrer* in (13a), *a la gola* in (13b)) to the Θ -grid of the prefixed V. However, such a proposal has to be rejected on the basis of the following examples:

- (14) a. El treballador subornà el cap.
 'The worker bribed the boss'
 b. El venedor ensibornà el client.
 'The seller fooled the client'
- (15) a. Ella va retirar els diners del banc.
 'She withdrew the money from the account'
 a' Va retirar la mà que jo li havia allargat.
 'He pushed away my approaching hand'
 a'' Ell es va retirar a un monestir.
 'He retreated to a monastery'
 b. Quan ell va allargar la mà, jo vaig enretirar la meva.
 'When his hand approached me, I moved my hand away'
 b' Ells van enretirar la taula.
 'They moved the table out of the way'
 b'' Si us enretireu, hi haurà prou espai per les taules.
 'If you throw yourself back, there will be enough space for the tables'

In both cases (14, 15), there is no addition of any Θ -role. Regarding *subornar/ensibornar* (14), both take the same roles: [+c+m] (agent) and [-c-m] (theme). As for *retirar/enretirar* (15), they show the opposite pattern of *travessar/entravessar*. When used transitively (15a, a' and 15b, b'), *retirar* has an extra Θ -role in (15a) (*del banc*) which would get reduced in the prefixed V. That is, the V in (15a) needs a locative source but that is not compulsory for *enretirar* (15b, b') or even for *retirar* in (15a'). The same holds for the reflexive variants: i.e. *retirar-se* in (15a'') needs a locative source (*a un monestir*) but *enretirar-se* in (15b'') does not. The examples just mentioned show that no systematic patterns between the two argument structures can be observed, i.e. the prefix does not seem to bring anything visible to the V.

The Θ -percolations in (16) will then be assumed for the previous Vs. Let us consider (*en*)*travessar* for concreteness sake.



Following Reinhart’s marking system, the [+c] role is assigned index 1, determining its external merger, and the [-c-m] role will get index 2, forcing internal merger of such role. I assume that some reduction process takes care of the reflexive variants.

At this point, the postulation of an empty suffix seems irrelevant and so does the question of whether *en-* is a left head. In V-to-V prefixations, the base is already a V and there is no conversion, for which the suffix or prefix can be made responsible. Further, there is no apparent change in the argument structure of prefixed Vs and those Vs without a prefix. With respect to the role of *en-*, one could suggest that the prefix does have some feature specification, but this does not percolate because the base V has the same features and, according to the Rel. RHR, the rightmost element specified for some features is the one which gets its features percolated. To check whether that can be the case, we will have to turn to the next sections which also provide an answer to the question of whether prefixed Vs are real exceptions to the RHR or not and to which approach to reflexives is the right one, given that in V-to-V prefixation there is only form (*endur-se* < *dur*) that admits the clitic *se* and nothing can be concluded on the basis of a single form.

3.1.2. *A-to-V prefixation*

A very common pattern for *en+A* Vs is that most of them allow a transitive (to make A) and an unaccusative (to become A) variant, the latter typically expressed with the reflexive clitic *se/es* (included within parentheses below). In the following examples, all of which allow the two verbal variants, the feature clusters of both the A’s Θ-role and the derived V’s Θ-roles have been placed next to them. The feature specification in parentheses indicates that this role is absent in the unaccusative variant of the V ((b.2) sentences), but present when the V is used transitively ((b.1) sentences).

- (17) a. *dolç*_A ‘sweet’ [-c-m]
 b. *endolcir(-se)*_V ‘to make/become sweet’ ([+c] [-c-m])
 b.1 (pro [+c+m]) *Vaig endolcir la llet* [-c-m]. ‘I sweetened the milk’
 b.2 *La llet* [-c-m] *s’ha endolcit*. ‘The milk became sweeter’
- (18) a. *negre*_A ‘black’ [-c-m]
 b. *ennegrir(-se)*_V ‘to make/become black’ ([+c] [-c-m])
 b.1 *Els núvols* [+c] *ennegriren el cel* [-c-m]. ‘The clouds blackened the sky’
 b.2 *El cel* [-c-m] *s’ennegrí*. ‘The sky turned blacked’
- (19) a. *ros*_A ‘blonde’ [-c-m]
 b. *enrossir(-se)*_V ‘to make/become blonde’ ([+c] [-c-m])
 b.1 *El tint* [+c-m] *l’* [-c-m] *ha enrossit*. ‘The dye made his hair turned blonde’
 b.2 *El seu cabell* [-c-m] *s’ha enrossit*. ‘His hair turned blonde’

To find out which role the prefix *en-* and the alleged \emptyset -suffix play in deadjectival Vs, their argument structures have to be compared with those of their corresponding As. In (17-19) the A from which the V is derived has a [-c-m] role, which is maintained in both transitive and unaccusative variants of the V. However, one needs to explain the presence and source of the extra Θ -role [+c] in the transitive variant. Although the prefix might look as the most obvious source, this analysis would run into problems when considering Vs like those in (20) and (21):¹¹

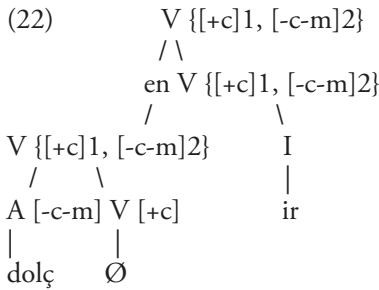
- | | | | |
|------|------------------------------|--|---------------|
| (20) | a. canut _A | ‘white-haired’ | [-c-m] |
| | b. encanudir _V | ‘to become white-haired’ | [-c-m] |
| (21) | a. cresp _A | ‘curly’ | [-c-m] |
| | b. cresp _V | ‘to curl one’s hair’ | [+c] [-c-m] |
| | c. encresp(-se) _V | ‘to curl one’s hair’ ‘to heighten the waves’ | ([+c]) [-c-m] |

Although we would be able to explain *encanudir* by saying that the [+c] role of the prefix is reduced, and that it is not in the case of *encresp* (on the transitive variant), Vs like *cresp* still cannot be accounted for, since there is no source for the unexpected [+c] role if we assume that such role originates in the prefix. Another source for the [+c] role needs to be found. One could entertain the idea that the [+c] role originates in the inflectional suffix (e.g. *-ar* in the case of *encresp*), but that option is a dead end, since inflectional suffixes, unlike derivational ones, do not contribute to Θ -grids. In addition, given that an inflectional suffix is present in each and every simplex V, such view implies that all unaccusative Vs are derived by a reduction operation, an option which needs to be investigated further. A \emptyset -suffix then seems to be the only possible candidate left. On the basis of examples like (20-21), I propose that the \emptyset -suffix always carries a [+c] role, although this is not active all the time (i.e. it can be reduced). I also propose that the same \emptyset -suffix is responsible for the conversion of As to Vs. The inflectional suffix and the prefix *en-* cannot be responsible for the conversion. Inflectional suffixes do not change category and the derived V *cresp* in (21) clearly illustrates that the prefix is not needed, since this is absent and a deadjectival V can still be derived.

Given that my analysis presupposes a specific direction of derivation: $A \rightarrow V \rightarrow en+V$ ($cresp_A \rightarrow cresp_V \rightarrow encresp_V$), one might think that a weakness of this analysis is that not always is it possible to derive existing intermediate Vs (marked as ‘!’ in $canut_A \rightarrow !canudir_V \rightarrow encanudir_V$), but the possibility of deriving possible but non-existent words has been established in other work (for instance, see Stiebels 1998, Ackema & Neeleman 2004 who argue for an overgenerating morphology).

To see how the analysis just proposed for deadjectival *en-*prefixations works, the Θ -percolations and marking procedures for (17), *endolcir(-se)*, will be presented.

¹¹ Go to Appendix B, the section of deadjectival Vs, to view other unprefixated Vs which contain a [+c] role, e.g. *agrir(-se)* ‘to sour’, *corbar* ‘to bend’.



The Θ -role of both the A and the conversion-suffix percolate, resulting in a transitive Θ -grid, where the marking procedures assign indices 1 and 2 to the [+c] and [-c-m] arguments respectively. The indices will, in turn, determine external merger for the [+c] role and internal merger for the [-c-m] argument. As noted, this V participates in the transitive-unaccusative alternation. I suggest that first a transitive Θ -grid is generated and then a process reducing the [+c] role takes place. In this case, it is clear that reflexives are the result of reducing the external argument, thus explaining why the (b.2) sentences in (17-19) do not have the role [+c], but only the [-c-m] one, which is inherited from the A. It could be said that the [+c] role is lexically absorbed, leaving *se* as the marker of such process, or that the clitic itself is the external argument containing the [+c] feature. Either view is compatible with my analysis. Otherwise, if one tried to derive the unaccusative variants by reducing the internal argument, the meaning of the sentences would not make much sense. Consider (23).

- (23) a. El tint s'ha enrossit.
 'The dye turned blonde'

On the basis of A-to-V prefixations, I conclude that the RHR can be maintained, since the \emptyset -suffix, and not the prefix, is responsible for the conversion of As to Vs and for providing the [+c] role sometimes present in deadjectival Vs, whether prefixed or not. The presence or absence of the [+c] role is in turn determined by the reduction operation. Given that it is the external argument that is reduced in the case of *en*+A Vs, the unaccusative approach to reflexives seems superior to the one which considers reflexives to be unergative entries (i.e. internal reduction has taken place). Again, the prefix does not have any visible effects on the resulting Θ -grid. One can only hypothesize that if the prefix has some features, these should be the same as those of the base (i.e. [-c-m]) and that the Rel. RHR determines that the features of the A, and not those of the prefix, percolate. The next section shows that the basic pattern found in *en*+A Vs will also hold for *en*+N Vs.

3.1.3. *N-to-V prefixation*

Three semantic patterns can be distinguished within *en*+N Vs: the first one means 'to put something/somebody in/onto/towards N' (24) (cf. location Vs); the second one has the opposite relation between the two arguments, i.e. 'to put N around/in

something/somebody' (25) (cf. locatum Vs);¹² and finally, the third semantic pattern involves the creation of the N, namely, 'to make N', which is the same pattern found with As (26). The feature specification for each Θ -role has been placed next to the derived V. The R-role is associated with every N (Williams 1981b) (see footnote 8).

- (24) a. caixa_N 'box' R
 b. encaixar_V 'to put something in a N' [+c] [-c-m]
 b. En Joan encaixà els llibres.
 'John packed the books away'
- (25) a. caputxa_N 'hood' R
 b. encaputxar(-se)_V 'to put the N on somebody's head' [+c] [-c-m]
 b. Ell encaputxà la Maria.
 'I put the hood on Mary's head' / 'I covered Mary's head with a hood.'
- (26) a. rai_N 'raft' R
 b. enraiar_V 'to make a N' [+c] [-c-m]
 a. Els homes enraïaren els troncs.
 'The men created the raft out of logs' / 'The men tied logs together to create a raft'

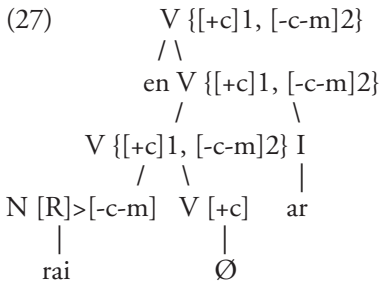
Although one could think that the semantics of the Vs in (24) and (25) looks quite distinct from each other, the division between the two may get blurred in some cases. This is made evident by Vs like *envinagrar* 'EN-vinegar_V', which can have both readings. That is, *envinagrar* can either mean 'to soak something (e.g. pickles) in N' or 'to pour N over something (e.g. food)'. (See Appendix B, section A/B for other words like *envinagrar*).¹³ This suggests that the division between the two groups may not be linguistically relevant after all, leaving us with two semantic patterns for denominal *en*-prefixations: one involving a change of location (24, 25) and the other involving the creation of the N (26).

Since Vs of creation have the same semantics as *en+A* Vs, it is logical to assume that similar Θ -percolations take place. Although the [+c] role of *enraiar* can come from the conversion affix, the [-c-m] role has no apparent source, since the N only has an R-role. However, Williams (1981b) argues on semantic grounds that the R-role could be interpreted as a theme, a view I adopt, given that it is in accordance with the feature specification we would expect ([-c-m]) from the N.¹⁴ Once the [+c] role and the reinterpreted R-role, [-c-m], are on the verbal node, they will get index 1 and index 2, which will determine external and internal merger respectively (see 27).

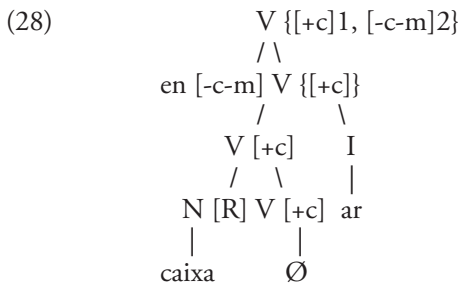
¹² For discussion about location and locatum Vs, see Clark & Clark (1979), Kiparsky (1997), HK (1993, 1998, 2002), Mateu (2001b, 2002), among many others. Regarding denominal Vs with a meaning of creation, see Clark & Clark (1979) and Gràcia et al. (2000) for example.

¹³ Mateu (2001b, 2002) also groups location and locatum Vs together, and treats them as 'change of state' Vs.

¹⁴ Williams rejects the option of considering the R-role an external theme, because then there would be two themes in a single Θ -grid. In this respect, Neeleman & Schipper's (1992) remarks about Θ -role reinterpretation are illustrative. According to them, a Θ -role can only be reinterpreted as a role that is semantically close, and they also consider that that is the case for themes and R-roles.



Although the reinterpretation of the R-role accounts for Vs of creation, it leaves the presence of the [-c-m] role in verbs like *encaixar* or *encaputxar* unexplained, since their meaning is not ‘to make/become a box/hood’. For these Vs, I propose that the prefix is responsible for the [-c-m] role found in the V’s Θ-grid. In the two previous sections I already hinted at the possibility of the prefix having some feature specification, but due to the Rel. RHR, the prefix’s features were always obscured. In en+N Vs, though, the prefix constitutes the rightmost head specified for the features [-c-m], since the Ø-suffix only has a [+c] role and the noun’s R-role is not reinterpreted. Hence, the features of the prefix [-c-m] percolate up to the V’s node, where they will get index 2 (internal merger) (see 28).



Interestingly, some Vs can have both a creation and ‘change of location’ reading:

- (29) *coixí_N* ‘cushion’ R
encoixinar_V ‘to make a N’ / ‘to put Ns in a place’ [+c] [-c-m]
- toia_N* ‘bouquet’ R
entoiar_V ‘to make a N’ / ‘to put Ns in a place’ [+c] [-c-m]

I suggest that they will have one or the other reading, depending on the source of the [-c-m] role. If the R-role is reinterpreted, the V will have a creation meaning, but if it is not, then the [-c-m] features will come from the prefix and this will result in a locative meaning.

To explain the existence of the intransitive variant (typically marked with the clitic *se*) of en+N Vs, I will adopt the reduction mechanism already used before. Since most intransitive variants clearly show reduction of the external argument (30), and in only a few cases is it difficult to tell which argument gets reduced like in (25) (it could be either the external or internal one), I assume that external reduction takes place in en+N verbs uniformly. Crucially, there are no cases with clear internal reduction.

- (30) a. *encoratjar*_V ‘to encourage’ [+c] [-c-m]
 b. *encoratjar-se*_V ‘to become encouraged’ [-c-m]

- a. El primer gol els encoratjà. ‘The first goal gave them courage’
 b. Amb el cinquè gol es (CL) van encoratjar. ‘With the fifth goal, they were encouraged.’

Like deadjectival *en*-prefixations, *en*+N Vs also seem to have an intermediate stage in which the N has become a V, but the prefix is not present as yet. Among the existing intermediate Vs, two groups can be differentiated. The first group includes those forms whose meaning is related to the prefixed V and the second one contains those intermediate forms which, according to the dictionaries, have the same meaning as the prefixed Vs. The two groups include location and locatum Vs, Vs that can have the two readings and Vs of creation (see Appendix B, section of denominal Vs). Whereas the first group does not pose a problem to my analysis, the second one does.

As for the first three types of Vs in the first group (i.e. location and locatum Vs, and Vs that can be interpreted either way), the locative meaning is associated with the prefixed V, in agreement with my analysis, according to which the prefix *en*- with the features [-c-m] contributes to such meaning. Compare *caminar*_V ‘to walk’ with *encaminar*_V ‘to put somebody in the correct path’.¹⁵ Regarding Vs of creation, no locative reading is involved, so whether the prefix is present or absent is irrelevant.

Regarding the second group, the Catalan dictionaries define some intermediate Vs as having the same meaning as their prefixed version. Here are included the Vs with a locative reading, i.e. location and locatum Vs, and those Vs that can have the two readings. All these intermediate forms with a locative meaning question my analysis, since there is no source for the [-c-m] role, due to the absence of the prefix. A possible explanation for these locative intermediate forms is given in section 3.3.

The conclusion from this section is that the basic generalizations established in the previous sections also hold for Catalan *en*+N Vs. The \emptyset -conversion affix is responsible for the [+c] role. Finally, we have seen that the prefix does have some features, and that these play a role in Vs expressing a change of location. That is, a denominal V will express a change of location if the features of the prefix percolate, but it will have a creation reading if the N’s R-role is reinterpreted.

3.2. English data

Although several authors have worked on unprefixated locatum and location denominal Vs and have classified them into extensive lists (see footnote 12), no classification has been provided for *en*-prefixated Vs, as far as I am aware of (except for authors like Marchand (1969) who deal with historical data). I hope then that my classification and my findings here will shed light on a not much worked on area.

¹⁵ At first sight, the intermediate V *sorrar* ‘to put sand on something’ would be an exception to my generalization. The locative reading has no source because the prefix is absent. However, Catalan speakers, when asked to choose between *sorrar* and *ensorrar*, prefer the prefixed version. According to the etymology of these words, first the V *sorrar* (century) XIV was formed out of the N, and then it may well be that speakers added *en*- to best express the locative meaning and the result was *ensorrar* (c. XVI).

3.2.1. *V-to-V prefixation*

Like in Catalan, there is no systematic change in the argument structures of the few examples of *V-to-V* prefixation in English with respect to their unprefixated version. Of the eight Vs on my list, one could argue that some should be removed because a N (and an A in one case), exists together with the unprefixated V, and could suggest that the N (or A), and not the unprefixated V, is the base on which the prefixated V is built.¹⁶ If that were correct, it would be the case for most of the Vs, e.g. *enact*_V<*act*_V<*act*_N, *enchant*_V<*chant*_V<*chant*_N, *enclose*_V<*close*_V<*close*_A. However, I do not think that is the correct approach. According to Corbin's (1976) [cited in Varela (1993)] semantic criterion, a N is derived from a V if it can be paraphrased by 'the act of Ving' and has no affix. A N which cannot have the previous paraphrasing and has no affix added to it precedes the V. If we apply this hypothesis to the pairs of Ns and Vs on my list, we will see that the V comes first and that the N and prefixated V are derived later. To exemplify, consider *chant*_{N/V}. Given that a *chant* is the act (result) of chanting, *chant*_V is the source on which the other forms are based. The result is that the prefixated Vs in question are not denominal Vs, as one might think at first sight, but rather are derived from a V. Although such forms do not have to be removed from the list, others need to, namely *engrave*, *enliven*, and *ensue*. The two first forms are based on archaic Vs: *grave*_V is rarely used and *liven*_V has been replaced by *liven up*. As for the V *ensue*, some speakers do not longer see its compositional structure, which would be related to the V *sue*.

Five Vs remain on the list of *en*-prefixated Vs, not enough to find a systematic pattern between the prefixated Vs and their unprefixated version, as will be seen shortly by the following examples. (To see the other Vs, go to Appendix A)

- (31) a. The young boy confessed his desire to act.
 b. Her husband acted in Roberto's films.
 c. The little child enacted old stories.
- (32) a. The people outside chanted mantras.
 b. Merlin enchanted¹⁷ the house.
- (33) a. I joined my sister in California.
 b. The actress joined a dance company.
 c. The boss enjoined him strictly not to tell anyone else.
 d. Islam enjoins tolerance.

By comparing (31a) with (31c), one could initially suggest that *en-* adds a role to the Θ -grid of the unprefixated V: (31c) contains a [+c+m] (agent) role and a [-c-m] (theme) role, the latter not present in (31a). However, this option has to be abando-

¹⁶ Some authors have resolved the issue of what comes first in a derivation by means of category indeterminacy. For instance, Marantz (1997, 2001) argues that roots are underspecified for syntactic categories like N and V and that the morpheme attaching to the root will provide the category.

¹⁷ Whereas English speakers can still perceive the compositional structure of the V *enchant*, Catalan speakers cannot do the same with the corresponding V *encantar*. That can be explained by a gradual process, according to which speakers would lose the sense of compositionality progressively, being faster with some speakers than others.

ned. The unprefixing V can also have two Θ -roles, as shown in (31b) and there is no change between the argument structures of the unprefixing and prefixing Vs in (32) and (33). *Chant* and *enchant* both take the same roles (i.e. a [+c+m] (agent) role and a [-c-m] (theme) role) and so do *join* and *enjoin*. They can both show up in structures with three (33a, c) and two (33b, d) roles.

The reduced number of Vs that can be prefixed with *en-* and the lack of any apparent link between the two argument structures both in English and Catalan (section 3.1.1) questions whether V-to-V prefixation really constitutes a proper class in the two languages. In fact, HK (1993, 1998, 2002) and Mateu (2005) do not predict their existence. For instance, according to HK's theory of argument structure, unergative Vs like *laugh* and *dance* are derived from an initial transitive structure involving incorporation of a nominal head N into an abstract V (cf. HK 1993, 1998). More recently (2002), there is no incorporation mechanism although HK still assume an initial transitive structure, where the V, filled through Vocabulary Insertion this time, governs an empty nominal complement, thus accounting for the relationship between *laugh* as a N and V, the two clearly related. Similarly, Mateu (2002, 2005) also reaches the conclusion that Ns are the real primitive elements taken as complements by apparently underived Vs. The syntactic analysis proposed by HK will be taken up in the Discussion section.

For the moment, the Θ -percolations in (34) will be assumed for the Vs discussed in this section (cf. (16) for Catalan). Let us consider how Reinhart's percolation system of Θ -roles would derive the Θ -grid of (*en*)*chant*.

$$(34) \quad \begin{array}{c} V \{ [+c]1, [-c-m]2 \} \\ / \backslash \\ \text{en } V \{ [+c]1, [-c-m]2 \} \\ | \\ \text{chant} \end{array}$$

The V *chant* has two Θ -roles a [+c+m] (agent) and a [-c-m] (theme), a sufficient number of roles to allow marking. Accordingly, the [+] role will get index 1 and will merge externally, and the [-] role will receive index 2 and will merge internally.

To summarize V-to-V prefixation in English, no systematic patterns between the two argument structures can be observed, i.e. the prefix does not seem to bring anything visible to the V. At this stage, the question of whether *en-* is a left head is redundant, and so is the postulation of a zero-suffix. The base in V-to-V prefixation is already verbal and no conversion process can be attributed either to the prefix or empty suffix. After all, V-to-V prefixation may not constitute a proper class in English nor in Catalan.

3.2.2. A-to-V prefixation

This section addresses the question of whether the same analysis for Catalan *en+A* Vs can also explain the different types of deadjectival Vs found in English. Examples like those in (35-37) show that the same analysis can be maintained, although in English *en+A* Vs only allow the transitive variant, illustrated in (a). The sentences in (b) indicate that the unaccusative variant is impossible and (c) provides some alternatives

to (b). Note that this type of deadjectival Vs is not really productive: seven en+A Vs are the only existing forms nowadays (Appendix A).

- (35) rich_A [-c-m]
 enrich_V [+c] [-c-m]
- a. She will enrich the country. b. *The country will enrich.
 c. The country will {be/become} rich.
- (36) large_A [-c-m]
 enlarge_V [+c] [-c-m]
- a. The reporter enlarged the picture. b. *The picture enlarged.
 c. The picture grew larger. / The picture was enlarged.
- (37) noble_A [-c-m]
 ennoble_V [+c] [-c-m]
- a. His willingness to help ennobled Steven enormously. b. *Steven ennobled.
 c. Steven was ennobled by his willingness to help.

Following the analysis proposed for Catalan deadjectival Vs, the extra Θ -role [+c] present in the derived V is provided by a zero-suffix and the [-c-m] role comes from the A. The Θ -percolations and marking procedures for any of the forms in (35-37) are the same as those for *endolcir(-se)* in (22), repeated as (38) here for the V *enrich* (<*rich*), the only difference being that there is no inflectional morpheme in English.

- $$\begin{array}{c}
 (38) \quad \quad \quad V \{ [+c]1, [-c-m]2 \} \\
 \quad \quad \quad / \quad \backslash \\
 \quad \quad \quad \text{en } V \{ [+c]1, [-c-m]2 \} \\
 \quad \quad \quad / \quad \backslash \\
 \quad \quad \quad A \quad [-c-m] \quad V \quad [+c] \\
 \quad \quad \quad | \quad \quad \quad | \\
 \quad \quad \quad \text{rich} \quad \quad \quad \emptyset
 \end{array}$$

Some evidence to confirm that the zero-suffix, and not the prefix, is responsible for the conversion process and the [+c] role of the prefixed V comes from a more productive type of English deadjectival Vs, namely those without prefix. Consider the As in (39a), their derived Vs in (39b) and some sentences ((40a) and (41a)) where the latter are used in context. The sentences in (40b) and (41b) show variability in behaviour with respect to the transitivity alternations (the unaccusative alternate is allowed by *clear*, but not by *clean*).

- (39) a. clean_A, clear_A, dirty_A, empty_A, narrow_A, thin_A [-c-m]
 b. clean_V, clear_V, dirty_V, empty_V, narrow_V, thin_V [+c] [-c-m]
- (40) a. The old lady cleaned her glasses with a napkin.
 b. *Her glasses cleaned with a napkin.
- (41) a. The cook thinned the sauce slightly. b. The sauce thinned slightly.

The A has the usual [-c-m] role and the V has a [+c] role, whose presence cannot be accounted for without a zero-suffix, since the prefix is not available. From these examples it is then clear that the prefix cannot contribute to the [+c] role present in the derived V, and hence cannot be a causativizer, as already noted earlier.¹⁸

Still there exists a third type of deadjectival Vs in English, which have no prefix *en-*, but instead they end with the suffix *-en*. (42) provides some examples, and the sentences in (43) and (44) show that this kind of Vs can participate in the transitive-unaccusative alternation.

- (42) a. black_A, bright_A, hard_A, sweet_A, thick_A, wide_A [-c-m]
 b. blacken_V, brighten_V, harden_V, sweeten_V, thicken_V, widen_V [+c] [-c-m]
- (43) a. The cook thickened the sauce. [+c] [-c-m]
 b. The sauce thickened. [-c-m]
- (44) a. The sun reddened the sky. [+c] [-c-m]
 b. The sky reddened. [-c-m]

From very early on, a number of authors (e.g. Halle 1973, Aronoff 1976, Siegel 1979, Scalise 1984, Fabb 1988) have noted their existence and have claimed that *-en* is the element triggering the conversion of As to Vs. If that is the correct approach, there are two sources for the [+c] role: a zero-suffix and the verb-forming suffix *-en*. That is an awkward situation for my analysis, and I suggest that there is a single suffix which sometimes has phonological content (the *-en* morpheme) and sometimes does not (the zero-suffix), thus avoiding the unwanted double forms for a unique function. The aforementioned authors agree that there are some constraints on the suffix's attachment. The suffix *-en* only attaches to monosyllabic As that on the surface end in a single obstruent, preceded by a vowel, which optionally may, in turn, be preceded by a sonorant. If an A violates the condition just stated and there is a related nominal form that satisfies it, then *-en* attaches to the N: e.g. *frighten_V* (*afraid_A* has two syllables), and *strengthen_V/lengthen_V* (*strong_A* and *long_A* end in a nasal). The focus here, though, is on A-based Vs derived by the suffix *-en* (cf. 42).

Although these *en*-suffixed forms have the same Θ -percolations as *enrich* (cf. 38), now the zero-suffix is replaced by *-en*. This third type of deadjectival Vs provides further evidence to say that the *en*-prefix does not have any features, or if it does, they are probably the same as those of the base, and due to the Rel. RHR, the features of the base get percolated, obscuring those of the prefix.

The present scenario predicts the existence of deadjectival Vs constituted by both prefixation and visible suffixation. As seen earlier, the source of the [+c] role is a suffix which can be full or empty of phonological content (*widen_V*, *sweeten_V*, *brighten_V* vs. *rich_V*, *noble_V*, *clear_V*). If that is correct, and the constraints on the suffix's attachment are satisfied, the prefix *en-* should be able to attach to bases with both types of suffixes, deriving prefixed deadjectival Vs, suffixed and non-suffixed. Considering

¹⁸ For a different view, see e.g. Zwanenburg (1988), and Grimshaw (1990) who have proposed that the prefix *en-* gives the causative reading ([+c] role in my analysis) to Vs like *ennoble* and *enrage*.

historical data (Marchand 1969), one observes that this prediction is borne out. There was one period where forms like *enwiden*, *ensweeten* and *embrighten* existed alongside of forms like *enrich*, *ennoble*, and *enclar*. However, of all these prefixed and suffixed forms, *embolden* is the only existing word nowadays. The productivity of the different types of deadjectival Vs will be discussed in section 3.3.

Regarding the disparate behaviour of English A-based Vs with respect to the transitive-unaccusative alternation, we have seen that the first group (e.g. *enrich*, *enlarge*) does not allow the V to have an unaccusative variant, an observation unnoticed until now as far as I know (cf. 35-37b), while the third group (e.g. *sweeten*, *redden*) does (cf. 43-44b). As for the second group, there is no uniform behaviour (40-41b). Vs like *clean* cannot have the two variants, whereas Vs like *clear* can. HK's (1993, 1998, 2002) syntactic theory of argument structure cannot explain this. According to their theory, all deadjectival Vs, i.e. Vs incorporating As, should participate in the transitive-unaccusative alternation. By looking at the numbers of Vs in each group that allow the alternation, one sees that their claim is generally true. Deadjectival Vs admitting both an unaccusative and transitive variants exceed those Vs that do not. However, there is still a group of Vs which need some explanation. As observed by Kiparsky (1997), the real generalization behind the transitivity alternations does not have to do with the category (A) which gets incorporated into the V, but with the notion of agentivity. In Kiparsky's words, 'the availability of the causative alternation depends on the nature of the Agent's involvement in the event' (p. 495). In other words, only those Vs denoting processes which can be initiated and continued without an agent will allow the causative alternation. This claim is confirmed by the data of this section. Vs like *ennoble* and *clean* require the presence of an agent, and accordingly do not permit an unaccusative variant, where the agent would be suppressed. By contrast, Vs like *clear* and *sweeten* do not need the participation of an agent for the process to initiate and continue, and admit both the transitive and unaccusative variants. The conclusion is that my data favour a semantic account, rather than a syntactic one. It is not the syntactic category of the element which gets incorporated (A) but the semantics of the V that determines whether a V will show the transitivity alternations.

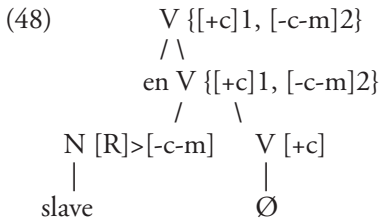
To summarize this section, we have seen that the analysis proposed for Catalan deadjectival Vs can be maintained for the three types (the type *embolden* does not constitute a fourth group due to its single membership) of deadjectival Vs found in English. Like in Catalan, the RHR can be observed: the prefix *en-* is responsible neither for the conversion of As to Vs nor for the [+c] role present in A-based Vs in their transitive variant. I have shown that the element responsible for the conversion and addition of the [+c] role is a suffix, which can be empty (zero-suffix) or full (*-en* suffix), depending on some phonological constraints. Now I would like to conclude this section by noting the productivity of each group briefly. The first one (type *enrich*) is non-productive, and closed (only seven forms). The zero-suffixed group which has no prefix (type *clean*) is more productive and finally the productivity of the *en-*suffixed group (type *sweeten*) is subject to some phonological constraints. I will take up the issue of productivity in the Discussion section.

3.2.3. *N-to-V prefixations*

At first sight, the picture for Catalan denominal prefixation is duplicated in the English data. Three different semantic patterns can also be differentiated: location Vs with the paraphrase ‘to put something/somebody in/onto/towards N’ (45), locatum Vs which can be paraphrased as ‘to put N around/in/into something/somebody’ (46), and Vs of creation with the semantic paraphrase ‘to make N’ (47).¹⁹

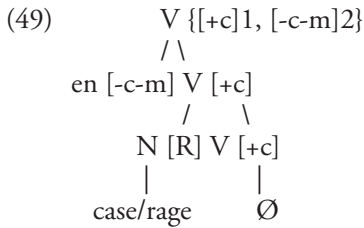
- (45) a. case_N R
 b. encase_V [+c] [-c-m]
 b. They encased the dangerous substance in a container.
- (46) a. rage_N R
 b. enrage_V [+c] [-c-m]
 b. He enraged the government by renouncing the agreement.
- (47) a. slave_N R
 b. enslave_V [+c] [-c-m]
 b. The captain enslaved the poor boy and he had to do what he was told.

As said before (for Catalan), Vs of creation like *enslave* (‘to make N’) involve the same semantics as deadjectival Vs (‘to make A’), suggesting that they should both have the same Θ -percolations. Recall that in deadjectival Vs the [+c] role came from the \emptyset -suffix and the [-c-m] role from the A. Although the same could be maintained for the [+c] role in Vs of creation, the [-c-m] role resulted from the reinterpretation of the N’s R-role (see section 3.1.3 for the explanation and footnote 14). Accordingly, (48) is the resulting structure for the V *enslave* (cf. 27). Note that the inflectional morpheme present in Catalan is now missing.



The same Θ -percolations cannot explain Vs like *encase* and *enrage*, given that they do not mean ‘to make a case/rage’. For Vs like these I proposed (for Catalan) that the [-c-m] role does not come from the reinterpreted R-role, but from the prefix, whose features have been obscured until now due to the Rel. RHR. In other words, the prefix in these locative Vs is the rightmost head marked with the features [-c-m], because the \emptyset -suffix has a [+c] role and the base N has an R-role, which is not reinterpreted. The [-c-m] features of the prefix will get index 2 once it is at the verbal node together with the [+c] role from the \emptyset -suffix, which will receive index 1. The resulting picture is illustrated in (49).

¹⁹ Appendix A shows the classification just mentioned. Although it contains another semantic pattern (i.e. ‘to give N’), that has been subsumed within the locatum Vs.



A single representation for location and locatum Vs may not look sufficient. How is one going to distinguish the two if both have the same representation? I suggest that the distinction between the two may not be linguistically relevant, given the existence of some Vs which can be interpreted either as location or locatum Vs. For instance, consider *entangle* which can mean ‘something (e.g. a whale) is caught in N’ or ‘to put N over somebody (e.g. people)’. The same phenomenon was found in Catalan. In the following section I will propose that the answer to the previous question has to do with pragmatics. If that view is correct, we are left with two semantic patterns with different Θ -percolations: one involving the creation of the N with the [-c-m] features coming from the reinterpreted R-role (cf. 48) and another one involving a change of location with the [-c-m] features coming from the prefix (cf. 49).

Now it remains to be seen whether the same analysis can explain other types of denominal Vs in English. Like in deadjectival Vs, locative en+N Vs also have a stage where the N has become a V but the prefix is still not present. Among the intermediate forms, some have a meaning related to the prefixed version and others have the same meaning, always according to the definitions given in the English dictionaries. After checking these definitions with the speakers’ judgements, the result is that of the intermediate forms whose meaning is related to the prefixed V, we get different patterns. First, one of the two forms may be non-existing, which can either be the unprefixed V (e.g. *crust*, *compass*) or the prefixed V (e.g. *entrain*, *engirdle*), the latter questioning my analysis, since there is no possible source for the [-c-m] features (i.e. the prefix is absent). Second, the intermediate form does not have a locative meaning (e.g. *list*, *trench*), which according to my analysis follows from the absence of the prefix, the source of the locative features. Third, there are some intermediate forms which have a locative reading (e.g. *snare*, *tangle*), clearly going against my proposal.

Of all locative intermediate forms which are listed in the dictionaries as having the same meaning as their prefixed version, only three are really synonyms for English speakers: *encode* (*code*), *encircle* (*circle*), and *entitle* (*title*). They all go against my analysis: there is no source for the locative features if the prefix is absent. According to English speakers, all the remaining intermediate forms which supposedly have the same meaning as their prefixed version are non-existing (e.g. *throne*, *shrine*), with the exception of four, namely *encipher*, *engraft*, *enshroud*, and *enwrap*. While the former do not pose any problem, the second ones do. Again, there is no source for the locative reading in *cipher*, *graft*, *shroud*, and *wrap*, the prefix being absent.

In short, all the unprefixed intermediate forms with a locative reading require some explanation. Although this set of Vs is small and could be disregarded (cf. in Appendix A, the section of denominal Vs lists all en+N Vs with a locative reading as well as all possible intermediate forms also having a locative interpretation), one still

wants some explanation for their behaviour, and for another more productive type of unprefixed denominal Vs with a locative reading. Consider the following location (50) and locatum (51) Vs (see footnote 12 which gives references for lists of other location and locatum Vs).

- | | | |
|---------|---|-------------|
| (50) a. | box _N , jail _N , kennel _N | R |
| | b. box _V , jail _V , kennel _V | [+c] [-c-m] |
| (51) a. | crown _N , curtain _N , chain _N | R |
| | b. crown _V , curtain _V , chain _V | [+c] [-c-m] |

Given that the prefix *en-* is the source of the [-c-m] features (the theme role) in 'change of location' Vs, there is no visible source for such features in (50-51). This problem would be resolved if a null prefix performed the function of the visible prefix. Is there any evidence to postulate a zero-prefix for English? Is there an abstract element responsible for the [-c-m] features present in these Vs? Padrosa (2005a) suggests that some historical reanalysis might have taken place, i.e. the prefix might have been attached to these Vs originally, and then for some reason, it was dropped, although the meaning remained the same. I pursue this idea here, and I suggest that the prefix originally attached to the V had the [-c-m] features still present in the prefix of Vs like *encase* and *enrage*. Historical data (cf. Marchand 1969) confirm my suggestion. All the forms in (50-51) were initially prefixed forms: *embox*, *enjail*, *enken- nel*, *encrown*, *encurtain*, and *enchain*.

Now it needs to be explained how all the forms which lost the prefix still have a [-c-m] role. I propose that the loss of the prefix has been a gradual process in which speakers have disassociated the [-c-m] features from the prefix and have relinked them to the base N (cf. autosegmental phonology, see e.g. Kenstowicz 1994, Roca et al. 2000). The prefix with no features of its own had no function in the word and was probably lost gradually. (Maybe phonological weakening helped to its loss). If that view is correct, one needs to explain how speakers can differentiate Vs of creation from locative Vs, because the [-c-m] features in both cases have the same source (the N), a question which will be discussed in the following section. Although the most productive type of denominal Vs in English seems to complicate the picture as for the source of the [-c-m] features, at the same time it provides some evidence to say that the Ø-suffix, apart from providing the [+c] role in Vs like (50-51), is the element responsible for the conversion of Ns to Vs.

The conclusion from this section is that the analysis proposed for Catalan *en+N* Vs can be maintained, but only for few denominal Vs in English, namely Vs of creation (e.g. *enslave*) and those prefixed Vs with a locative reading (e.g. *encase*, *enrage*). In both cases the [+c] role originates in the Ø-suffix, also responsible for the conversion (N→V). The [-c-m] features come from the reinterpreted R-role in Vs of creation and from the prefix in locative denominal Vs. However, another type of denominal Vs was found, viz. those that have no prefix but have a locative reading. For those, I proposed that the [-c-m] features are contained in the base N. Crucially, in all cases, the RHR is observed: the element containing a specific feature specification constitutes the rightmost element marked with those features.

3.3. Discussion

This section is mainly devoted to explaining some points left unresolved from the previous sections. First, I will explain how speakers can distinguish locatum Vs from location Vs given that they have the same structure. Second, I will present how English speakers can derive the locative reading in unprefixed Vs like *box*. Third, I will discuss how my analysis can explain the fact that the prefix *en-* seems to potentiate the suffix *-ment* if the prefix is not a head. Finally, HK's (1993, 1998, 2002) proposal will be briefly presented to see whether it can handle the data of my study satisfactorily.

As seen in the two previous sections, there are some Vs which can be interpreted as locatum and location, like *envinagrar* and *entangle* with the paraphrases 'to put N into/over something' and 'to put something into the N'. Although the division between the two readings may not be linguistically relevant, one has to explain the fact that speakers can differentiate the two and assign the appropriate meaning (either locatum or location) to any denominal Vs given a specific context. According to Clark & Clark (1979), the characterization of denominal Vs into locatum and location depends on their predominant features. If the source N denotes things which are conventionally placed with respect to other objects (i.e. placeables in their terminology), then the locatum reading will be derived. If the source N denotes things which are used as places with respect to which other objects are placed, we will get the location interpretation. Clark & Clark also note that some Vs may have more than one predominant feature, thus giving rise to Vs like *envinagrar* and *entangle*.

Kiparsky (1997) reaches a similar conclusion by a conceptually-knowledge based principle making use of the canonical use (instead of Clark & Clark's predominant features) of the N on which the V is built. He derives the following fixed meanings for the two locative relations (p. 482):

- (52) a. Locatum verbs: putting x in y is a canonical use of x.
 b. Location verbs: putting x in y is a canonical use of y.

Kiparsky explains that some Vs will be able to be interpreted either way if the object the source N denotes can have the two canonical uses, namely 'to be put on something' and 'to have something put on it'.

Although Clark & Clark and Kiparsky acknowledge the existence of Vs with two possible relations of location and explain them by the N having more than one predominant feature or canonical use respectively, nothing is said about how the speaker identifies which of the two locative relations is meant by a denominal V given a context. I assume speakers will resolve these ambiguities by looking at the context in which the V is uttered and by selecting the interpretation most relevant accordingly. This view is in line with Relevance Theory (RT) (cf. e.g. Sperber & Wilson 1986/1995, Wilson 1994, Wilson & Sperber 2004), which is based on some simple assumptions. Every utterance has several linguistically possible interpretations, not all of which occur to the hearer simultaneously. Hearers are assumed to be equipped with a criterion for evaluating (accepting or rejecting) interpretations, as they occur to them. This criterion excludes all interpretations, except for one at most. So, the hearer can assume that the first acceptable interpretation they find is the intended

one. In other words, the hearer considers interpretations in order of accessibility and stops when they find one that is relevant enough to satisfy their expectation of relevance, with the result that the first satisfactory interpretation is the only acceptable one. The criterion is ultimately based on the cognitive principle of relevance: human cognition is relevance-oriented (Wilson 1994: 17).

Following a relevance-theoretic account, when listeners are presented with the utterance 'to shelve the books' for example (*shelve* being a V that admits the two locative readings), the first interpretation they will consider will be that of 'putting books on the shelves' and not the other way round. Similarly, if they are given the utterance 'to shelve the closet', the first satisfactory interpretation they will find will be that of 'putting shelves in the closet', and not 'putting the closet on shelves'. By simply looking at the direct object of the V, hearers can pick out the interpretation they think the speaker intended on that occasion, the most relevant interpretation for them. In 'to shelve the books/closet', the interpretations 'putting shelves on the books' and 'putting the closet on shelves' are not relevant enough to satisfy the hearer's expectation of relevance (and will be rejected). In short, the hearer can readily identify which locative relation is intended (locatum or location) within RT.

Another question which remained unresolved from the previous section was how English speakers (and Catalan speakers to a much lower degree) can derive the locative reading in prefixless Vs like *box*, *crown*, *circle* and *snare* (the two last Vs being intermediate forms in the derivations of their prefixed version), if the prefix *en-* is the element responsible for such reading ([*-c-m*] features). I suggested that the [*-c-m*] role once associated with the prefix was relinked to the base N, with the consequence that the prefix was gradually lost. The prefix had no function to perform, i.e. the semantic content it had before (it contained the [*-c-m*] role) was affected and so was its productivity. As a result, native speakers refrained themselves from coining new members with the prefix, a tendency which led to its disappearance. If we look at numbers, the locative Vs without prefix largely exceeds those with prefix. That proposal seems to find further confirmation in the fact that locatum/location Vs that once had a prefix now do not have it any more. What I am implying here is a contrast between Catalan and English with respect to the productivity of *en-*prefixation. While it seems that this morphological process was and is still active in Catalan, it has become unproductive in English.

If it is true that the [*-c-m*] features of English locative denominal Vs and Vs of creation both come from the same source, namely the base N, one also has to explain how English speakers differentiate the two. Again, I think a relevance-theoretic account has the answer. The hearer will interpret a denominal V as locative if that is the first interpretation that satisfies their expectation of relevance. Similarly, a V will be interpreted as V of creation if that is the first acceptable interpretation for the listener. To illustrate the point, consider *to box the apples*. The first satisfactory interpretation will not be that of a V of creation, i.e. 'to make a N (box)', but that of a locative V (a location V in this case), 'to put the apples in the N (box)'. In short, one can readily pick out the interpretation intended by the speaker within RT.

A different question which also needs to be addressed in the Discussion is how my analysis can explain the fact that the suffix *-ment* seems to be potentiated by the prefix *en-* (if the prefix is not a head with respect to the category-changing ability, as I have defended). It is generally agreed that affixes may be sensitive to other affixes in

their base (cf. Fabb 1988, Hoeksema 1988). In line with this generalization some authors (cf. Aronoff 1976, Williams 1981a, Scalise 1984) have proposed that the suffix *-ment* attaches most productively to Vs of the form *en+X* (e.g. *encroachment*), claiming that the prefix *en-* potentiates the suffix *-ment* because the prefix is the head. One could claim that the same phenomenon exists in Catalan, given the large quantity of words with the form *en+X+ment* (e.g. *encoratjament* ‘the act of encouraging’). The GDLC lists more than 250 words with this form. However, the prefix *en-* is not the only element able to potentiate the suffix *-ment*. In English the prefix *be-* has the same ability (e.g. *bedazzlement*). The CCED and COD include more than 50 *en+X+ment* forms, and more than 30 *be+X+ment* forms. Although the number of the latter is lower, it is still significant. In Catalan it seems that several prefixes like *a-* and *des-* can also potentiate the suffix *-ment* (e.g. *allargament* ‘the act of lengthening’, *descargolament* ‘the act of unscrewing’). In this case, the GDLC lists more than 200 words for *a+X+ment* forms and more than 150 for *des+X+ment* forms, both numbers being substantial. All these numbers (always relatively speaking) seem to indicate that the suffix *-ment* is not favoured by a particular prefix but simply by the presence of a prefix (see Scalise 1988b for the same conclusion for Italian). To explain this fact I can only suggest that *-ment* has a particular feature [F] which needs to be satisfied and that the prefix has the relevant feature [F]. However, I am aware that this suggestion is only descriptively adequate since it explains why *-ment* seems to be potentiated by *en-*, but it does not say anything about the nature of the feature.

After discussing some points left unresolved from the previous sections, and before ending the present one, now I would like to briefly present another proposal, i.e. HK’s (1993, 1998, 2002), to see whether it can handle the data satisfactorily. HK adopt a syntactic approach to the representation of lexical argument structure. Vs are derived by conflation²⁰ of a N or A into an empty phonological V base, thus giving it phonological content. The structural types of lexical argument structure relevant here are those associated with the morphosyntactic category A and N, given that now I will focus on how HK’s theory can derive deadjectival and denominal Vs in English and Catalan. Recall that in V-to-V prefixation no systematic pattern was found, which explains its omission in the following discussion. Although HK’s theory may seem to cope with the data adequately at first sight, there are some questions which cannot be answered within their syntactic approach.

As already said, conflation explains the formation of deadjectival Vs. The phonological matrix of the A replaces that of the V, which can be empty like in *clean* (53a), or partially empty as in *enrich* which has a prefix or *thicken* which has a suffix. For the latter cases, HK assume that the host V is bipartite, consisting of an empty phonological matrix together with an overt matrix corresponding to that of the prefix or suffix (53b, c) (HK 1998: 85).

²⁰ Note that the discussion that follows is based on HK (1993, 1998). The same results, though, would be obtained by using HK’s more recent version. Let me just point out one remarkable difference between their earlier and later accounts, namely their use of the term conflation. In the more recent version, it does not refer to a movement operation. In HK’s terms, ‘it is merely the binding relation that holds between the semantic features of a V (phonologically overt now) and features of the nominal head of its complement’ (HK 2002: 103).

- (53) a. V
 / \
 V A
 | |
 [Ø] clean
- b. V
 / \
 V A
 / \
 pref [Ø]
- c. V
 / \
 V A
 / \
 [Ø] suf

HK's treatment of deadjectival Vs can then explain the three types found in English. Regarding Catalan deadjectival Vs, they can also be accounted for by (53). The structure in (53a) would explain Vs like *agrir* (<*agre*_A 'sour'), and the structure in (53b) could derive Vs like *endolcir(-se)* (<*dolç*_A 'sweet'). Although the English and Catalan deadjectival Vs can be explained on the whole, Vs like *embolden* cannot be derived, because they involve simultaneous prefixation and suffixation, implying ternary branching. Although this weakness could be solved by appealing to the non-productivity of the type *embolden*, HK still have to explain it. The type *embolden* was once an active process.

Denominal locative Vs present a similar scenario. The V will be bipartite in the case of Vs like *encase* and *encaixar* (<*caixa*_N 'box'), and it will not be so in the case of *box* and *registrar* (<*registre*_N 'register'). The same problem presented for deadjectival Vs is also present now. Again, HK need some account for the existence of Vs like *enlighten*, whose formation was once productive. In addition, HK suggest that the distinction between location and locatum Vs is not one of structure (which is what one would expect from their account) but derives from the semantic properties of the head. Apparently, their P (the prefix in my terminology) distinguishes terminal and central coincidence. If that is true, HK's claim that the properties of word meaning follow from syntactic constraints can no longer be observed. In addition, I assume HK would use some kind of semantics to derive denominal Vs expressing creation (e.g. *enslave*, *enraiar*) given that P can only express terminal and central coincidence and there is no other element available in their analysis to account for the correct reading.

Also, HK would probably resort to a semantic account to explain the fact that some Vs can be interpreted as a locative V and as a V of creation. For instance, the V *encoixinar* could be interpreted as 'putting cushions in a place' and as 'making a cushion'. In the former interpretation the P would be the element responsible for such reading, but in the latter there would be no source for such reading, unless they resort to some semantics.

An additional problem for their analysis is the origin of Θ -roles. According to my approach, in the case of denominal Vs the [-c-m] features originate in the prefix in locative Vs, but in the N's reinterpreted R-role in Vs of creation. (Recall that this picture derives some English denominal Vs (the prefixed ones) and most Catalan denominal Vs). If my analysis is correct, then HK's approach cannot be on the right track. They cannot explain the non-uniform source of Θ -roles given their adherence to the UTAH²¹ (cf. Baker 1988), according to which there is direct mapping between thematic roles and syntactic structure. More specifically, each thematic role must be

²¹ UTAH stands for Uniformity of Theta Assignment Hypothesis and is defined by Baker (1988) in the following terms: 'Identical thematic relationships between items are represented by identical structural relationships between those items at the level of D-structure'.

linked to a single position in D-structure. A related problem that is a consequence of HK's configurational model of thematic relations is that the lower thematic VP only allows two theta-roles (generally the [-c-m] (theme) role and [-m] (locational) role), with the result that other roles, such as [+c-m] (instrument), cannot be represented although they cannot be considered adjuncts. In addition, the role assigned to the subject cannot be represented either.

In short, I think enough problems have been found in HK's account to pursue their approach here (see e.g. Di Sciullo 1997, Kiparsky 1997, Stiebels 1998 for other criticisms). The conclusion is that a syntactic account has not proved sufficient to account for the data presented in the previous sections. Next a brief summary and the main conclusions of my study will be presented.

4. Conclusions

In this paper, I considered a potential class of counterexamples to the RHR, namely the class of prefixes in English and Catalan. More specifically, I looked at how the prefix *en-* present in the two languages apparently converts As and Ns to Vs in a productive way. However, on the basis of A/N-to-V prefixations, I argued that complex words derived by *en-*prefixation are not really exceptions to the (Relativized) RHR.

I showed that a \emptyset -suffix is responsible for the conversion of Ns and As to Vs, a process which takes place before the prefix is attached, thus not incurring any violation to the RHR (see e.g. Neeleman & Schipper 1992, Gràcia 1995, Stiebels 1998 for similar views). The crucial argument for the postulation of the \emptyset -suffix comes from the Θ -grid of the Vs. The \emptyset -suffix is responsible for the [+c] role, whose presence would be unaccounted for without the postulation of the conversion-suffix. The \emptyset -suffix also accounts for the observation that *en-X* words are always verbal.

As for the role of the prefix *en-*, we have seen that it is responsible for the [-c-m] role in the case of *en+N* Vs with a locative meaning. However, I have shown that the formation of locative *en+N* Vs is no longer productive in English. Although the presence of *en-* was once felt compulsory for the formation of locative denominal Vs both in English and Catalan, which according to my analysis follows from the fact that the prefix gives the locative reading to the V, there is now a contrast between speakers of the two languages. Catalan speakers still require the presence of the prefix to express both locatum and location N-based Vs suggesting that *en-*prefixation is still an active process. In contrast, English speakers prefer denominal locative V without prefix, which I explained by disassociation of the [-c-m] role from the prefix and re-associating it to the base N.

Although the Rel. RHR can still be maintained for English unprefixated Vs, because the N constitutes the rightmost element specified for those features, the fact that some Vs can have a locative interpretation and a creation reading becomes difficult to explain. Both interpretations depend on the [-c-m] role now present in the same node, the base N. To solve this problem, I make use of a relevance-theoretic account, according to which hearers evaluate interpretations in order of accessibility (e.g. context, disambiguation, etc.) and stop considering them when their expectation of relevance is satisfied. The result is that the first adequate interpretation

satisfying the hearer's expectation of relevance is the only possible one given a specific context. When a listener is presented with a V which can be interpreted with a locative and creation reading, they will readily pick out the interpretation intended by the speaker, the only satisfactory interpretation on a particular occasion. I also showed that RT can explain the distinction between the location and locatum interpretations of some Vs.

If the use of semantics and pragmatics is necessary to explain some basic contrasts which otherwise would remain a mystery, a syntactic theory of argument structure like that of HK's (1993, 1998, 2002) is not sufficient. In fact, HK themselves recognize the need for some semantics in their account. For instance, they admit that the prefix distinguishes terminal and central coincidence, clearly two semantic notions.

Another reason to reject HK's approach is their direct mapping between configurational positions and specific roles. As I have shown, semantics cannot be read off the structure. Recall that the [-c-m] features in Catalan denominal Vs can come from the prefix (when there is a locative reading) or the base N (when a creation reading is implied). This limitation shows that the framework adopted here, Reinhart's (2000, 2001), is superior to that of HK's at least in the sense just discussed.

Although the present study has adopted Reinhart's theta-system, her approach to reflexives has been rejected. On the basis of Catalan deadjectival and denominal Vs (e.g. *endolcir(-se)*, *encoratjar(-se)*), the approach of reflexives as unaccusatives (cf. Grimshaw 1990, Sportiche 1998) has proved to deal with the data more satisfactorily than the view which favours reflexives as unergative entries (cf. Reinhart & Siloni 1999). In most of the cases, the external argument is clearly reduced (i.e. the [+c] role undergoes reduction in Vs like *endolcir(-se)*). In other cases, it is hard to tell which argument has undergone reduction (e.g. *encaputxar(-se)*). Crucially, there are no cases of clear internal reduction. Hence, I proposed that it is the external argument that is always reduced.

Another question to which I intended to provide an answer in my study was whether a Θ -role percolation approach to the inheritance of thematic information (cf. Booi 1988, Levin & Rappaport 1988, Gràcia 1992, 1995 and Neeleman & Schipper 1992) could be confirmed. I think the data have amply corroborated this question as well as Mateu's (2001a, 2002) view of complex denominal Vs in German, according to which the preverb (the prefix in my case) is part of the resulting thematic structure, thus also contributing to the Θ -grid of the predicate.

As for the remark made by several authors (see e.g. Williams 1981a) that the prefix *en-* potentiates the affix *-ment* in English and Catalan because the former is a head, I argued that it is not the presence of *en-*, but simply the presence of any prefix which triggers the suffix *-ment*. For that fact I suggested that the suffix *-ment* has a certain feature [F] which needs to be satisfied, and that the prefix *en-* has the relevant feature [F] (cf. Fabb 1988). Obviously, this option needs to be further investigated to find the real feature behind the potentiation of the suffix *-ment*.

Other questions also need more study. One has to do with the existence of apparent synonyms with a locative reading in the English data (e.g. *(en)circle*). Do they show that the [-c-m] features of the prefix are still available and that the process of relinking these features to the base N has not died out completely? A further question which also needs to be addressed is whether the process of relinking the [-c-

m] role to the base N also takes place in Catalan? If it does, why is delinking more common in English than in Catalan? For the moment I leave all these questions for future research.

Appendix A

This appendix classifies the *en*-prefixed Vs in English into three different groups depending on the base on which they are built: (a) a V, (b) an A, and (c) a N. The base from which the V is derived is given within parentheses after the prefixed V. The Vs in each group have been divided into semantic paraphrases. The Vs in (a) have not been classified due to the lack of semantic regularity. On the whole, all deadjectival Vs below follow the semantic pattern 'to make (something/somebody) A'. Regarding denominal Vs, they have been divided into four groups: location Vs (A), locatum Vs (B), Vs which can have the two interpretations (A/B), and Vs of creation (C). Each group includes subgroups where intermediate Vs have been listed. According to my analysis, in the formation of prefixed deadjectival and denominal Vs there is an intermediate stage in which the A and N have become a V, but the prefix is not present as yet. These intermediate forms have been grouped depending on whether they have the same meaning as that of the prefixed version or a related one. Note that in the group of locatum Vs, the locatum can either be a physical object (e.g. *venom* in *envenom*) or an abstract one (e.g. *danger* in *endanger*). Lexicalised prefixed forms have not been taken into account in the study.

$V \rightarrow [en+V]_V$

Enact (act), enchant (chant), enclose (close), engrave (grave), enjoin (join), enlist (liven) ensue (sue), and entreat (treat).

$A \rightarrow [en+A]_V$

Trans [+c] [-c-m] 'to make A'

Enable (able), endear (dear), enfeeble (feeble), enlarge (large), ennoble (noble), enrich (rich), and ensure (sure).

$N \rightarrow [en+N]_V$

A) *Trans* [+c] (usually [+c+m]) [-c-m] 'to put something around/in/onto/towards N'
Location Vs.

Encapsulate (capsule), encase (case), encode (code), encyst (cyst), engorge (gorge), enlist (list), enmesh (mesh), enplane (plane), enrobe (robe), enshrine (shrine), enshroud (shroud), ensile (silo), ensnare (snare), enthrone (throne), entomb (tomb), entrain (train), entrance (trance), entrap (trap), entrench (trench), and enurn (urn).

- *Intermediate Vs which have meanings related to the prefixed Vs (A.1):* Engorge (gorge), enlist (list), enmesh (mesh), ensnare (snare), entrap (trap), entrain (train), and entrench (trench).
- *Intermediate Vs which have the same meaning as the prefixed Vs (A.2):* Encode (code), enrobe (robe), enshrine (shrine), enshroud (shroud), enthrone (throne), entrance (trance), and enurn (urn).

B) *Trans [+c]* (usually [+c+m]) [-c-m] 'to put N around/in/into/on something/somebody'
Locatum Vs.

Encircle (circle), encompass (compass), encrust (crust), endanger (danger), enfold (fold), enforce (force), engirdle (girdle), engraft (graft), engulf (gulf), enlighten (light), enrage (rage), entrust (trust), and evenom (venom).

→ [+c] (usually [+c+m]) [-c-m] 'to give N' (= 'to put N in somebody')

Encourage (courage), enfeoff (fief), enfranchise (franchise), enrapture (rapture), and entitle (title).

➤ *Intermediate Vs which have meanings related to the prefixed Vs (B.1):*

Encompass (compass), encrust (crust), enfold (fold), enforce (force), engirdle (girdle), enlighten (light), enrage (rage), and entrust (trust).

➤ *Intermediate Vs which have the same meaning as the prefixed Vs (B.2):*

Encircle (circle), engraft (graft), enfranchise (franchise), and entitle (title).

A/B) Some verbs fit into either group (A or B):

Although all 'locative' verbs have been placed either in group (A) or (B), some could be argued to belong to both groups. For instance, consider the verbs *encrypt* (*crypt*), *encipher* (*cipher*), *entangle* (*tangle*), and *enwrap* (*wrap*).

➤ *Intermediate Vs which have meanings related to the prefixed Vs (A/B.1):*

Entangle (*tangle*).

➤ *Intermediate Vs which have the same meaning as the prefixed Vs (A/B.2):*

Encipher (*cipher*), and enwrap (*wrap*).

C) N as a result 'to make N' Vs of creation

Encamp (camp), encash (cash), enslave (slave), and envision (vision).

➤ *Intermediate Vs which have meanings related to the prefixed Vs (C.1):*

Envision (vision)

➤ *Intermediate Vs which have the same meaning as the prefixed Vs (C.2):*

Encamp (camp), encash (cash), and enslave (slave).

Appendix B

Appendix B classifies the *en*-prefixed Vs in Catalan following the same criteria established in Appendix A. Complex words derived by *en*- have been divided into three groups, these being determined by the category of the base (V, A, N). Concerning the prefixed Vs whose source is a V, they have not been divided into different semantic groups due to its variability in meaning. The base V is included within parentheses after each prefixed V, and due to its reduced number, all *en*-prefixed Vs have been included on the first list. That is, the first list does not dis-

tinguish prefixed Vs with a lexicalized meaning. For instance, speakers do no longer associate the Vs *encantar* or *endreçar* with *cantar* and *dreçar* respectively. Also, there are a few pairs of Vs (i.e. with and without the prefix) of which the speaker only uses one form and not the other for different reasons: one of the two forms may belong to Old Catalan (e.g. *encercar*, *enseguir*) or to one specific dialect (*enfondre*, *engronsar*, *enxautar-se*) and these have also been included on the first list. Finally, there is a third group of Vs (i.e. with and without the prefix) of which speakers do not use any of the two forms and these have not been filtered out from the first list either (e.g. *enforollar* (*forollar*), *ensulsi(a)r-se* (*sulsir*)). However, the first list is followed by a second list from which all the previous forms have been removed. The reduced number of verbs on the second list shows that it is very difficult to find a systematic pattern similar to those found in deadjectival and denominal Vs.

In the classification of deadjectival Vs, the A from which the V is derived is given in the masculine form within parentheses after the V. The clitic *se* within parentheses () indicates that the V can be either transitive (without *se*) or unaccusative (with *se*). The Vs with clitic can only be unaccusative and those without are mostly transitive. The clitic *se* within square brackets [] indicates that the V can be used transitively (without the clitic), and intransitively (as an unaccusative) either with the clitic or without. The same holds for denominal Vs.

On the whole, all deadjectival Vs below follow the semantic pattern 'to make (something/somebody) A' and '(something/somebody) becomes A', when used transitively and intransitively, respectively. If some Vs slightly differ from this pattern (one on this list), their behaviour can still be explained. For example, the V *enaltir* (*alt*) 'praise' 'tall' can be understood as 'making someone high/putting someone in a high position by prasing him'. Regarding denominal Vs, four groups can be distinguished: location Vs (A), locatum Vs (B), Vs which can have the two previous patterns (A/B), and Vs of creation (C). Each group includes subgroups where intermediate Vs have been listed. Recall that according to my analysis in the formation of prefixed deadjectival and denominal Vs there is a stage in which the A and N have become a V, but the prefix is still not present. These intermediate forms have been grouped depending on whether they have the same meaning as that of the prefixed V or a related one. Note that in the group of locatum Vs, the locatum can either be a physical object (e.g. *caputxa* 'hood' in *encaputxar*) or an abstract one (e.g. *amor* 'love' in *enamorar*).

Deadjectival Vs like *fosquejar*, *groguejar*, *lluentejar* and *rossejar* have not been taken into account, since they all contain the suffix *-ej-* between the adjectival base and the inflectional morpheme. The same applies to denominal Vs and prefixed Vs whose source is already verbal (i.e. verbs like *encamellar* (<*cama*_N) and *endormiscar-se* (<*dormir*_V) have also been avoided because they contain suffixes (*-ell-* and *-isc-* respectively) intervening between the nominal/verbal base and the inflectional element, although most of them do not seem to affect the resulting argument structure of the V).

Lexicalized deadjectival and denominal Vs have not been included in this survey. For instance, denominal Vs like *ensenyar*<*senya* and *enviar*<*via* have been disregarded.

$V \rightarrow [en+V]_V$

List 1

Encantar (cantar), encarregar (carregar), encercar (cercar), encavalcar (cavalcar), encarregar (carregar), encobrir (cobrir), encomanar (comanar), encórrer (córrer), endreçar (dreçar), endurar (durar), endur-se (dur), enfondre (fondre), enfonyar (fonyar), enforollar (forollar), enfugir-se (fugir), engronsar (gronxar), enlluir (lluir), ennavegar-se (navegar), enveixinar (veixinar), enretirar (retirar), enseguir (seguir), ensibornar (subornar), ensomniar (somniar), ensostrar (sostrar), ensulsi(a)r-se (sul-sir), entallar (tallar), entorcir (tòrcer), entravessar (travessar), envolar-se (volar), and enxautar-se (xautar-se).

List 2

Encarregar (carregar), encloure (cloure), encobrir (cobrir), endur-se (dur), enretirar (retirar), ensibornar (subornar), and entravessar (travessar).

$A \rightarrow [en+A]_V$

Transitive [+c] [-c-m] 'to make A' / Reflexive 'to become A' [-c-m]

Enagrir(-se) (agre), enaltir (alt), enardir(-se) (ardit), enasprar(-se)/enasprir(-se) (aspre), encalbir(-se) (calb), encalentir (calent), encalmar-se (calm), encanutir (canut), encarir[-se] (car), encegar (cec), encertir(-se) (cert), encoixir(-se) (coix), encrespar(-se) (cresp), encruar-se (cru), encruelir(-se) (cruel), encuriosir (curiós), endoblar (doble), endoblar-se (doble), endolcir(-se) (dolç), endolentir(-se) (dolent), endropir(-se) (dropo), endurir(-se) (dur), enfadeir(-se) (fat), enfellonir(-se) (felló), enfereir-se (fer), enferestir-se (ferest), enferotgir-se (ferotge), enfolllir[-se] (foll), enfondir(-se) (fondo), enfortir(-se) (fort), enfoscar[-se] (fosc), enfosquir[-se] (fosc), enfranquir (franc), enfredar(-se) (fred), enfredolicar(-se) (fredolic<fred), engalanar (galà), engallardir(-se) (gallard), engallofir(-se) (gallof), engandulir(-se) (gandul), engegantir (gegant), engelosir(-se) (gelós), engolosir (golós), engordir(-se) (gord), engormandir(-se) (gormand), engrandir[-se] (gran), engrevir(-se) (greu), engroguir(-se) (groc), engrossir(-se) (gros), enguerxir(-se) (guerxo), enjogassar(-se) (jogasser<joc), enjovenir (jove), enllefernar(-se) (llefre), enllefiscar(-se) (llefiscós), enllepissar(-se) (llepissós<llepar), enllepolir(-se) (llepol), enllestir(-se) (llest), enlletgir(-se) (lleig), enllordar(-se) (llord), enlluentir (lluent), enllustrar-se (llustre), ennegrir(-se) (negre), ennoblir(-se) (noble), ennovar(-se) (nou), enrancir(-se) (ranci), enrarrir(-se) (rar), enrellentir(-se) (rellent), enrigidir(-se) (rígid), enriquir(-se) (ric), enrobustir(-se) (robust), enroguir(-se) (roig), enronquir(-se) (ronc), enrossir(-se) (ros), ensalvatgir(-se) (salvatge), enseriosir-se (seriós), ensordir(-se) (sord), ensuperbir(-se) (superb), ensutzar/ensutzir/ensutzir* (sutze), entebeir(-se) (tebi), entebionar (tebió), entendre(-se) (tendre), enterbolir(-se) (tèrbol), entoixar (toix), entorpir (*Spanish* torpe), entossudir-se (tossut), entristar(-se)* (trist), entristir(-se) (trist), entumir(-se) (túmid), envalentir(-se) (valent), envanir(-se) (va), envellir(-se) (vell), enverdir(-se) (verd), enverinosar (verinós), envermellir(-se) (vermell), envilanir(-se) (vilà), envilir(-se) (vil), and enxiquir (xic).

*Note that the verbs *ensutzar/ensutzir/ensutzir* and *entristar(-se)* belong to Old Catalan.

➤ *Here is a list of some of the existing intermediate Vs:*

Agre → agrir(-se) → enagrir(-se)
 Cec → cegar → encegar
 Corb → corbar(-se) → encorbar(-se)
 Cresp → crespar → encrespar(-se)
 Doble → doblar(-se) → endoblar
 Guerxo → guerxar(-se) → enguerxir(-se)
 Rellent → rellestar → enrellestar(-se)

$N \rightarrow [en+N]_V$

A) *Trans [+c] (usually [+c+m]) [-c-m] 'to put something around/in/onto/towards N'*
Location Vs.

enarbrar(-se) (arbre), encabassar (cabàs), encadellar (cadell), encaixar (caixa), encaixonar (caixó), encalaixonar (calaixó), encambrar(-se) (cambra), encaminar(-se) (camí), encanalar (canal), encanastrar (canastra), encanonar (canó), encanyonar (canyó) encapçalar (capçal), encapsar (capsa), encapsular(-se) (càpsula), encarcanyar (carcanyell), encarcerar (càrcer), encarrerar(-se) (carrera), encarrilar(-se) (carril), encartar (carta), encartutxar (cartutx), encasar (casa), encasellar (casella), encastellar(-se) (castell), encauar(-se) (cau), encelar-se (cel), encinglar-se (cingle), encistellar (cistell), enclaperar-se (clapera), enclaustrar(-se) (claustre), enclotjar(-se) (clot), encoblar (cobla), encofinar (cofi), encofrar (cofre), encofurnar(-se) (cofurna), encollar (coll), encorrallar (corral), encossiar (cossi), encotxar-se (cotxe), encovar-se (cova), encovenar (cove), encubar (cup), encubellar (cubell), endollar (dolla), endossar(-se) (dors), enfilosar (filosa), enfonsar(-se) (fons), enforatar (forat), enfornar (forn), enfotjar (fotja), enfundar (funda), engabiar(-se) (gàbia), engaltar (galta), engalzar (galze), engargamellar (gargamella), engarjolar (garjola), engatjar (gatge), englotir(-se) (glotis), engolar(-se) (gola), engolir(-se) (gola), engorgar-se (gorg), engorjar(-se) (gorja), engraelar (graella), engranar (graner<gra), enguardiolar (guardiola), enguierar (guier), enjovar (jou), enllistar (llista), enllitar(-se) (llit), enllomar (lloç), enqueixalar (queixal), enquistar-se (quist), enregistrar (registre), enriuar (riu), enrocar(-se) (roca), enrodar (roda), enrolar(-se) (rol), ensacar (sac), ensarriar (sàrria), ensarrionar (sarrió), ensarronar (sarró), ensenderar (sender), ensitjar (sitja), ensobrar (sobre), ensolcar (solc), ensotar(-se) (sot), entaular (taula), entinar (tina), entrampar(-se) (trampa), entrapar (trapa), entrullar (trull), envaixellar (vaixell), envalisar (valisa), envasar (vas), and envergar (verga).

➤ *Intermediate Vs which have meanings related to the prefixed Vs (A.1):*

Arbre → arbrar(-se) → enarbrar(-se)
 Camí → caminar → encaminar(-se)
 Càpsula → capsular → encapsular(-se)
 Clot → clotar → enclotar(-se)
 Coll → collar → encollar
 Llista → llistar → enllistar
 Llit → llitar → enllitar(-se)
 Queixal → queixalar → enqueixalar

Roda → rodar → enrodar
 Solc → solcar → ensolcar
 Trull → trullar → entrullar

➤ *Intermediate Vs which have the same meaning as the prefixed Vs (A.2):*

Registre → registrar → enregistrar

B) Trans [+c] (usually [+c+m]) [-c-m] 'to put N around/in/into/on something/somebody'
Locatum Vs

enaiguar(-se) (aigua), enamorar(-se) (amor), enarçar (arç), enartar (art), enasprar(-se) (aspre), encabestrar (cabestre), encabironar (cabiró), encadarnar (cadarn), encadenar(-se) (cadena), encadirar (cadira), encaironar (cairó), encalcinar (calcina<calç), encalimar (calima), encalitzar(-se) (calitja), encalmar-se (calma), encamisar(-se) (camisa), encanyar (canya), encanyissar (canyís<canya), encaparrar(-se) (caparra<cap), encapellar(-se) (capell), encaperonar(-se) (caperó), encaperullar(-se) (caperull), encaperutxar(-se) (caperutxa), encapirotar(-se) (capirot), encapotar(-se) (capot), encapotar (capota), encapritxar(-se) (capritx), encapullar(-se) (capulla), encaputxar(-se) (caputxa), encaramel·lar (caramel), encarbonar(-se) (carbó), encarestiar (carestia), encasquetar (casquet), encatifar (catifa), encausar (causa), encendrar (cendra), encerar (cera), encerclar (cercle), encercolar (cèrcol), encimbellar(-se) (cimbell), encimolsar (cimolsa), encintar (cinta), enciriar (ciri), enclavar (clau), enclavillar (clavilla), encobertar (coberta), encobertorar (cobertora), encoblar (coble), encofiar (còfia), encoixinar (coixí), encolar (cola), encolerir-se (còlera), enconxar (conxa), encoratjar(-se) (coratge), encordar (corda), encordillar (cordill), encordonar (cordó), encortinar (cortina), encotillar (cotilla), encotonar (cotó), encrestar (cresta), encrococar (croca), encrostar(-se) (crosta), encrostimar(-se) (crostim), encuirar (cuir), encuirassar (cuirassa), enderiar-se (dèria), endeutar(-se) (deute), endiablir (diable), endimoniar (dimoni), endogalar (dogal), endolar (dol), endomassar (domàs), endosserar (dosser), endrapar (drap), enfaixar (faixa), enfardar (farda), enfarinar (farina), enfebrar-se (febre), enferrar (ferro), enferritjar-se (ferritja), enfervorir(-se) (fervor), enfeudar (feu), enfilar (fil), enflocar(-se) (floc), enflorar(-se) (flor), enfocar (focus), enfredorar(-se) (fredor<fred), enfredorir(-se) (fredor<fred), enfrenar (fre), enfuriar(-se) (fúria), enfurir (fúria), enfusellar (fusell), enfustar (fusta), engafar (gafa), engafetar (gafet), engalbar (galba), engallinar (gallina), engalonar(-se) (galó), engalvanir(-se) (galvana), engandallar (gandalla), enganxar (ganxo), engarlandar (garlanda), engarroter (garrot), engassar (gassa), engavatxinar (gavatxi), engelabrir-se (blend of *gel* + *gebre*), engolfar (golfo), engomar (goma), engranar (gra), engravar (grava), engredar (greda), engreixar (greix), engreixinar (greixina<greix), engrescar(-se) (gresca), engrillonar (grilló), engronyar (grony), engualdrapar (gualdrapa), enguantar(-se) (guant), enguixar (guix), enherbar(-se) (herba), enjardinar (jardí), enjoiellar (joiell), enjoncar (jonc), enjovar (jou), enjullar (jull), enjuncar (junc), enlacrar (lacre), enllagrimar-se (llàgrima), enllaminir (llamí), enllandar (llanda), enllangorir(-se) (llangor), enllardar(-se) (llard), enllardonar (llardó<llard), enllatar (llata), enlleganyar-se (lleganya), enlliçar (lliç), enllistonar (llostó), enllosar (llosa), enllotar(-se) (lлот), enllustrar(-se) (llustre), enneguitar-se (neguit), ennigular-se (nígul), ennuvolar(-se) (núvol), enorgullar(-se) (orgull), enorgullir(-se) (orgull), enquimerar(-se) (quimera), enquitrancar (qui-

trà), enrabiarse (ràbia), enrajar (raig), enrajolar (rajola), enramar (ram), enramellar (ramell<ram), enrampar(-se) (rampa), enrandar (randa), enredoltar (redolta), enredortar (redorta), enreixar (reixa), enriallar(-se) (rialla), enrivetar (rivet), enrogallar-se (rogall), enrondar (ronda), enrosar(-se) (ros), enrovinar (rovina), enrubinar (rubina), ensabonar (sabó), ensafranar (safrà), ensaginar (sagí), ensagnar(-se) (sang), ensalivar(-se) (saliva), ensamarrar-se (samarra), ensellar (sella), ensementar (sement), enserrellar (serrell), ensetinar (setí), enseuar (sèu), ensivellar (sivella), ensucrar (sucre), ensulfatar (sulfat), ensutjar (sutja), entacar (taca), entapissar (tapís), entarimar (tarima), entatxar (tatxa), entaulellar (taulell<taula), entelar(-se) (tel), entelar (tela), entendre (tenda), entenebrar(-se) (tenebra), entenebrir(-se) (tenebra), enteranyinar-se (teranyina), enterrosar(-se) (terròs), entintar (tinta), entoïar (toïa), entovar (tova), entuixegar (túixec), enturar (turo), envelar (vel), envelar (vela), envellutar (vellut), enverdescar (verdesca), enverinar(-se) (verí), envermellonar (vermelló), envernissar (vernís), envescar(-se) (vesc), envetar (veta), envidrar (vidre), envidriar (vidre), envigorir(-se) (vigor), envinar (vi), envinyar (vinya), enviollar (violla), enviscar(-se) (visc), envitrallar (vitral), enxarolar (xarol), and enxavetar (xaveta).

➤ *Intermediate Vs which have meanings related to the prefixed Vs (B.1):*

Calma → calmar(-se) → encalmar-se
 Carbó → carbonar → encarbonar(-se)
 Cendra → cendrar → encendrar
 Clau → clavar → enclavar
 Corda → cordar → encordar
 Ferro → ferrar → enferrar
 Fil → filar → enfilar
 Fre → frenar → enfrenar
 Garrot → garrotar → engarrotar
 Gra → granar → engranar
 Greix → greixar → engreixar
 Guix → guixar → enguixar
 Ros → rosar → enrosar(-se)
 Sagí → saginar → ensaginar
 Sang → sagnar → ensagnar(-se)
 Saliva → salivar → ensalivar(-se)
 Taca → tacar → entacar

➤ *Intermediate Vs which have the same meaning as the prefixed Vs (B.2):*

Aspre → asprar → enasprar(-se)
 Cabestre → cabestrar → encabestrar
 Cairó → caironar → encaironar
 Cercle → cerclar → encerclar
 Cèrcol → cercolar → encercolar
 Cinta → cintar → encintar
 Clavilla → clavillar → enclavillar
 Cuirassa → cuirassar → encuirassar
 Drap → drapar → endrapar

Faixa → faixar(-se) → enfaixar
 Lacre → lacrar → enlacrar
 Lustre → llustrar → enllustrar(-se)
 Orgull → orgullar(-se) → enorgullir(-se)
 Rivet → rivetar → enrivetar
 Setí → setinar → ensetinar
 Sulfat → sulfatar → ensulfatar
 Tela → telar → entelar
 Verí → verinar → enverinar(-se)
 Xarol → xarolar → enxarolar

A/B) Some verbs fit into either group A or B:

Enastar (ast), encarar(-se) (cara), encarnar(-se) (carn), encartonar (cartó), encastellar(-se) (castell), encepar (cep), endentar(-se) (dent), enfangar(-se) (fang), enforçar (força), enforquillar (forquilla), enformar (forma), enfrontar(-se) (front), engarbullar (garbull), engrapar (grapa), enguerrrar (guerra), enjoiar(-se) (joia), enjudiciar (judici), enllaunar (llauna), enrastellar (rastell), enroscar (rosca), ensabar (saba), ensorrar (sorra), enterrar (terra), entonar (to), entubar (tub), envinagrar (vinagre), and enxarxar (xarxa).

➤ *Some intermediate Vs have related meanings to the prefixed Vs (A/B.1):*

Dent → dentar → endentar(-se)
 Grapa → grapar → engrapar
 Rastell → rastellar → enrastellar
 Sorra → sorrar → ensorrar

➤ *Some intermediate Vs have the same meanings as the prefixed Vs (A/B.2):*

Forma → formar(-se) → enformar
 Rosca → roscar → enroscar

**C) N as a result 'to make N' [+c] [-c-m] / Reflexive 'to become N' [-c-m]
Vs of creation**

Enarcar(-se) (arc), encadastrar (cadastre), encallir(-se) (call), encanallar-se (canalla), encarrellar (carrell), encartonar-se (cartó), encirar-se/enciriar-se (ciri), encistar(-se) (cist), encoixinar (coixí), enconcar(-se) (conca), encordonar (cordó), encrestar (cresta), endosserar (dosser), enfarcellar (farcell), enfardar (farda), enfardellar (fardell), enfardar (farda), enfeixar (feix), enfistular(-se) (fistula), enfolcar (folc), enforçar (forc), engallar-se/engallir-se (gall), engarbullar (garbull), engolfar-se (golf), engorgar-se (gorg), engracellular (graella), engruixar (gruix), engruixir(-se) (gruix), enjardinar (jardí), enllacar (llac), enllaçar (llaç), enquadernar (quadern), enraiar (rai), enrastellerar (rastellera), enrinxolar(-se) (rinxol), enrotllar (rotlle), enrullar(-se) (rull), enrunar(-se) (runa), ensenyorir(-se) (senyor), entoïar (toia), entollar(-se) (toll), entortellar (tortell), entorxar (torxa), entrunyellar (trunyella), envesprir (vespre), envetar (veta), envidreir-se (vidre), envidriar-se (vidre), enviduar (vidu), and enviudar (viuda).

- *Some intermediate Vs have related meanings to the prefixed Vs (C.1):*
 Feix → feixar → enfeixar
 Rotlle → rotllar → enrotllar(-se)
- *Some intermediate Vs have the same meanings as the prefixed Vs (C.2):*
 Arc → arcar-se → enarcar(-se)
 Llaç → llaçar → enllaçar
 Rínxol → rinxolar(-se) → enrinxolar(-se)
 Rull → rullar → enrullar(-se)

References

- Ackema, P. and Neeleman, A., 2004, *Beyond Morphology: Interface Conditions on Word Formation*. Oxford: Oxford U. P.
- Aronoff, M., 1976, *Word Formation in Generative Grammar*. Cambridge Mass.: MIT Press.
- Baker, M., 1988, *Incorporation: A Theory of Grammatical Function Change*. Chicago & London: University of Chicago Press.
- Booij, G., 1988, «The Relation between Inheritance and Argument Linking: Deverbal Nouns in Dutch». In Everaert, M. et al. (eds.), *Morphology and Modularity*. Dordrecht: Foris, pp. 57-73.
- Cabré, M. T., 1988, «La prefixació en català». In Staczeck, J. J. (ed.), *On Spanish, Portuguese and Catalan Linguistics*. Washington: Georgetown U. P., 47-63.
- 1994, *A l'entorn de la paraula*. València: Servei de Publicacions de la Universitat de València (2 vol.).
- and Rigau, G., 1986, *Lexicologia i Semàntica*. Barcelona: Enciclopèdia Catalana.
- Carrier-Duncan, J., 1985, «Linking of Thematic Roles in Derivational Word Formation», *LI* 16: 1-34.
- Chomsky, N., 1970, «Remarks on Nominalization.» In R. Jacobs and Rosenbaum, P. (eds.), *Readings in English Transformational Grammar*. Waltham (Mass.): Ginn and co. pp. 184-221.
- 1995, *The Minimalist Program*. Cambridge, Mass.: MIT Press.
- Clark, E. and Clark, H. 1979. «When Nouns Surface as Verbs», *Lg* 55: 767-811.
- COD: Pearsall, J. et al., 2001, *Concise Oxford Dictionary*. Oxford: Oxford U. P.
- DGLC: Fabra, P., 1932, *Diccionari general de la llengua catalana*. Barcelona: Edhasa. (ed. 1984).
- DLC: Institut d'Estudis Catalans, 1995, *Diccionari de la llengua catalana*. Barcelona: Enciclopèdia Catalana.
- Di Sciullo, A. M. and Williams, E., 1987, *On the Definition of Word*. Cambridge Mass.: MIT Press.
- 1997, «Prefixed-verbs and Adjunct Identification». In Di Sciullo, A. M. (ed.), *Projections and Interface Conditions*. New York / Oxford: Oxford U. P., 52-73.
- Fabb, N., 1984, *Syntactic Affixation*. Doctoral dissertation, MIT.
- 1988, «English suffixation is constrained only by selectional restrictions», *NLLT* 6: 527-539.
- GDLC: Giralt, J. et al, 1998, *Gran Diccionari de la llengua catalana*. Barcelona: Enciclopèdia Catalana.

- Gràcia, Ll., 1992, «-ble adjectives and middle constructions: a problem for inheritance», *Catalan Working Papers in Linguistics*: 163-182.
- , 1995, *Morfologia Lèxica: l'herència de l'estructura argumental*. València, Servei de Publicacions de la Universitat de València.
- et al., 2000, *Configuración morfológica y estructura argumental: léxico y diccionario*. Zaratuz (Guipúzcoa): Euskal Herriko Unibertsitateko Argitalpen Zerbitzua.
- Grimshaw, J., 1990, *Argument Structure*. Cambridge Mass.: MIT Press.
- Gruber, J., 1965, *Studies in Lexical Relations*. Doctoral Dissertation, MIT.
- Hale, K. and Keyser, S. J., 1993, «On Argument Structure and the Lexical Expression of Syntactic Relations». In Hale, K. and Keyser, S. J. (eds.). *A view from Building 20: Essays in Linguistics in Honor of Sylvian Bromberger*. Cambridge Mass.: MIT Press. pp. 53-109.
- , 1998, «The Basic Elements of Argument Structure». In Harley, H. (ed.). *MIT Working Papers in Linguistics 32, Papers from the Upenn/MIT Roundtable on Argument Structure and Aspect*. Cambridge, Mass.: MIT Press. pp. 73-118.
- , 2002, *Prolegomenon to a Theory of Argument Structure*. Cambridge Mass.: MIT Press.
- Halle, M., 1973, «Prolegomena to a Theory of Word Formation», *LI* 4: 3-16.
- Hoeksema, J., 1988, «Head-types in morpho-syntax». In Booij, G. and Marle, J. van (eds.), *Yearbook of Morphology*. Dordrecht: Foris. pp. 123-137.
- Hoekstra, T. and Putten, F. van der, 1988, «Inheritance Phenomena». In Everaert, M. et al. (eds.), *Morphology and Modularity*. Dordrecht: Foris. pp. 163-186.
- Jackendoff, R., 1975, «Morphological and Semantic Regularities in the Lexicon», *Lg* 51: 639-671.
- 1990, *Semantic Structures*. Cambridge, Mass.: MIT Press.
- 1997, *The Architecture of the Language Faculty*. Cambridge, Mass.: MIT Press.
- 2001, *Foundations of Language. Brain, Meaning, Grammar, and Evolution*. Oxford/New York: Oxford U. P.
- Kenstowicz, M., 1994, *Phonology in Generative Grammar*. Cambridge, Mass.: Blackwell Publishers.
- Kiparsky, P., 1997, «Remarks on Denominal Verbs». In Alsina, A. et al. (eds.), *Complex predicates*. Stanford, CA: CSLI Publications, pp. 473-499.
- Levin, B. and Rappaport, M., 1988, «Non-event -er Nominals: a Probe into Argument Structure», *Linguistics* 26: 1067-1083.
- & Rappaport Hovav, M., 1995, *Unaccusativity. At the Syntax-Lexical Semantics Interface*. Cambridge, Mass.: MIT Press.
- Lieber, R., 1989, «On Percolation». In Booij, G. and Marle, J. van (eds.), *Yearbook of Morphology*. Kluwer, 95-138.
- 2004, *Morphology and Lexical Semantics*. Cambridge/New York: Cambridge U. P.
- Marantz, A., 1997, «No Escape from Syntax: Don't Try Morphological Analysis in the Privacy of Your Own Lexicon», *UPenn Working Papers in Linguistics* 4: 201-225.
- 2001, *Words*. Unpublished ms. MIT.
- Marchand, H., 1969, *The Categories and Types of Present-Day English Word-Formation*. München: Georg Appl, Wending.
- Mascaró, J., 1986, *Morfologia catalana*. Barcelona: Enciclopèdia Catalana.
- Mateu, J., 2001a, «Preverbs in Complex Denominal Verbs: Lexical Adjuncts or Core Predicates?», *Catalan Working Papers in Linguistics* 9: 37-51.

- , 2001b, «Locative and Locatum Verbs Revisited. Evidence from Romance». In D'Hulst, Y. et al. (eds.), *Romance Languages and Linguistic Theory 1999*. Amsterdam: John Benjamins. pp. 223-244.
- , 2002, *Argument Structure: Relational Construal at the Syntax-Semantics Interface*. Doctoral Dissertation, Universitat Autònoma de Barcelona.
- , J., 2005, «Impossible Primitives». [To appear in Werning, M. et al. (eds.). *The Composition of Meaning and Content. Volume I: Foundational Issues*. Germany: Ontos Verlag] MWOD: <http://www.m-w.com/cgi-bin/dictionary>
- Neeleman, A. and Schipper, J., 1992, «Verbal prefixation in Dutch: thematic evidence for conversion». In Booij, G. and Marle, J. van (eds.), *Yearbook of Morphology*. Kluwer, 57-92.
- and Koot, H. van de, 2002, «Configurational Matrix», *LI* 33 (4): 529-574.
- Padrosa, S., 2005a, «Catalan Verbal Prefixation». In E. Daskalaki et al. (eds), *CamLing 2004: Proceedings of the University of Cambridge Second Postgraduate Conference in Language Research*. Cambridge: CILR. pp. 372-379.
- 2005b, *Argument Structure and Morphology: the Case of en- Prefixation in English and Catalan*. MA Dissertation, Universitat Autònoma de Barcelona.
- Pesetsky, D., 1995, *Zero syntax: Experiences and Cascades*. Cambridge, Mass.: MIT Press.
- Reinhart, T., 2000, «The Theta System: Syntactic Realization of Verbal Concepts». In OTS Working Papers in Linguistics. [To appear (extended) in the LI Monographs Series, MIT Press]
- 2001, «A Synopsis of the Theta system». Unpublished ms. Utrecht University. [An extended version can be found in: Reinhart, T., 2002, «The Theta System —an Overview», *Theoretical Linguistics* 28(3): 229-290.]
- and Siloni, T., 1999, «Against the Unaccusativity Analysis of Reflexives». [To appear in Alexiadou, A. (ed.) *Studies on Unaccusativity: the Syntax-Lexicon Interface*, CUP].
- Roca, I. and Johnson, W., 2000, *A Course in Phonology*. Oxford: Blackwell Publishers.
- Samek-Lodovici, V., 2003, «The internal structure of arguments and its role in complex predicate formation», *NLLT* 21: 835-881.
- Scalise, S., 1984, *Generative Morphology*. Dordrecht: Foris.
- , 1988a, «The notion of 'head' in morphology». In Booij, G. and Marle, J. van (eds.), *Yearbook of Morphology* 1. Dordrecht: Foris. pp. 229-245.
- , 1988b, «Inflection and derivation», *Linguistics* 26: 561-581.
- Selkirk, E., 1982, *The Syntax of Words*. Cambridge: MIT Press.
- Siegel, D., 1979, *Topics in English Morphology*. New York/London: Garland. CCED: Sinclair, J. et al., 1995, *Collins Cobuild English Dictionary*. Great Britain: HarperCollins Publishers.
- Sperber, D. and Wilson, D., 1986/1995, *Relevance: Communication and Cognition*. Oxford: Blackwell.
- Sportiche, D., 1998, *Partitions and Atoms of Clause structure: Subjects, agreement, case and clitics*. London: Routledge.
- Stiebels, B., 1998, «Complex Denominal Verbs in German and the Morphology-Semantics Interface». In Booij, G. and J. van Marle (eds.), *Yearbook of Morphology 1997*. Kluwer. pp. 265-302.
- Varela, S. (ed.), 1993, *La formación de palabras*. Madrid: Taurus Universidad.
- and Haouet, L., 2001, «For a Morphological Analysis in the Privacy of the Lexicon: Prefixed Verbs», *Cuadernos de Lingüística del I.U. Ortega y Gasset* 8: 53-69.

- Williams, E., 1981a, «On the Notions ‘Lexically Related’ and ‘Head of a Word’», *LI* 12: 245-274.
- 1981b, «Argument Structure and Morphology», *The Linguistic Review* 1: 81-114.
- 2004, «Dumping Lexicalism». Unpublished ms. Princeton University.
- Wilson, D., 1994, «Relevance and Understanding». In Brown, G. et al. (eds.). *Language and Understanding*. Oxford: Oxford U. P., 1-30.
- and Sperber, D., 2004, «Relevance Theory». In Ward, G. et al. (eds.). *Handbook of Pragmatics*. Oxford: Blackwell, 607-632.
- Zwanenburg, W., 1988, «Morphological Structure and Level Ordering». In Everaert, M. et al. (eds.), *Morphology and Modularity*. Dordrecht: Foris, 395-410.

A CROSSLINGUISTIC PERSPECTIVE ON N-WORDS

Doris Penka

University of Tübingen

Abstract

The semantic status of so-called n-words in Negative Concord languages has been under considerable debate. This paper takes a new perspective on this problem by bringing Negative Concord together with two different phenomena that n-words give rise to in non-Negative Concord languages, namely scope splitting in German and distributional restrictions in the Scandinavian languages. I argue that all this taken together reveals the common nature of n-words across languages. These phenomena suggest that n-words should not be analysed as negative quantifiers. Rather, n-words are morpho-syntactic markers of sentential negation. I present a cross-linguistic analysis of n-words and show how the three phenomena discussed follow from it. This analysis is based on the assumption that n-words are semantically non-negative and must be licensed by a (possibly abstract) negation. It is proposed that n-words cross-linguistically are of essentially the same nature and that differences between languages regarding their behaviour are due to parametric variation.

1. Introduction

This paper is concerned with the syntax and semantics of words that (in a pre-theoretical sense) have both a negative and some other meaning component, usually indefinite. Since Laka (1990) these words are called n-words. The use I make of the term ‘n-word’ in this paper might be more liberal than is customary. First, I will take n-words to comprise not only negative forms of indefinites (or ‘negative quantifiers’), but also certain other items, such as the conjunction *ni... ni* ‘neither... nor’ in Spanish as argued for by Herburger (2001). Second, while the term ‘n-word’ was introduced as a theory-neutral name for these words in Negative Concord languages, in which their nature is notoriously unclear (see section 2), I will not restrict it to Negative Concord languages, but their pendants in non-Negative Concord languages (also called Double Negation languages) will also be called n-words. The reason for this is that the difference between n-words in Negative Concord languages and non-Negative Concord languages is much smaller than generally assumed, as will be shown in this paper. A sample of the n-word inventory of several languages is given in Table 1.

Table 1
n-word inventory of some languages

	English	German	Italian	Spanish	Polish
person	<i>nobody</i>	<i>niemand</i>	<i>nessuno</i>	<i>nadie</i>	<i>nikt</i>
thing	<i>nothing</i>	<i>nichts</i>	<i>niente</i>	<i>nada</i>	<i>nic</i>
determiner	<i>no</i>	<i>kein</i>	<i>nessuno</i>	<i>ningún</i>	<i>zaden</i>
time	<i>never</i>	<i>nie (mals)</i>	<i>mai</i>	<i>nunca</i>	<i>nigdy</i>
place	<i>nowhere</i>	<i>nirgendwo</i>	—	—	<i>nigdzie</i>
conjunction	<i>neither... nor</i>	<i>entweder... oder</i>	<i>né... né</i>	<i>ni... ni</i>	<i>ani... ani</i>

A characteristic of n-words is that they can be used as negative fragmentary answers:

- (1) a. Who came to the party? - Nobody.
b. Who came to the party? - *Anybody.

Since this contrasts with Negative Polarity Items (NPIs), as shown in (1b), the ability to constitute negative fragmentary answers makes a useful criterion to distinguish n-words from NPIs (see Giannakidou 2002). In the literature, n-words have sometimes been subsumed under NPIs (e.g. in Laka 1990; Giannakidou 1997) and this has led to confusion. It is important to keep n-words and NPIs apart, since as will be argued, they are subject to different licensing conditions.

The standard analysis of (indefinite) n-words is as nominal or adverbial negative quantifiers, i.e. their lexical entry expresses negated existential quantification, as exemplified for *nobody* in (2):

- (2) $[[\textit{nobody}]] = \lambda P. \neg \exists x [\textit{person}(x) \ \& \ P(x)]$

However, there are reasons to doubt that the negative quantifier analysis constitutes the whole story. In this paper, I discuss three phenomena that n-words give rise to in different languages. All of them are problematic for the negative quantifier approach. While they have so far been discussed independently of each other, the aim of this paper is to bring them together and thus derive conclusions on the common nature of n-words across languages.

2. Negative Concord

2.1. Data

The first phenomenon arising in connection with n-words has been extensively discussed in the literature and is known as Negative Concord (NC) (see Laka 1990, Zanuttini 1991, Haegeman 1995, Zeijlstra 2004, among many others). In languages that exhibit NC, multiple negative expressions yield an interpretation with only one negation as shown by the following examples.¹

¹ \$ is used to indicate that the sentence does not have the reading paraphrased.

- (3) Gianni non ha visto nessuno. Italian
 Gianni neg has seen n-person
 ‘Gianni hasn’t seen anybody.’
 \$‘Gianni hasn’t seen nobody.’ = ‘Gianni has seen somebody.’
- (4) Nikt nie przeczytał tego artykułu. Polish
 n-person neg read-3SG.PAST this-GEN article-GEN
 ‘Nobody has read this paper.’
 \$‘Nobody has not read this paper.’ = ‘Everybody has read this paper.’

It is useful to distinguish between strict and non-strict NC-languages (see Gianakidou 2002). In strict NC languages, an n-word is obligatorily accompanied by the sentential negative marker, independently of the position of the n-word. The Slavic languages are strict NC languages, as can be seen for Polish in the following example (from Błaszczak 2001: (217)).²

- (5) a. Żadne dziecko *(nie) wyjechała na wakacje. Polish
 no child neg go-3SG.PAST on holiday
 ‘No child went on holiday.’
 b. *(Nie) wyjechało żadne dziecko na wakacje.
 neg go-3SG.PAST no child on holiday
 ‘No child went on holiday.’

On the other hand, Romance languages are non-strict NC languages, since only postverbal n-words require the presence of the negative marker (6b). A preverbal n-word plus a negative marker is ungrammatical, or at best yields a reading with double negation (6a).³

- (6) a. Nadie (*no) vino. Spanish
 n-person neg came
 ‘Nobody came.’
 b. *(No) vino nadie.
 neg came n-person
 ‘Nobody came.’

2.2. Approaches to n-words in NC languages

Due to the confusing behaviour of n-words in NC languages—in some cases such as (6a) they seem to contribute a negation to the semantics, in others such as (6b) they apparently do not—there is no consensus on their semantic status. I cannot possibly do justice to the considerable literature on NC in this paper and will only give a brief overview on the main positions held.

One line of research considers n-words to be negative quantifiers (a. o. Zanuttini 1991, Haegeman 1995, de Swart and Sag 2002). In these accounts, the behaviour of preverbal n-words in non-strict NC languages follows immediately, but additional

² *(x) is used to signify that the sentence is judged grammatical with x and ungrammatical without. (*x) on the other hand means that the sentence is judged grammatical without x and ungrammatical (under the reading paraphrased) with x.

³ There are, however, some exceptions to this claim: in some (varieties of) languages in the Romance family, e.g. Catalan, preverbal n-words can optionally be accompanied by a negative marker.

assumptions are needed to account for postverbal n-words. In order to explain that n-words can lose their negative force, a mechanism called polyadic quantification is used, which absorbs the negative component of an n-word if it is in a certain configuration with another negative element.

Another set of accounts takes the fact that n-words in strict NC languages and postverbal n-words in non-strict NC languages do not seem to have negative force to reveal the nature of n-words. Accordingly, they assume that n-words are semantically non-negative and must be licensed by a negation. These accounts differ in how this is exactly spelled out. For Laka (1990) n-words in Spanish are NPIs, while Ladusaw (1992), whose analysis will be discussed in more detail in section 4.5, proposes that n-words are indefinites that must be bound by a negation operator. Common to both of them is that in order to account for preverbal n-words in non-strict NC languages they assume that the negation operator can be abstract.

Finally, Herburger (2001) takes the two-sided behaviour of n-words in non-strict NC languages at face value and argues that they are lexically ambiguous between negative quantifiers and NPIs.

The approach taken in this paper is that NC should be seen in the light of other phenomena n-words exhibit in non-NC languages. Rather than assuming that n-words in NC languages are special, the ability to participate in NC should follow from their common cross-linguistic nature. As we will see, NC is only one of the reasons to believe that the negative quantifier analysis is not a good candidate.

3. Scope Splitting

Although n-words in non-NC languages at first glance clearly seem to be negative quantifiers, we find a related problem in them. In this case the problem is not that the negative quantifier analysis results in too many negations, but rather that the negation is in the wrong position.

3.1. German data

In German, in certain environments n-words can split their scope in the sense that an operator takes scope in between the negation and the indefinite meaning component (see Bech 1955/57, Jacobs 1980). Consider the following example:

- (7) Du musst keine Krawatte anziehen.
 you must n-Det tie wear
- | | |
|---|--------------|
| a. 'It is required that you don't wear a tie.' | must > ¬ > ∃ |
| b. 'There is no tie that you are required to wear.' | ∃ > ¬ > must |
| c. 'It is not required that you wear a tie.' | ¬ > must > ∃ |

Under the assumption that n-words in German are negative quantifiers only the readings (7a,b) are derived. (7a), in which the negative quantifier is interpreted with surface scope, is hard to get and only available with help from the context, because it runs against the strong tendency of modals in German to be in the scope of negation

rather than vice versa. The only way for the modal to get in the scope of negation is LF-movement of the negative quantifier across it. This results in the reading paraphrased as (7b). But this reading is quite weak since it says that there is no specific tie you are required to wear. This does not exclude that you might have to wear some tie or other because the occasion requires it. However, the sentence (7) is usually understood to convey that it is fine if you do not wear a tie. So in the salient reading, paraphrased in (7c), the negation has wide scope over the modal whereas the indefinite is interpreted within the scope of the modal (de dicto reading of the indefinite). Thus for the interpretation the n-word is split into a negative and an indefinite part, so that the modal can take scope in between the two. Under the negative quantifier analysis this is not possible, since the negation and the indefinite are part of the meaning of the lexical unit *kein*.

Let me convince you that the split reading is real and cannot be reduced to one of the two readings derived by the negative quantifier analysis. In sentences with expletive *es* ('there') an indefinite subject of a modal verb only has the narrow scope reading:

- (8) Es muss ein Arzt anwesend sein.
 there must a physician present be
 a. 'It is required that there be a physician present.' must > \exists
 b. \$ 'There is a physician who is required to be present.' \exists > must

This also holds if the subject consists of an n-word. But although such a sentence does not have a reading in which the negative quantifier takes wide scope, its salient reading is nevertheless (9c), in which the negation outscopes the modal.

- (9) Es muss kein Arzt anwesend sein.
 there must n-Det physician present be
 a. 'It is required that there be no physician present.' must > \neg > \exists
 b. \$ 'There is no physician who is required to be present.' \neg > \exists > must
 c. 'It is not required that there be a physician present.' \neg > must > \exists

It is also possible to construct examples for which the split reading is the only possible one. These involve the modal verb *brauchen* ('need'), which is an NPI and must therefore be interpreted in the scope of a negative item, thus excluding the narrow scope reading of the negative quantifier. But simultaneously, because of expletive *es* ('there'), *kein Arzt* ('no doctor') is required to have narrow scope with respect to the modal:

- (10) Es braucht kein Arzt anwesend sein.
 there need n-Det doctor present be
 a. \$ 'It is required that there be no physician present.' must > \neg > \exists
 b. \$ 'There is no physician who is required to be present.' \neg > \exists > must
 c. 'It is not required that there be a physician present.' \neg > must > \exists

Besides the context of modal verbs, n-words give also rise to a reading with split scope when they are the object of transitive intensional verbs, such as *suchen* ('seek').

- (11) Peter sucht kein Einhorn.
 Peter seeks n-Det unicorn
 a. \$ 'Peter is trying not to find a unicorn.' seek > ¬ > ∃
 b. 'There is no unicorn that Peter is trying to find.' ¬ > ∃ > seek
 c. 'Peter is not trying to find a unicorn.' ¬ > seek > ∃

For n-words as objects of transitive intensional verbs the narrow scope reading of a negative quantifier (11a) is not available. The wide scope reading (11b), again, is rather weak since it is already true if unicorns do not exist in the evaluation worlds, independently of Peter's activities.

Not only verbs can take scope in between the negation and the indefinite part of n-words, but also nominal quantifiers. But while scope splitting with respect to verbs expressing intensional operators is generally possible, this is restricted to sentences bearing topic-focus-accent. Under this rise-fall-contour a universal DP in topic position has scope in between the negation and the indefinite contributed by an n-word in the Mittelfeld. This time the split reading is the only available one.

- (12) JEDER/ Student hat KEIN\ Auto.
 every student has n-Det car
 'It is not true that every student has a car.' ¬ > ∀ > ∃

3.2. Previous accounts of scope splitting

The data presented in the last subsection pose a problem for the assumption that n-words in German are negative quantifiers. There are, however, analyses that derive the split reading of n-words while retaining this assumption. In these accounts special mechanisms are proposed to handle scope splitting. Geurts (1996) assumes that the split reading is due to quantification over abstract individuals, while de Swart (2000) employs quantification over higher types. But both of these analyses face a serious problem: The mechanisms proposed apply unrestrictedly and thus overgenerate. For instance, if the split reading of (13) is assumed to be due to a special interpretation of *kein Buch* 'no book' nothing prevents this interpretation from applying to it in (14) as well, thus deriving a split reading that is not available.

- (13) /ALLEN Studenten habe ich KEIN\ Buch empfohlen.
 all students. DAT have I no book.ACC recommended
 'It isn't true that for every student there is a book such that I recommended it to him.' ¬ > ∀ > ∃
- (14) Ich habe kein Buch allen Studenten empfohlen.
 I have no book. ACC all students. DAT recommended
 'There is no book that I recommended to every student.' ¬ > ∃ > ∀
 \$ 'It isn't true that for every student there is a book such that I recommended it to him.' ¬ > ∀ > ∃

The mechanisms derive split readings for n-words in contexts where they do not have such readings. It remains unaccounted for that scope splitting is restricted to particular environments such as topic-focus-accent.

4. Analysis

4.1. Conclusion from the data

We have seen two phenomena that arise in connection with n-words in different languages. So what lesson can we draw from them regarding the nature of n-words? NC shows that n-words are not always semantically negative. The data with split scope of n-words in German demonstrate that the negative meaning component of n-words can take scope independently of the indefinite meaning component. Both of these properties are unexpected under the assumption that n-words are negative quantifiers and cannot easily be handled by such an analysis. Therefore, I conclude that n-words should not be analyzed as negative quantifiers. Rather, I propose that the discussed properties are part of the true nature of n-words. In the remainder of this section, an analysis that implements this idea is presented in detail.

4.2. N-words: semantically non-negative elements licensed by negation

The main conclusion that can be drawn from the data discussed in the previous sections is that n-words themselves are not bearers of semantic negation. Rather, they are semantically non-negative, which means that the meaning of an n-word is the same as for its positive pendant.

$$(15) \text{[[nobody]]} = \text{[[somebody]]} = \lambda P. \exists x [\text{person}(x) \ \& \ P(x)]$$

From this semantics the phenomenon of NC follows immediately: n-words in NC constellations do not contribute a negation to the meaning of the sentence, simply because their semantics does not contain a negation.


Another ingredient of the analysis is needed to explain that n-words only occur in negative sentences. This is achieved by the requirement that n-words must be licensed by negation. One way to spell out this licensing requirement would be to assume that n-words are NPIs (taken by Laka 1990, Giannakidou 1997). But this is problematic since n-words and NPIs are not licensed in exactly the same contexts: n-words cannot (without contributing negative force) occur in all contexts in which NPIs are allowed.⁴ And what is worse, NPIs are not acceptable in all contexts in which n-words are licensed, for example in negative fragmentary answers, as has been demonstrated in (1). Furthermore, a line of thinking about NPIs has been established (Kadmon and Landman 1993, Krifka 1995, Lahiri 1998) that derives the need for a negative context from the fact that the use of an NPI makes a statement stronger. Since no such strengthening is associated with n-words, their licensing requirements must have a different source.

I follow Zeijlstra (2004), who argues that the licensing of n-words in NC languages is a form of syntactic agreement. N-words carry an uninterpretable feature [uNEG] that must be checked against an interpretable feature [iNEG] carried by a negative operator. For instance, in the Italian example (16) the n-word *nessuno* has the feature [uNEG], which must be licensed. As the sentential negation marker *non* ('not') is semantically negative it has the feature [iNEG], which checks the [uNEG]-feature on *nessuno* (cf.

⁴ Giannakidou (1997) accounts for this fact by proposing that n-words have stronger licensing requirements: while NPIs are licensed in non-veridical contexts, n-words require anti-veridical contexts.

17). On the other hand, if the negative marker is not present, as in (18), there is no semantically negative element carrying [iNEG] and thus the [uNEG]-feature on *nessuno* is not licensed.

(16) Gianni non telefona a nessuno. Italian
 Gianni neg call to n-person
 ‘Gianni doesn’t call anybody.’

(17) Gianni non_[iNEG] telefona a nessuno_[uNEG]


(18) *Gianni telefona a nessuno_[uNEG]


To explain the fact that more than one n-word can be licensed by the same negation, as in (19), Zeijlstra (2004) proposes that n-word licensing is subject to Multiple Agree (Haraiwa 2001), i.e. several [uNEG]-features can be checked by one and the same [iNEG]-feature, as shown in (20).


(19) Maria non ha detto niente a nessuno. Italian
 Maria neg has said n-thing to n-person
 ‘Maria hasn’t said anything to anybody.’

(20) Maria non_[iNEG] ha detto niente_[uNEG] a nessuno_[uNEG]


4.3. Abstract negation

But what about preverbal n-words in non-strict NC languages, which do not co-occur with the negative marker (recall (6a))? Ladusaw (1992) proposes that in these cases the sentential negation is realized abstractly. In his terminology, n-words are self-licensing in the sense that an n-word can license itself by introducing an abstract negative operator.⁵ But sentential negation may only be abstract if its presence is marked by an element preceding the verb. Assuming that a feature [iNEG] on a semantic negation can only check a [uNEG]-feature carried by an n-word if the negation c-commands the n-word (cf. Zeijlstra, 2004), preverbal n-words must be c-commanded by an abstract negation. This means that the underlying structure of (21) is (22), where NEG is an element that is semantically interpreted as sentential negation and not realized phonologically.

(21) Nessuno telefona a Gianni. Italian
 n-person call to Gianni
 ‘Nobody calls Gianni.’


(22) [NEG_[iNEG] [nessuno_[uNEG] telefona a Gianni]]


⁵ This ability for self-licensing can also be made responsible for the fact that n-words on their own can be used as negative fragmentary answers (see (1)).

This also explains why in non-strict NC languages preverbal n-words co-occurring with a negative marker yield a reading with double negation (if they receive prominent stress; otherwise such sentences are judged as ungrammatical). Since preverbal n-words are licensed by a c-commanding abstract negation, in this case the negative marker is the second semantically negative element in the structure of the sentence.

In contrast to non-strict N-languages like Italian, n-words in strict NC languages like Polish are also accompanied by a negative marker when they are preverbal (see (5a)). So why is it that in these languages this constellation does not result in a double-negation reading (or ungrammaticality)? Zeijlstra (2004) argues that in strict NC languages the negative marker on the verb itself is not semantically negative and carries a feature [uNEG]. Thus the semantic negation is always abstract in strict NC languages.

(23) Nikt nie przeczytał tego artykuł u. Polish
 n-person neg read-3SG.PAST this-GEN article-GEN
 'Nobody has read this paper.'

(24) NEG_[iNEG] nikt_[uNEG] nie_[uNEG] przeczytał tego artykuł u


So the difference between strict and non-strict NC languages is reduced to the status of the negative marker in a language: in strict NC languages it is semantically negative, whereas in non-strict NC languages it is not.

One may find it odd that something as important to the meaning of a sentence as negation can be realized abstractly. But then, why not? As long as there is always clear indication of the presence of negation, it does not really matter whether the negative operator itself is expressed overtly or covertly. Due to their licensing conditions, n-words are automatically connected to a negation and so there is no need for the negation itself to be present overtly. There are thus two strategies in natural language to express sentential negation: the first is using a negative marker corresponding to semantic negation, the second is using n-words that mark the presence of a possibly abstract negation.

4.4. N-words in non-NC languages

So far the theory of Zeijlstra (2004) on NC. But whereas he assumes a dichotomy between n-words in NC languages, which are semantically non-negative and subject to syntactic licensing conditions, and n-words in non-NC languages, which he considers as negative quantifiers, I argue that all n-words are essentially the same in nature.

An analysis according to which n-words are semantically non-negative and must be licensed by a possibly abstract negation explains straightforwardly the phenomenon of scope splitting that n-words in German show as discussed in section 3 (such an analysis was proposed in Penka and von Stechow 2001). As the negation and the indefinite do not form a semantic unit, it follows immediately that some other operator can take scope in between the two. For non-NC languages in

most cases it does not make a difference whether an n-word is analyzed as negative quantifier or an indefinite plus a sentential negation. But in cases where some other semantic operator takes scope in between the negation and the indefinite, the difference becomes crucial.

Consider again the example (7), which is repeated as (25) below, this time as embedded clause to abstract away from V2 movement. Recall that the salient reading is the split reading as paraphrased in (25). Responsible for this reading is the fact that the abstract negation licensing the n-word can be in a position high enough to also c-command the modal verb, as illustrated in the structure (26a). From the surface structure (26a) the LF (26b), which expresses the intended truth conditions, is immediately derived (by reconstruction of the subject to a position within the embedded vP).

- (25) ... dass du keine Krawatte anziehen musst German
 ... that you n-Det tie wear must
 ‘... that it is not required that you wear a tie’

- (26) a. ... dass du NEG [[keine Krawatte anziehen] musst]
 b. LF: NEG [[du keine Krawatte anziehen] musst]

Note that there is no need to move the negation to the position from which it takes scope, since it is already there in the surface structure.⁶ That the LF (26b) corresponds to the salient reading is due to the fact that modal verbs in German show a strong tendency to be in the scope of negation rather than vice versa (see de Haan 1997), and this does not depend on whether the negation is overt or abstract. But if the context requires it, NEG —just as the negation marker *nicht* (‘not’)— can also be in the scope of the modal, i.e. adjoined to the embedded vP, yielding a reading in which the modal outscopes both the negation and the indefinite (‘It is required that you don’t wear a tie.’):

- (27) a. .. dass du [NEG [keine Krawatte anziehen] musst]
 b. LF: [NEG [du keine Krawatte anziehen] musst]

To obtain the wide scope reading in which both the negation and the indefinite have wide scope with respect to the modal (‘There is no tie that you are required to wear.’), I assume that QR can also target vP. Thus the LF expressing this reading is derived from the surface structure in (27a) by adjoining the quantifier *keine Krawatte* (‘no tie’) to the embedded vP in the scope of NEG:

- (28) LF: [NEG keine Krawatte λ_1 [du 1 anziehen] musst]

The fact that an n-word must be interpreted in the scope of its licensing negation can be reduced to a general constraint on LF-movement, according to which a negation operator constitutes a barrier for upward movement (see Beck 1996 for German).

The assumption that n-words in NC and non-NC languages have the same nature immediately raises the question how the different behaviour n-words shown in the two types of languages is accounted for. I propose that the difference is due to parametric

⁶ Assuming LF-movement of the negation would be undesirable for two reasons. First, it would be hard to motivate, since adverbs always seem to have surface scope. Second, the movement of a propositional operator like negation does not have a semantic effect at all, unless such movement would be stipulated to not leave a semantically interpreted trace, which would result in a rather strange kind of movement.

variation of Multiple Agree with respect to [NEG]-features. NC languages have Multiple Agree and thus several [uNEG]-features can be checked by one [iNEG]-feature (see (20)). In contrast, non-NC languages do not have Multiple Agree and accordingly the ratio of semantic negations to n-words is 1:1. This means that in non-NC languages each n-word is licensed by its own c-commanding abstract negation:⁷

- (29) a. ... dass niemand kein Auto hat. German
 ... that n-person n-Det car has
 ‘... that nobody has no car’ = ‘... that everybody has a car’
 \$ ‘... that nobody has a car’
- (30) ... dass NEG_[iNEG] [niemand_[uNEG] NEG_[iNEG] [kein_[iNEG] Auto hat]]



Furthermore, in German the negative operator licensing n-words must be abstract and cannot be realized overtly. Otherwise we would expect the sentential negative marker, which is assumed to be semantically negative and hence to have the feature [iNEG], to license n-words, parallel to Italian *non* (‘not’). That this is not the case is evident from the fact that the following sentence only has a reading with double negation, if it is acceptable at all:

- (31) ?... Peter nicht kein Auto hat. German
 ... Peter neg n-Det car has
 ‘... that Peter does not have no car’ = ‘... that Peter has a car’
 \$ ‘... that Peter does not have a car’

It is clear that n-words in non-NC languages can only be licensed by an abstract negation, since otherwise they would show a form of negative concord holding between the negative marker and an n-word (provided that the negative marker in non-NC languages is semantically negative). But the licensing conditions in these languages are even stricter. While n-words in NC languages are licensed in the entire domain c-commanded by a negative operator, n-words in non-NC languages must be immediately surface-adjacent to NEG, i.e. no phonologically realized element may intervene between an n-word and its licensing negation.⁸ This explains why scope splitting in German is restricted to certain environments, which were discussed in section 3. Recall that scope splitting is generally possible with respect to verbal operators such as modal and object intensional verbs. Given that the basic word order in German corresponds to SOV, this is expected because an abstract negation licenses an n-word in the leftmost position within the VP under surface adjacency, even

⁷ This implies that there is no fixed position (NegP) for the negation operator. For German this is in line with Jacobs (1982), who argues that negation is a sentential adverb that can be adjoined to any node at least as high as VP.

⁸ This formulation of the requirement is actually too strict, since there is one type of elements that may intervene between an n-word and its licensing NEG, namely prepositions. N-words may be embedded in PPs, even under a split reading:

- (i) Peter sucht nach keinem Einhorn. German
 Peter seeks after n-Det unicorn
 ‘Peter is trying not to find a unicorn.’

if NEG is in a structurally higher position also c-commanding the verb. So whenever the surface structure corresponds to (32) a reading is available in which the verb takes scope in between negation and the n-word.

- (32) [NEG [_{VP} [n-word...] V]]

While the fact that scope splitting with respect to intensional verbs is always possible is put down to the availability of a surface structure in which the corresponding scope relations hold, the case of scope splitting with respect to nominal operators is different. DPs can only take scope in between an n-word and its licensing NEG if they get into their scope position during the derivation of LF, since at the surface the two have to be adjacent. This explains why split readings with respect to universal quantifiers are restricted to the context of topic-focus-accent. According to Büring (1997), the only available reading for sentences bearing this intonation pattern is one in which the topicalised constituent is reconstructed to its base position. Given this, we can now explain how the split reading for a sentence such as (33) comes about. In the surface structure (34a) the object *kein Auto* ('no car') is immediately adjacent to NEG dominating the vP, because the subject has moved to the topic position. But due to the topic-focus-accent, the only LF expressing an available reading is the one in which the subject is reconstructed to its base position within vP (34b). Hence at LF, the universal subject intervenes between NEG and the n-word, yielding the split reading.

- (33) JEDER/Student hat KEIN\ Auto.
 every student has n-Det car
 'It is not true that every student has a car.'

- (34) a. [_{CP} [_{DP} jeder Student]_i hat_j [NEG [_{vP} t_i kein Auto t_j]]]
 b. LF: [NEG [_{vP} jeder Student kein Auto hat]]

Usually in German the scope relations at LF correspond to the order of elements at the surface. So it is only if something, such as the meaning of topic and focus, forces the scope relations at LF to be different, that a nominal quantifier can take scope in between the negation and the n-word. Thus the analysis of n-words presented here can not only straightforwardly derive split readings, but also provides an explanation for why they are restricted to certain environments. The problem of overgeneration that alternative accounts face does not arise in the first place.

4.5. Comparison to Ladusaw (1992)

Many of the ideas which the analysis presented here is based on are already present in Ladusaw (1992, 1994, 1995). But Ladusaw's proposal is programmatic in nature and is therefore hard to interpret. In this section, I want to argue against one way in which his proposal can be interpreted. The central idea is stated in the following quote from Ladusaw (1992: 254):

- (35) «Assuming that all the negative argument expressions are univocally indefinites which are strong NPIs, i.e. must be roofed in lf by a negative operator, we have an account of the pattern of negative concord.»

“Negative argument expressions” refers to the expressions that are called n-words here and the roof of an indefinite is defined as “the operator that triggers the anchoring or binding of an indefinite” (Ladusaw, 1992: 245).

The analysis sketched in the quote above makes crucial use of the fact that n-words are the negative forms of indefinites, and relates them to the semantics of indefinites. According to Heim (1982) indefinites are not existential quantifiers, but rather open propositions consisting of a free variable and a predicate over this variable, e.g. *a boy* is translated as ‘ x is a boy’. The free variables introduced by indefinite expressions are bound by a semantic operator. Such an operator can either be a (nominal, adverbial or verbal) quantifier in the sentence, a covert modal operator associated with conditionals or an existential closure operator. The licensing relation between n-words and a negation can now be regarded as binding of an indefinite variable by a negation operator, i.e. the free variables introduced by n-words must be bound by a negation.

But such an analysis of n-words becomes problematic when their ability for scope splitting is taken into account. This also holds for NC languages, in which scope splitting is transparent in the sense that in constructions with the split reading the negative marker precedes the verb. Thus the order of the elements at the surface overtly reflects the scope relations at LF:

- (36) Ty ne dolzhen mne darit nikakich podarkov. Russian
 you neg must me-DAT give n-Det.GEN.PL present-GEN.PL
 ‘It is not necessary that you give me presents.’

Under the assumption that the variables introduced by n-words must be bound by the negation operator, the following semantic representation for the sentence (36) is derived, where $\text{Acc}(w, w')$ means that a possible world w' is accessible from the evaluation world w under a certain (deontic, circumstantial etc.) interpretation of the modal:⁹

- (37) $\neg \exists x [\forall w': \text{Acc}(w, w') \rightarrow x \text{ are presents in } w' \ \& \ \text{you give me } x \text{ in } w']$

But (37) expresses exceedingly weak truth conditions. It is true whenever there is no group of things in the real world for which it follows from the modal background that these things are presents. This does not correspond to a natural reading of (36). The problem with (37) is that the operator binding the indefinite variable has wide scope over the modal while the restrictor has narrow scope. The representation expressing the split reading correctly is (38), where the variable introduced by the n-word is existentially bound within the scope of the modal while the negation has wide scope.

- (38) $\neg \forall w': \text{Acc}(w, w') \rightarrow \exists x [x \text{ are presents in } w' \ \& \ \text{you give me } x \text{ in } w']$

These considerations show that a semantic licensing condition for n-words according to which the indefinite variables introduced by n-words must be bound by a negation cannot be correct.

⁹ I assume that negation triggers existential closure in its scope and that n-words are licensed more precisely if they are bound by an existential closure operator triggered by negation (see the above definition of ‘roof’).

Note that the analysis I presented in the last section does not make any commitment regarding the quantificational status of n-words: they can be translated as Heimian indefinites as well as existential quantifiers.¹⁰ This constitutes another advantage over the negative quantifier approach, since n-words can also occur in contexts for which it has been argued that indefinites should be interpreted as properties rather than quantifiers, such as existential constructions (see McNally 1998):

- (39) Es gibt hier keine Gespenster. German
 there are here n-Det.PL ghost-PL
 'Ghosts do not exist here.'

According to the analysis presented here, the negation associated with the n-word in (39) refers to the verb. Since the semantics of *keine Gespenster* ('no ghosts') is the same as for the corresponding positive indefinite, it can express a property, which constitutes the argument of the existential verb.

5. Distributional restrictions in Scandinavian

The assumption that n-words cross-linguistically are licensed by negation and that this licensing is of syntactic nature is confirmed by a third phenomenon exhibited by n-words. In the Scandinavian languages, n-words are restricted in their syntactic distribution (see Christensen 1986, Kayne 1998, Svenonius 2002). An n-word cannot occur in object position if the clause is embedded or if the verb form is composed of a participle, as the following Norwegian examples (from Christensen 1986) illustrate:

- (40) Jon leser ingen romaner. (41) *Jon har lest ingen romaner.
 Jon reads n-Det novels Jon has read n-det novels
 'Jon doesn't read (any) novels.' 'Jon hasn't read (any) novels.'
- (42) *Dette er en student som leser ingen romaner.
 this is a student who reads n-Det novels
 'This is a student who doesn't read (any) novels.'

Norwegian

The generalisation underlying this pattern of restricted distribution is that n-words in Norwegian are only grammatical if they are adjacent to the canonical position of the negative marker *ikke* ('no'). In cases in which an n-word is ungrammatical another element intervenes between it and the position of *ikke*, as can be seen in the grammatical pendants of these sentences, in which the negative marker plus an indefinite is used:

- (43) Jon leser ikke noen romaner. (44) Jon har ikke lest noen romaner.
 Jon reads neg some novels Jon has neg read some novels
 'Jon doesn't read (any) novels.' 'Jon hasn't read (any) novels.'

¹⁰ The lexical entry in (15) should be understood as simplification to abstract away from the problems of the semantics of indefinites.

- (45) Dette er en student som ikke leser noen romaner.
 this is a student who neg reads some novels
 'This is a student who doesn't read (any) novels.'

In the embedded clause in (42), the finite verb intervenes between the position the negative marker would occupy and the n-word. As in main clauses the finite verb is subject to V2 movement, it gets out of the way and does not intervene anymore in the licensing of the n-word (cf. 40). But if part of the verb, e.g. a participle, stays behind as in (41), there is still material intervening and an n-word is not licensed.

These restrictions on the distribution of n-words in Scandinavian are actually predicted by the analysis of n-words presented in the last section. As the Scandinavian languages do not exhibit NC, the licensing conditions for n-words are the same as in German, i.e. n-words must be surface adjacent to an abstract negation. But in contrast to German, which is SOV, the basic word order in these languages is SVO. So in basic word order, the verb intervenes in the licensing of n-words. An n-word in object position is not adjacent to NEG, which must c-command the verb to express sentential negation, and thus yields ungrammaticality. This is illustrated in (46) for the structure underlying (42):

- (46) *... som NEG [leser ingen romaner]

But if the verb moves out in main clauses as in (40), adjacency holds and an n-word is properly licensed:

- (47) [_{CP} Jon leser_i [NEG [*t_i* ingen romaner]]]

The syntactic restrictions n-words are subject to in the Scandinavian languages thus follow immediately from the licensing conditions that were put forward for German.

6. Conclusions

In this paper, I have discussed three phenomena n-words give rise to in different languages and used them to derive conclusions on the cross-linguistic nature of n-words. The fact that n-words show NC indicates that they are semantically non-negative. Rather, they are licensed by sentential negation. That n-words refer to sentential negation is also manifest in the phenomenon of scope splitting. The distributional restrictions n-words show in the Scandinavian languages confirm that n-words are subject to licensing conditions that are syntactic in nature.

Each of these phenomena is unexpected under the assumption that n-words are negative quantifiers. It is nevertheless possible to retain the negative quantifier analysis and employ a special mechanism to handle each of these phenomena, e.g. polyadic quantification for NC (Zanuttini 1991, Haegeman 1995, de Swart and Sag 2002); special kind of quantification for scope splitting (Geurts 1996, de Swart 2000); additional assumptions regarding syntactic structures to account for restricted distribution (Kayne 1998). But such a proceeding would simply seem to miss the generalisation.

In the approach argued for here, the three phenomena are all manifestations of the same underlying nature of n-words: n-words themselves are semantically non-ne-

gative and must be syntactically licensed by negation. Thus they correspond to morphosyntactic markers of sentential negation.

The cross-linguistic perspective taken on n-words has a further implication. In simple cases that do not exhibit NC or scope splitting, the analysis above is equivalent to the assumption that n-words are negative quantifiers. But if the phenomena discussed here are taken into account, such an analysis is superior in empirical coverage and can thus be taken to constitute the true nature of n-words. This means that there are no elements in natural language that correspond to negative quantifiers.

References

- Bech, G., 1955/57, *Studien über das deutsche Verbum infinitum*, Danske Videnskabernes Selskab, Kopenhagen.
- Beck, S., 1996, *Wh-movement and Transparent Logical Form*, PhD thesis, Universität Tübingen.
- Błaszczak, J., 2001, *Investigation into the Interaction between the Indefinites and Negation*, PhD thesis, Humboldt-Universität Berlin.
- Büring, D., 1997, *The Meaning of Topic and Focus: The 59th Street Bridge Accent*, Routledge, London/New York.
- Christensen, K. K., 1986, «Norwegian “ingen”: A case of post-syntactic lexicalization», in O. Dahl and A. Holmberg, (eds.), *Scandinavian Syntax*, Institute of Linguistics, University of Stockholm, pp. 21-35.
- de Haan, F., 1997, *The Interaction of Modality and Negation: A Typological Study*, Garland Publishing, New York/ London.
- de Swart, H., 2000, «Scope Ambiguities with Negative Quantifiers», in K. von Stechow and U. Egli (eds.), *Reference and Anaphoric Relations*, Kluwer Academic Publishers, pp. 109-132.
- and I. A. Sag, 2002, «Negation and Negative Concord in Romance», *Linguistics and Philosophy* 25, 373-417.
- Geurts, B., 1996, «On No», *Journal of Semantics* 13, 67-86.
- Giannakidou, A., 1997, *The Landscape of Polarity Items*, PhD thesis, Rijksuniversiteit Groningen.
- 2002, «N-words and negative concord». Ms, University of Chicago, to appear in *The Linguistics Companion*, Blackwell.
- Haegeman, L., 1995, «The Syntax of Negation» (Cambridge “Studies in Linguistics” 75), Cambridge U. P.
- Haraiwa, K., 2001, *Multiple agreement and the defective intervention effect*. Ms, MIT.
- Heim, I., 1982, *The Semantics of Definite and Indefinite Noun Phrases*, PhD thesis, University of Massachusetts at Amherst.
- Herburger, E., 2001, «The Negative Concord Puzzle Revisited», *Natural Language Semantics* 9, 289-333.
- Jacobs, J., 1980, «Lexical decomposition in Montague grammar», *Theoretical Linguistics* 7, 121-136.
- 1982, «Syntax und Semantik der Negation», *Deutschen Studien zur Theoretischen Linguistik* 1, Fink, München.
- Kadmon, N. and F. Landman, 1993, «Any», *Linguistics and Philosophy* 16(4), 354-422.
- Kayne, R., 1998, «Overt vs. covert movement», *Syntax* 1, 128-191.

- Krifka, M., 1995, «The semantics and pragmatics of polarity items», *Linguistic Analysis* 25, 209-258.
- Ladusaw, W. A., 1992, «Expressing Negation», in C. Barker and D. Dowty, eds, «Proceedings of SALT II», Ohio State Working Papers in *Linguistics* 40, Ohio State University, pp. 237-259.
- 1994, «Thetic and Categorical, Weak and Strong, Stage and Individual», in L. Santelman and M. Harvey, eds, *Proceedings of SALT IV*, Cornell University, Ithaca, NY, pp. 220-229.
- 1995, «Configurational Expression of Negation», in J. van der Does and J. van Eijck, eds, *Quantifiers, Logic and Language*, CSLI Publications, Stanford, pp. 203-223.
- Lahiri, U., 1998, «Focus and Negative Polarity in Hindi», *Natural Language Semantics* 6, 57-123.
- Laka, I., 1990, *Negation in Syntax: On the Nature of Functional Categories and Projections*, PhD thesis, MIT.
- McNally, L., 1998, «Existential Sentences without Existential Quantification», *Linguistics and Philosophy* 21, 353-392.
- Penka, D. and A. von Stechow, 2001, «Negative Indefinita unter Modalverben», in R. Müller and M. Reis, (eds.), «Modalität und Modalverben im Deutschen», *Linguistische Berichte, Sonderheft* 9, 263-286.
- Svenonius, P., 2002, *Licensing Negation in Norwegian*. Working Papers in Scandinavian Syntax.
- Zanuttini, R., 1991, *Syntactic Properties of Sentential Negation: A Comparative Study of Romance Languages*, PhD thesis, University of Pennsylvania.
- Zeijlstra, H., 2004, *Sentential Negation and Negative Concord*, PhD thesis, University of Amsterdam.

TYPES OF SYNCRETISM IN THE CLITIC SYSTEMS OF ROMANCE

Diego Pescarini

University of Padova

Abstract

In this paper I will discuss the hypothesis according to which every clitic system bears an elsewhere item, i.e., a non-specific clitic that can be inserted when the insertion of more specific items is blocked by independent constraints.

In my opinion the insertion of an elsewhere clitic accounts for different phenomena such as synthetic clusters (Bonet 1991, Harris 1994, Pescarini to appear) and absolute syncretisms (Calabrese 1994, Loporcaro 1995). I will support this claim on the basis of the data displayed by some Italian dialects.

0. Introduction

In this paper I present part of a wider research (Pescarini 2005, to appear) dealing with *synthetic clusters*, i.e. sequences of clitics displaying a mismatch between their morphological form and their syntactic functions (Bonet 1991, 1995; Harris 1994, 1997). For example in Italian (1) a cluster formed by a reflexive *si* and an impersonal *si* is not realized with a regular *si si* sequence, but as *ci si*, where an unexpected clitic *ci* appears instead of the reflexive one.

- (1) *Nel medioevo si si lavava raramente.
 √Nel medioevo ci si lavava raramente.
 in the Middle Ages REFL IMP washed rarely
 ‘In the Middle Ages they washed rarely’

In section 1 I will sketch briefly the analysis of synthetic clusters I have already presented and discussed in Pescarini (to appear). According to the basic claim of my analysis, the clitic inserted in the synthetic clusters is a non-specific exponent that is inserted by default when specific items are blocked by independent constraints.

In sections 2-4 I will explore a consequence of this hypothesis on the basis of the data shown by some Italian varieties. The prediction I will test is that the morpheme appearing in synthetic clusters is a syncretic exponent too.

1. Synthetic clusters

In order to account for synthetic clusters, we need to answer two distinct questions:

- a. Why cannot two identical clitics co-occur?
- b. Why is a given clitic (e.g. *ci*) inserted?

The tentative answer to the first question is partially consistent with Grimshaw (1997) suggesting that synthetic clusters are mainly due to a markedness constraint disallowing the same clitic to be inserted more than once in the same cluster. Her proposal is that this markedness constraint is a morphological form of OCP (Obligatory Contour Principle) and in my opinion this principle can be better defined if it is divided into two distinct conditions:

- a. *morphological consistency*: in order to trigger OCP, two items have to belong to the same morphological class (for example pronominal clitics);
- b. *phonological identity* (maybe just similarity): it is worth noting that this condition is not enough to trigger a morphological substitution like the one in (1).

Moreover, OCP does not apply directly on clitics, but on the morphemes forming clitics. Indeed, following Harris (1994) and Kayne (2000) I will suggest that clitics are decomposable and that their morphological structure is based on a template like (2) setting the order of person, gender, number and case exponents. Moreover, in my opinion, this template could be derived from a split-DP architecture—like the one in (3)—via head to head movement.

- (2) person # gender # number # case
 ↑ ↑
 1st, 2nd and non-person clitics
 3rd reflexive (locative, partitive, etc.)

- (3) [_{KP} case [_{NumP} number [_{GenP} gender [_{PersP} person]]]]

Clitics can exploit just parts of the template in (3). Indeed, first, second and third person reflexive clitics do not bear gender and number features—therefore they are realised as bare person exponents plus an epenthetic vowel—while non-person clitics (like locative and partitive) can be analysed as bare case exponents. 3rd person clitics (non reflexive) are supposed to exploit a larger portion of the structure as shown by the morphology of the Spanish clitic *las* where the 3rd person exponent is *l*, the feminine one is *a* and *s* is the plural marker.

Finally, it is worth noting that OCP is a markedness constraint, therefore it can be violated as shown by several Romance varieties displaying marked sequences of identical clitics. For example in my variety (in most Veneto dialects) the translation of the cluster in (1) is realized with a sequence of identical *se*. These clusters are not counterexamples, but just marked constructions that in other languages are blocked by OCP.

The second step of the analysis of clitic clusters (question b.) accounts for the substitution. Patterns of substitution have been usually accounted for through post-syntactic operations like those suggested by Bonet (1991, 1995) and Harris (1994, 1997). These operations would be responsible for changes of the feature representation of clitics before PF allowing the insertion of different and unexpected morphemes as the *ci* in (1).

But these operations are just language-specific assumptions and moreover, since each variety needs a particular set of operations, the whole inventory of operations

accounting for the Romance domain would be too wide and heterogeneous. Therefore, I have suggested an alternative explanation (Pescarini, to appear) based on a single and universal principle such as the Subset Principle, following Halle & Marantz (1993). This principle states that

- (4) an item is inserted in a syntactic node when:
 - a. the features representing the item are a subset of the features characterizing the node;
 - b. it is the most specific item among the underspecified ones.

Thus the process of insertion selects a finite number of under-specified candidates and then—in accordance with the second part of the Subset Principle—the most specific item wins.

The Subset principle allows a simple explanation for synthetic clusters: when OCP blocks the insertion of the optimal candidate, a less specific one is automatically inserted. My hypothesis is even stronger: I suggest that we normally insert a clitic without specifications, i.e. a clitic that is characterized by no morpho-syntactic feature. Indeed—in accordance with the Subset Principle—such a clitic is always under-specified, therefore it can be inserted by default in all the contexts where the insertion of more specific exponents is blocked. In accordance with this property, these items are labelled *elsewhere morphemes* (hereafter only *elsewhere*) and their main characteristic is their wide and complex distribution.

For instance, in the paradigm (5) there is no way to capture the distribution of *s* with a single feature matrix, but, if *-s* has no specification, it will be automatically inserted everywhere zero is too specific and the paradigm (5) will be thus economically captured.

(5)

Old French	sg.	pl.
nom.	-s	-s
obl.	-∅	-s

Inventory: -∅ ↔ sg. obl.
 -s ↔ elsewhere

In my opinion, the same machinery can account for synthetic clusters too: when a clitic is blocked by OCP, the elsewhere clitic is automatically inserted (6).

- (6) *clitic + clitic
 ↓
 elsewhere

This hypothesis can be supported by an independent piece of evidence. Indeed, on the basis of the Subset Principle, a diagnostic test can be formulated detecting the elsewhere. I have claimed that an elsewhere can replace other clitics because it lacks a specification, but how can we repair an OCP violation due to the co-occurrence of two elsewhere clitics?

If we replaced an elsewhere with another clitic, we would violate the Subset Principle since we would insert an overspecified clitic as shown in (7).

- (7) *~~elsewhere~~ + elsewhere
 ↓
 *clitic

Therefore, the only available repair is the deletion of an elsewhere (8).

- (8) *~~elsewhere~~ + elsewhere
 ↓
 *clitic

Summing up, this corollary of the Subset Principle entails that the *elsewhere* clitic is the clitic that cannot be replaced, but just deleted when it violates OCP.

Italian is consistent with this test: indeed, *si* is substituted by *ci* when it co-occurs with another *si*—as in (1)— but when two *ci* co-occur, one of them has to be deleted as shown in (9). (9) shows that a locative *ci* can occur with a 1st person singular pronoun *mi*, while it cannot with a 1st person plural *ci*.

- (9) A Roma mi ci porta Mario.
 A Roma ci (*ci) porta Mario.
 ‘Mario brings me/us to Rome’.

In conclusion, there is a perfect correlation between the pattern displayed in (1) and the test in (9).

2. Prediction

In this section I will explore another independent piece of evidence supporting the main hypothesis just discussed. According to their label, the evident characteristic of the elsewhere morphemes is their distribution: for instance, in the Old French paradigm in (5), the elsewhere exponent is indeed the best candidate for syncretism.

Indeed, we can consider synthetic clusters as cases of contextual syncretism since the same clitic covers different syntactic functions only in certain syntactic contexts. For instance, in Italian the *ci* exponent acquires the function of reflexive, only in the specific context described in (1).

It seems to me that the same process is responsible for absolute syncretism too, i.e. in a given dialect the same clitic can cover syntactic functions that in another dialect are covered by different exponents. For instance in the Brindisino dialect—spoken in the South-East of Italy—the *nci* exponent is used as a 1st person plural, 3rd person dative, locative and partitive clitic, while in the proto-Romance paradigm (and in many contemporary dialects) these functions are expressed by different exponents. Compare the paradigm in (10) and (11), both from Calabrese (1994).

(10)

Proto-Romance	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.					*(i)llu/a	*(i)llus/as
Indir obj.	*me/i	*nos	*te/i	*vos	*(i)lli	*(i)llis
Reflexive	*se					
Partitive	*(i)nde					
Locative	*hic / *(ec)ce+hic / *(i)nc+[i] / *(i)bi					

(11)

Brindisi	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.					lu/la	li/le
Indir obj.	me	nci	te	bbu	nci	
Reflexive	si					
Partitive	nci					
Locative	nci					

According to the Subset Principle, we can account for the distribution of syncretic exponents— like the Brindisino *nci*— assuming that they are elsewhere. On the basis of the discussion above I can indeed sketch the hypothesis that during the evolution from the proto-Romance system to the contemporary Brindisino one, some constraints (cf. Calabrese 1994) blocked the insertion of some etymological forms (usually 1st person plural and 3rd person dative) while the elsewhere clitic—that in Brindisino derives from the Latin locative particle **ince*— automatically replaced the blocked clitics giving rise to the paradigms of the Brindisi type.

Therefore, I am claiming that, if the Subset Principle is correct, there must be a deep relation between the processes responsible for contextual syncretisms (namely synthetic clusters) and those responsible for absolute syncretism: both processes are indeed based on the insertion of an elsewhere morpheme when external and independent reasons block the insertion of the appropriate one.

This correlation entails that the exponent involved in synthetic clusters is a syncretic marker too, in other words I predict that *in each variety there is a single clitic involved in both absolute and contextual syncretisms, namely synthetic clusters*. Italian seems to verify this

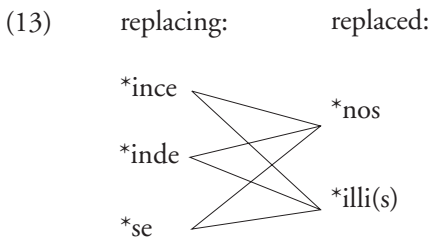
prediction because *ci* is the clitic appearing in synthetic clusters (1), *ci* is deleted when it violates OCP (9) and, finally, *ci* is the syncretic exponent of the Italian paradigm in (12).

(12)

Italian	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	mi	ci	ti	vi	lo/la	li/le
Indir obj.					gli/le	gli
Reflexive					si	
Partitive	ne					
Locative	ci					

3. Absolute syncretism

Before testing this prediction on the basis of the data I have collected, a general point has to be made on the patterns of absolute syncretism displayed by Italian dialects. Almost all the cases of absolute syncretism can be captured according to a simple scheme like the one in (13) that has to be read in parallel to the Proto-Romance paradigm in (10).



**Nos* and **illi* (on the right in the scheme) are the target of the syncretism, i.e. the exponents that are usually blocked and replaced, while the exponents on the left are the replacing items, i.e. the potential elsewhere.

We can capture the different patterns of substitution displayed by Italian dialects assuming that **ince*, **inde* and **se* have replaced one or two targets (**nos* and **illi*) giving rise to different and heterogeneous patterns like those in (14), (15) and (16).

(14)

Bologna	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	<i>m'</i>	<i>s'</i>	<i>t'</i>	<i>v'</i>	<i>al / la</i>	<i>i</i>
Indir obj.					<i>i</i>	
Reflexive					<i>s'</i>	
Partitive	<i>n', in'</i>					
Locative	<i>i</i>					

(15)

Palermo	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	mi	ni	ti	vi	u/a	i/(le?)
Indir obj.					ci	
Reflexive					si	
Partitive	ni					
Locative	ci					

(16)

Torino	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	me	ne	te	ve	lu/la	je
Indir. obj.					je	
Reflexive		se			se	
Partitive	ne					
Locative	je					

4. Data

In this section I will test the prediction formulated in section 2 regarding the correlation between absolute and contextual syncretisms. For each variety I will verify:

- a. which clitic appears in the opaque clusters;
- b. which clitic is deleted when it violates OCP;
- c. which clitic is a syncretic exponent.

If these three conditions are satisfied by a single clitic, it will be an important piece of evidence in favour of the main hypothesis discussed in section 1, i.e. the presence of an elsewhere morpheme in the clitic system of each variety.

The first dialects I will analyse are characterized by a single syncretic exponent. For example, the paradigm of the dialect spoken in Sarroch (province of Cagliari, Sardinia) shows a syncretic clitic deriving from the Latin reflexive pronoun **se*. Indeed, in the Sarroch dialect *si* is used as 3rd person reflexive and 1st and 2nd person plural.

(17) Sarroch

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	mi	si	ti	si	ddu/dda	Ddus/ddas
Indir. obj.					ddi	
Reflexive					si	
Partitive	ndi					
Locative	(n)ci					

This pattern of absolute syncretism correlates with the patterns of contextual syncretism shown in (18).

- (18) **ddi* + *ddu* → *si ddu* (**ddi ddu*)
 3.dat 3.acc
 ddi* + *ndi* → *si ndi* (ddi ndi*)
 3.dat part.

In the variety of Sarroch the regular form of these clusters would be *ddi ddu* and *ddi ndi*, but —like in Spanish— the dative clitic is replaced by the reflexive one. At the same time, the co-occurrence of two *si* markers is ruled out as predicted by my hypothesis. Indeed the translation of an Italian sentence with a reflexive *si* and an indirect object *si* is impossible.

Also the Vailate dialect (spoken in the province of Cremona, North Italy) shows a clear case of syncretism since the *ga* clitic —that derives from the Latin locative particle **hic*— is used as locative, 1st person plural and 3rd person dative (19).

(19) Vailate

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	ma	ga	ta	va	al/la	i /le
Indir. obj.		sa			ga	
Reflexive					sa	
Partitive	no					
Locative	ga					

In this dialect a reflexive + impersonal clitic is realized as *ga sa* where the reflexive clitic is represented by an unexpected *ga* exponent as in the Italian example in (1). At the same time two *ga* cannot co-occur. Therefore I conclude that in the Vailate dialect *ga* is the elsewhere morpheme.

The Fiorentino dialect (20) shows the same syncretism and the same clusters of Italian, therefore I will not repeat here the whole analysis (cf 1, 9, 12) even if I will briefly account for a point I have left open.

(20) Firenze

	1		2		3		
	sg	pl	sg	pl	sg	pl	
Dir. obj.	mi	ci	ti	vi	lo/la	li/le	
Indir. obj.		ci			gli/le		gli
Reflexive					si		
Partitive	ne						
Locative	ci						

In accordance with my proposal, the Italian cluster *le lo* should be realized as *ce lo*, because the elsewhere clitic of Italian is *ci*, cf. (9). But, actually, the spell out is *glielo* where *gli* (/ʎ/) is an allomorph of the pan-Romance 3rd person exponent *l*.

In my opinion this pattern is consistent with the second part of the Subset Principle stating that the most specific vocabulary item (among the underspecified ones) is inserted. Indeed, when *l* violates OCP, it is substituted by its allomorph *gli* instead of the elsewhere clitic *ci* because the former is, by definition, more specific than the latter. From this observation we can set different strategies of substitution. When OCP is violated, a clitic is substituted by:

- a. an allomorph, e.g. the Italian *gli*;
- b. an elsewhere clitic, e.g. *ci* in Italian, *si* in Sarroch, *ga* in Vailate;
- c. \emptyset , if the clitic is an elsewhere.

The main point is that the ranking of these strategies is not a stipulation, but it is due to the Subset Principle.

The Napoli paradigm and clusters—in (21) and (22)— show the same pattern of Fiorentino and Standard Italian, but with an interesting complication. Indeed the item inserted in the synthetic clusters (*nce*) is slightly different from the syncretic item of the paradigm that is *ce*.

(21) Napoli

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.					l _o – l _a	l _e
Indir. obj.	me	ce	te	ve	le – la	le - lloro
Reflexive					se	
Partitive	ne					
Locative	ce					

- (22) *le + l_o → nce lo
 3.dat 3.acc
- *le + ne → nce ne
 3.dat part.
- *se + se → nce se
 rifl. imp.

The presence of a nasal phoneme in the exponent in (22) would indeed suggest that the clustered item inserted in the clusters is more conservative than the one in isolation, as shown by the process in (23).

- (23) *ince > nce > ce

This situation is similar to the one displayed by two dialects spoken in Puglia (in the South East, examples in (24) and (25)); indeed, in the Barletta and Alberobello

dialects the form of the locative clitic is *ci* in isolation and *nci* in the clusters. At the moment I cannot account for this asymmetry.

- (24) ...ma nessuna *nge ne* dave (Barletta)
 ...annucite a veste ccu bbelle e mmettitangille. mettitece n-anedde u disete...
- (25) ...ma nessuna *ci* vulei da. (Alberobello)
 ...a cc-agge a ddice...
 ...a ggokka ccu bbelle mettitanigille. mettitece n-aniedde u disete...

The next dialects show two potential elsewheres in their paradigms, therefore they do not allow us to make a clear prediction, but they offer a piece of negative evidence. Indeed, a paradigm presenting two syncretic exponents (e.g. **ince* and **se*) allows us to exclude the third item (namely **inde*) as a potential elsewhere item. For instance, in the Arce dialect (spoken in Southern Italy, (26) the reflexive exponent is not a syncretic one, therefore —if my hypothesis is correct— I can exclude the possibility that it will be inserted in synthetic clusters. The contrary would be a strong counterexample falsifying my prediction and weakening the general hypothesis discussed in section 1.

(26) Arce

	1		2		3	
	Sg	pl	sg	pl	sg	pl
Dir. obj.	me	ne	te	ve	glie/la	glie/le
Indir. obj.		ci			glie	
Reflexive					se	
Partitive	ne					
Locative	ce					

Fortunately, Arce does not show clitic clusters where an unexpected reflexive marker is inserted, therefore there is a correlation between the clitic appearing in the clusters and one of the two candidates indicated by the paradigm (*ci* and *ne*).

- (27) **glie glie* → *ce glie*
 3.dat 3.acc
**gliene* → *ce ne*
 3.dat 3.acc

It is worth noting that both Napoli and Arce —(21,22) and (26,27)— replace the third person marker *l* with a locative elsewhere while Italian inserts the *gli* allomorph. These patterns are consistent with the discussion above since Napoli and Arce, unlike Italian, do not have any third person allomorph, therefore they must exploit the elsewhere.

Also the paradigm of the Catanzaro dialect (28) is characterized by two syncretic exponents.

(28)

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.		(n)ci			(l)u/(l)a	i/li
Indir. obj.	mi	ndi	ti	vi	(n)ci	
Reflexive					si	
Partitive	ndi/nda					
Locative	ci					

Moreover, Catanzarese shows two interesting clusters: in (29) the clusters are not repaired by a substitution, but one of the clitics is completely deleted.

- (29) **n*ci + ^lu → *n*ci
 3.dat 3.acc
- **n*ci + *ndi* → *n*ci
 3.dat part.

This pattern resembles those displayed by the dialect of Mascioni (Manzini & Savoia 2004) and by Barceloni, the Catalan variety analysed by Bonet (1991). At the moment, I have not sketched a single model accounting for both substitution (Sarroch, Vailate, Napoli, Arce, etc.) and deletions (Catanzaro, Mascioni, Barceloni), but, on the basis of the data I have collected, I can make a generalization: indeed, *the substitution always affects the clitic on the left of the cluster, while the deletion always affects the one on the right*. A model accounting for this descriptive generalization is still in progress.

5. Types of absolute syncretism

Neapolitan and the Arce dialect show the same kind of substitution displayed by standard Italian, even if their patterns of syncretism involve two clitics. This situation is quite frequent: in general when a dialect shows the co-occurrence of two syncretic exponents, the locative morpheme is preferred as elsewhere clitic.

This generalization is consistent with the patterns shown by the dialects of Puglia (South East). The syncretisms of these dialects can be described as a competition between two replacing items: the locative *n*ci and the partitive (*ndi*, *nde* or *nì*). In few cases one of the competitors wins as happens in Barese and Otrantino (30 and 31, from Calabrese 1994) where large portions of the paradigms are neutralized. But usually the paradigms of Puglia display a mixed distribution like the one in (32), from Calabrese (1994).

(30) Bari

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.					u/la	lə
Indir. obj.	mə	nǵə	tə	və	nǵə	
Reflexive					sə	
Partitive	nnə					
Locative	nǵə					

(31) Otranto

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.					lu/la	li/le
Indir. obj.	me	nde	te	bbu	nde	
Reflexive					si	
Partitive	nde					
Locative	nci					

(32) Campi - LE

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.					lu/la	li/le
Indir. obj.	me	nne	te	bbu	nci	
Reflexive					si	
Partitive	nne					
Locative	nci					

The main point is that, although the paradigms are characterized by this competition, synthetic clusters show a consistent preference for the locative clitic. Indeed, in almost all the dialects listed in table (33), the third dative clitic *li* is replaced by the locative *nci*.

(33)	<i>variety</i>	3.dat+3.acc	Spinazzola	nille, le_
	Vico Garganico	ce le	Laterza	ngi-i/ci le
	Vieste	cille	Martina Franca	ngille, ce ne
	Monte S. Angelo	celle	Grottaglie	nilo
	S. Marco la Catola	cele	Taranto	ncc ne
	Trinitapoli	ce/nge le	Maruggio	nci lu, nci ni
	Cerignola	ce li	Oria	nciulu, ncini
	Candela	ngille, nge ne	Maglie	sela
	Molfetta	ngiuue (?), ngere		

Only in three dialects —in grey in table (33)— the replacing clitics do not derive from **ince*. I have not collected the paradigms of these dialects yet, but my prediction is

that here the elsewhere clitics are *ne* and *si* respectively. Other cases of substitution from *li* to *ni* are displayed by some dialects of Calabria as Castrovillari, see Loporcara (1995).

This evidence allows me to restrict the area of my research: indeed, it seems to me that the dialects that have developed a *ne* elsewhere are spoken in the so-called Lausberg area, where cases of syncretic *ne* and *si* are actually found. Therefore, I can suggest that the lack of evidence showing an elsewhere *ne* is mainly due to a lack in the sample of dialects I have analysed.

However, there are theoretical reasons suggesting that, in general, the *ne* exponent is not a good candidate for the role of elsewhere. Firstly, I have already suggested (Pescarini 2005) that it can be due to the feature representation of these exponents, since in my opinion the partitive clitic is more specific than the locative one, therefore the locative is a better candidate in accordance with the Subset Principle.

But there is another point that has to be considered: maybe in some dialects the **nos* exponent has not been replaced by the partitive one —deriving from **inde*— but **nos* and **inde* have converged towards a common phonological form. In some dialects this process of mutual attraction is still in progress as shown by the Lecce and the Palermo dialects, (34) and (35). In these varieties there are indeed phonological differences between the form of the partitive clitic and the 1st person plural one: in Lecce (see the paradigm 34) the partitive displays a conservative *-nd-* consonantal cluster from **inde*, while Palermitan (35), that has assimilated *-nd-*, marks the distinction through the gemination of the nasal phoneme. The output of this diachronic process —in (37)— is represented by some Veneto dialects —in (36)— showing a perfect syncretism. My hypothesis is that in this case the syncretism is not due to a morphological substitution like those analysed in section 4, but it is caused by a phonological process giving rise to homophony.

(34) Lecce

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	me	ni	te	bu	lu/la	li/le
Indir. obj.					ni	
Reflexive					se	
Partitive	nde					
Locative	nci					

(35) Palermo

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.	mi	ni	ti	vi	u/a	i/(le?)
Indir. obj.					ci	
Reflexive					si	
Partitive	nni					
Locative	ci					

(36) Veneto

	1		2		3	
	sg	pl	sg	pl	sg	pl
Dir. obj.					o/a	i/e
Indir obj.	me	ne	te	ve	ghe	
Reflexive	se					
Partitive	ne					
Locative	ghe					

(37)

stage 1	stage 2	stage 3	stage 4
proto-Rom.	Lecce type	Palermo type	Veneto type
*nos >	ne >	ne >	ne
*inde >	nde >	nne >	ne

On the basis of these data I can suggest that sometimes the *ne* exponent is not an elsewhere clitic even if it is involved in syncretic patterns. The answer to this apparent paradox is that the syncretism of *ne* is due to a phonological evolution and not to the presence of an elsewhere morpheme.

This can account nicely for the *ne* vs *nci* asymmetry: since the *nci* syncretism is due to morphology, *nci* is a real elsewhere, therefore there is a high correlation between absolute and contextual syncretism because both require the insertion of an elsewhere clitic.

On the contrary, since some cases of *ne* syncretism are due to phonology, *ne* is not a real elsewhere and therefore it does not appear in synthetic clusters.

6. Paradigms without syncretisms

Finally, I will give some remarks on some Sardinian dialects even if they do not show any case of syncretism. Indeed, they are characterized by a conservative clitic paradigm, similar to the Proto-Romance one shown in (10). However, the Sardinian dialects I have observed show frequent cases of synthetic clusters as those in (38).

(38)

Nende. <i>bi.lu</i>	appo fattu un ibbagliu	(Ossi - SS)
Nanne. <i>bi.lu</i>	appo attu unu irbagliu	(Bitti - NU)
Nende. <i>bi.lu</i>	appo isbagliadu	(Posada - NU)
Narando. <i>si.ddu</i>	happo fattu unu sbagliu	(Baunei - NU)
Telling. <i>to-him.it</i>	aux make a mistake	

In three varieties the dative *li* is replaced by the locative *bi* deriving from the Latin locative **ibi*, while in Baunei the same exponent is replaced by the reflexive marker as in the Spanish *spourios se* phenomenon. These patterns are fully consistent with the cases presented by the Italian dialects examined above.

Maybe the Sardinian dialects represent an early stage of the evolution of Romance showing contextual syncretisms without absolute ones. But the synchronic data do not support this claim since you can find dialects characterized by the oppo-

site pattern as well. Perhaps trying to derive one phenomenon from the other is not very promising because they can be independent processes due to the same general principle that do not feed each other. Moreover they seem to operate in competition: indeed absolute syncretisms would be enhanced by contextual ones (both phonologically and morphologically based), while the presence of syncretic items in a system automatically increases the number of contexts violating OCP.

From a theoretical point of view, this competition is highly desirable, because it allows us to postulate a markedness constraint like OCP. Indeed, OCP would not make sense in a system without a tendency to assimilation. But, when the system—as in the case of Romance clitics—developes patterns of phonological attraction and morphological substitution, a markedness constraint arises as a natural counter reaction.

7. Conclusion

In this paper I have explored some consequences of a general hypothesis according to which every clitic system bears an elsewhere item, i.e. a non-specific clitic that can be inserted when the insertion of more specific items is blocked by independent constraints.

This repairing strategy accounts for synthetic clusters that, in synchrony, give rise to contextual syncretisms like those discussed in section 1. Moreover, in diachrony, it accounts for absolute syncretisms too (cf. section 2). Indeed, in my opinion, both contextual and absolute syncretisms are due to the insertion of an elsewhere clitic where independent constraints block the insertion of the appropriate item.

This claim is supported by the data presented and discussed in section 4 showing that in many Italian dialects there is a single clitic involved in both contextual and absolute syncretisms.

Finally, I have suggested that sometimes patterns of absolute syncretism are not due to a morphological substitution replacing an item with an elsewhere morpheme, but that they can be due to a phonological process giving rise to homophony. This hypothesis accounts for patterns of absolute syncretisms that do not involve any elsewhere clitics.

References

- Bonet, E., 1991, *Morphology after syntax: Pronominal clitics in Romance*. MIT: Doctoral dissertation.
- 1995, «Feature Structure of Romance Clitics», *NLLT* 13, 607-647.
- Calabrese, A., 1994, «Syncretism phenomena in the clitic systems of Italian and Sardinian dialects and the notion of morphological change». In J. N. Beckman, ed., *Proceedings of NELS 25: 2*. GLSA, University of Massachusetts, Amherst, 151-174.
- Grimshaw, J., 1997, «The Best Clitic: Constraint Conflict in Morphosyntax». In L. Haegeman, *Elements of Grammar*. Kluwer.
- Halle, M. & A. Marantz, 1993, «Distributed Morphology and the Pieces of Inflection». In *The View from Building 20*, ed. K. Hale and S. Jay Keyser. MIT Press, Cambridge, 111-176.
- Harris, J., 1994, «The syntax-phonology mapping in Catalan and Spanish clitics». In *MIT-WPL* 21.

- , 1997, «Why n'ho is pronounced [li] in Barceloní Catalan», *MITWPL* 30: 451-479.
- Kayne, R., 2000, *Parameters and Universals*. Oxford U. P.
- Loporcaro, M., 1995, «Un capitolo di morfologia storica italo-romanza: it. ant. *ne* 'ci' e forme meridionali congeneri», *L'Italia dialettale* 58: 1-48.
- Manzini, M. R. & Savoia, L., 2004, «Clitics: Cooccurrence and mutual exclusion patterns», in Rizzi, L. (ed.), *The structure of CP and IP*, New York, Oxford U. P.: 211-250.
- Pescarini, D., 2005, «Clitic Clusters and Morphological Repairs. Evidence from Italian, Spanish and Barceloní», *Studi linguistici e filologici on line* 3.1: 239-264.
- to appear, «The Morphology of Romance Clitic Clusters», *Proceedings of Camling 2005*, Cambridge Occasional Papers in Linguistics.

ARGUMENT-MAPPING AND EXTRACTION

Omer Preminger

Tel-Aviv University

Abstract

In this paper, I propose a unified account for argument-mapping and islandhood in the verbal domain, while casting new light on the notion *external argument*, as well as the interaction of Case and argument-mapping.

I argue that both types of syntactic merger (*set-merge* and *pair-merge*; Chomsky, 2004) are used for the merger of verbal arguments. The type of merger determines the islandhood of the argument at its base position, and along with Case-checking, it determines the internal or external mapping of the argument. Choice of the type of merger is governed by the feature-composition of the thematic role assigned to an argument, using the thematic feature system developed by Reinhart (2000).

This approach is shown to have clear empirical advantages, when compared to existing frameworks. In addition, it provides answers for previously unresolved questions about argument externality.¹

1. Introduction

This paper begins by examining *external arguments* in contemporary linguistic theory. Empirically, I will demonstrate that their distribution is not handled correctly within existing frameworks. Specifically, I will show that for the case of Object-Experiencer verbs and their intransitive alternates, no existing approach correctly predicts which argument will be external and when.

From a theoretical standpoint, I will demonstrate that the most basic question regarding *external arguments* has yet to be answered adequately —the question of what is special about their syntactic mapping.

In addition to these unresolved issues, I will show that some internal arguments behave syntactically as if they were external (a fact also exemplified by Object-Experiencer verbs).

To address these issues, I propose a system in which both types of syntactic merger assumed in minimalist syntax (*set-merge* and *pair-merge*; Chomsky 2004) are used for the merger of verbal arguments. The type of merger determines the islandhood of

¹ This work was originally based upon joint research with Alona Belikova, and has benefited greatly from comments and data provided by Eugenia Birger, Irena Botwinik-Rotem, Alex Grosu, Julia Horvath, Tal Kedar, Marijana Marelj, Aya Meltzer, Tanya Reinhart, and Tal Siloni.

the argument at its base position (arguments that have been moved are discussed separately). In addition, I argue that the interaction of *pair-merge* and accusative Case is what determines which (if any) of the arguments will be external.

This approach provides answers for the unresolved questions above, while also accounting for the surprising behavior of Object-Experiencer verbs.

Next, I will investigate the islandhood of arguments which are no longer at their base position, and argue for the empirical equivalent of the Freezing Principle (Wexler & Culicover 1977, 1980), which can be subsumed under the restriction of *Internal Merge* to *pair-merge*. This has the advantage of allowing the Subject Condition (Chomsky 1986, Huang 1982, Kayne 1984) to be derived instead of being a primitive, while also accounting for surprising facts regarding extraction in the Dative Shift paradigm.

Finally, I will show how this proposal, coupled with the late-merger approach of Fox (2002) and Fox and Nissenbaum (1999), allows *set-merge* and *pair-merge* to be construed as minimally different, contrasting only in extractability, without losing crucial predictions regarding the interaction of adjunction and Condition C of the binding theory.

2. Empirical and Theoretical State of Affairs

The first part of this paper will explore *external arguments*, as they stand in contemporary linguistic theory. I am not dealing here with the notion “subject” in general. Rather, I am referring to *external arguments* as identified by Belletti and Rizzi (1981) and Burzio (1986), *inter alia* —namely, the subjects of transitive verbs and unergatives, but not the subjects of unaccusatives or verbal passives. Of course, the two notions are not unrelated; an external argument, if present, will invariably be the argument that surfaces in subject position.

2.1. The Problem with Externality

Linguistic theory has explicitly recognized the importance of the distinction between external and internal arguments, at least as far back as the seminal work of Williams (1980). So much so, that in contemporary syntactic theory, a separate projection has often been posited for the sole purpose of merging the external argument into syntax: the little-*v* projection.²

Despite rich linguistic literature on external arguments, I will show that the following basic questions regarding argument externality have not yet been given satisfactory answers:

- (1) a. How is the external argument chosen from among the verb’s arguments?
- b. Once merged, what accounts for its particular syntactic behavior?
 In other words, what is syntactically special about external arguments?
- c. Why do certain internal arguments pattern with external arguments, in terms of syntactic behavior (see below)?

² As noted by Horvath and Siloni (2002), this projection has gone by many names: vP (Chomsky 1995b), VoiceP (Kratzer 1996), TrP (Collins 1997), and PredP (Bowers 1993). In the course of this work, I will be referring to it simply as “little-*v*”.

2.1.1. *How is the External Argument Chosen?*

2.1.1.1. Some Cross-Linguistic Data

Observe the following paradigm, showing an Object-Experiencer verb and its intransitive alternate, in English and Hebrew:

- (2) a. It worried the children that John was smoking.
 b. *hid'ig et ha-yeladim she-Dan me'ashen* (Hebrew)
worried ACC the-children that-Dan smoking
 'It worried the children that Dan was smoking.'
- (3) a. The children worried (that John was smoking).
 b. *ha-yeladim da'agu (she-Dan me'ashen)* (Hebrew)
the-children worried that-Dan smoking
 'The children worried (that Dan was smoking).'

This alternation provides several insights regarding the question in (1a), namely how the external argument is chosen. However, it is first necessary to establish which of the arguments in (2-3) are external and which are internal.

English does not mark the alternation in (2-3) morphologically. Therefore, it might be unclear which of the two versions is present in a given derivation. Hebrew proves helpful in this respect. The derivation in (3), in which the EXPERIENCER argument surfaces as a subject, is possible only with the form *da'ag(u)*. Conversely, the derivation in (2), in which the EXPERIENCER argument does not surface as a subject, is possible only with the form *hid'ig*. I will therefore use Hebrew to apply diagnostics of argument externality to each derivation.

I will start by examining the derivations in (3).

The default word order in Hebrew is SV(O). As shown by Reinhart and Siloni (2005) and Shlonsky (1987), the verb can precede the subject in one of two cases: *triggerred inversion*, in which some clause-initial XP licenses the inverse order ([XP V S]), or *simple inversion*, in which nothing precedes the verb ([V S]). Simple inversion is possible only when the subject is an internal argument. Thus, verbal passives (4a) and unaccusatives (4b) allow it, while unergatives (4c) do not:

- (4) a. *putru shlosha morim* (Hebrew)
fired.PASV three teachers
 'Three teachers were fired.'
- b. *higi'u shlosha necigim*
arrived three representatives
 'Three representatives arrived.'
- c. **rakdu shlosha yeladim*
danced three children

As shown below, the verb in (3b) (*da'agu* 'worried') patterns with the unergative in (4c) —it does not allow simple inversion, indicating that its EXPERIENCER argument is external:

- (5) **da'agu shlosha studentim* (Hebrew)
worried three students

Another diagnostic for argument externality in Hebrew is modification by a possessive dative constituent. As noted by Borer and Grodzinsky (1986), a dative consti-

tuent can serve as the possessor for the subject only if the subject is an internal argument. Therefore, it can serve as the possessor for the subjects of verbal passives (6a) and unaccusatives (6b), but not for the subjects of unergatives (6c):

- (6) a. le-mi butal ha-shi'ur? (Hebrew)
to-who cancelled.PASVt he-lesson
 'Whose lesson was cancelled?'
 b. le-mi nishbera ha-kos? c. *le-mi axlu ha-'orxim?
to-who broke the-glass to-who ate the-guests
 'Whose glass broke?'

As shown below, the verb in (3b) (*da'agu* 'worried') patterns with the unergative in (6c) —the dative constituent cannot be the possessor of the EXPERIENCER argument, indicating once again that the EXPERIENCER argument is external:

- (7) * le-mi da'ag ha-student (me-ha-macav) (Hebrew)
to-who worried the-student from-the-situation

The picture that emerges is therefore that in the derivations in (3), the EXPERIENCER argument is an external argument.

I will now turn to the derivations in (2).

When an argument is a clause instead of a DP, it is exempt from the (overt) Case requirements that apply to DP's. If the argument is also internal, it can form an expletive-associate chain with an expletive in subject position. In such a configuration, the argument remains in-situ in its internal position. Crucially, this option is not available for an external argument, whether it is clausal or not. Thus, this option is available with verbal passives (8a) and raising predicates (8b), but unavailable when the clause is an external argument (8c-d) (Reinhart 2001):

- (8) a. It was said [that John would be late].
 b. It seems (to Mary) [that John is late].
 c. *It biased the judge [that the defendant was wealthy].
 d. *It broke the window [that we were throwing rocks at it].

As shown by Reinhart (2001), the SUBJECT MATTER argument in (2) (*that John was smoking*) patterns with the arguments of verbal passives and raising predicates (8a-b), allowing the expletive-associate construction:

- (9) It worried the children [that John was smoking].
 This indicates that the SUBJECT MATTER argument is internal.

Another diagnostic, used by Reinhart (2001), involves so-called "backward anaphora":

- (10) a. ?? [His_i doctor] visited [every patient]_i.
 b. [His_i health] worried [every patient]_i.

The marginality of (10a) is a standard case of weak-crossover. Following Reinhart (2001), what salvages (10b) is that the SUBJECT MATTER argument (*his health*) is an internal argument. This can bleed weak-crossover effects, since as an internal argument, *his health* is base-generated in a position which is c-commanded by *every*

patient, and it can then reconstruct to that position to receive its bound-variable interpretation at LF. In that case, *every patient* no longer needs to undergo Quantifier Raising, and weak-crossover is averted.

The picture that emerges is that in the derivations in (2), the SUBJECT MATTER argument is internal. As for the post-verbal accusative-marked EXPERIENCER argument in (2), it is internal as well. Its accusative marking may be sufficient evidence of this, but the same can be shown using the possessive dative test (similar to (6)):

- (11) le-mi ha-macav hid'ig et ha-yeladim (Hebrew)
to-who the-situation worried ACC the-children
 'Whose children did the situation worry?'

The felicity of the possessive dative construction indicates that the possessed argument (in this case, the EXPERIENCER argument *ha-yeladim* 'the children') is indeed an internal argument.

To summarize, the derivations in (2) lack external arguments. Specifically, the EXPERIENCER arguments in (2a-b) are internal. The EXPERIENCER arguments in (3a-b), on the other hand, are external.

2.1.1.2. Possible Explanations

In this section, I will examine several possible explanations for the data presented above. Specifically, the aim is to predict the distribution of *argument externality*: under which conditions a given argument will be mapped as external, and under which conditions it will be mapped as internal.

2.1.1.2.1. Thematic Explanations

The thematic roles involved in (2) and in (3) are the same (presumably, EXPERIENCER and SUBJECT MATTER; see Pesetsky 1995, Reinhart 2001). This means that question (1a) (how the external argument is chosen) cannot be answered in terms of thematic information alone.

First, consider positing a projection such as *little-v*, and restricting the set of thematic roles it can assign to the argument it merges (the external argument). In such a system, if a thematic role is part of the given set, it will be merged by *little-v*, and if not, it will be merged by the verb itself. For example, the AGENT thematic role will almost certainly be part of this set, as AGENT arguments are invariably mapped as external arguments. However, as pointed out by Horvath and Siloni (2002), the EXPERIENCER role is either part of this set or not, so such an approach cannot explain why the EXPERIENCER argument is internal in (2), but external in (3).³

Second, since there is no difference between (2) and (3) in any of the thematic roles involved, even accounts in terms of thematic hierarchies will fail to explain these facts.

³ As Horvath and Siloni (2002) point out, this state of affairs represents more than just a case that *little-v* cannot account for. It is in fact a counter-argument for the *Little-v* Hypothesis altogether — the two verbs in (2-3) are clearly derivationally related, and it would be completely ad-hoc to assume that in (2), the Experiencer role is associated with the verbal head, while in (3), the same role is introduced by a separate head (*little-v*).

2.1.1.2.2. *Burzio's Generalization*

Since the formulation of Burzio's generalization (Burzio 1986), the presence of an external argument has been tied to the existence of accusative Case on the verb. In some cases, the little-*v* projection has been the mechanism used to encode this generalization in the grammar (cf. *v**P vs. *v*P; Chomsky 2001, 2004).

However, as noted by Reinhart (2001), the picture that emerges in (2-3), repeated below, constitutes an exception to Burzio's generalization:

- (12) a. It worried the children that John was smoking.
 b. hid'ig et ha-yeladim she-Dan me'ashen (Hebrew)
worried ACC the-children that-Dan smoking
 'It worried the children that Dan was smoking.'
- (13) a. The children worried (that John was smoking).
 b. ha-yeladim da'agu (she-Dan me'ashen) (Hebrew)
the-children worried that-Dan smoking
 'The children worried (that Dan was smoking).'

The verbs in (12) lack external arguments but have accusative Case (overtly manifested in Hebrew (12b)), while the verbs in (13) have external arguments but lack accusative Case. Therefore, an approach which associates externality with the presence of accusative Case, while capturing an important linguistic tendency, will fail to account for the facts above.

2.1.1.2.3. *The Theta System*

In the Theta System, as developed by Reinhart (2002), arguments are given syntactic mapping based on thematic information and derivational relations between lexical entries. Without going into the details of the analysis here, the result is that EXPERIENCER arguments are mapped as external arguments, unless some other argument preempts this mapping. In the Theta System, the thematic roles which can preempt an EXPERIENCER's external mapping are AGENT, CAUSE, and SENTIENT —none of which are present in the derivations in (2-3). Thus, the different behavior of the EXPERIENCER argument in (2) and in (3) poses a problem for the Theta System as well (as Reinhart herself notes; see Reinhart 2001).

2.1.1.3. Intermediate Summary

It therefore appears that there is no analysis currently available which is capable of dealing with the mapping facts exemplified in (2-3). This, despite the fact that the constructions in (2-3) would hardly be considered cumbersome or uncommon.

2.1.2. *What is syntactically special about External Arguments?*

Another problem facing the notion of argument externality is that of the syntactic encoding of this property. Namely, once syntactic structure is formed, what is the inherent difference between the mapping of an external argument and an

internal one, which causes the two to react differently to various syntactic operations?

The framework of Bare Phrase Structure (Chomsky 1995a) aims to eliminate stipulated levels of X-bar structure. Thus, it abandons the primitive distinction between specifier and complement, viewing them instead as derivative structural observations.

Horvath and Siloni (2002) argue for the rejection of the Little-*v* Hypothesis. However, as they point out, this leaves open the question of how to map the sole argument of an unergative verb in a position different from the sole argument of an unaccusative verb (given the assumptions of Bare Phrase Structure), as the behavior of external and internal arguments is known to differ. Consider two simple cases:

- (14) a. John ran. b. John arrived.

Linear order is not considered to be part of narrow-syntax. Thus, given the rejection of Little-*v* Hypothesis, the VP in each of the cases (14a-b) would be as shown in (15a-b), respectively:

- (15) a. [_{VP} John [_V run]] b. [_{VP} John [_V arrive]]

What differentiates *John* in (15a) from *John* in (15b), in terms of their syntactic status? If both are the only argument of V^0 , what accounts for the difference in syntactic behavior between the two?

2.2. Extraction: More Data

Yet another piece of the puzzle, which turns out to be related to argument-mapping, can be found in some subtle properties of extraction from verbal arguments.

It is well known that external arguments block extraction. This is covered, though not exhaustively of course, by the Subject Condition (Chomsky 1986, Huang 1982, Kayne 1984). The picture regarding internal arguments, however, is more complex:

- (16) a. Who₁ did the counselor meet [teachers of t₁]?
 b. *Who₁ did the situation worry [teachers of t₁]?
 (adapted from Johnson 1992, and Landau 2001)

- (17) a. [Iz kakogo universiteta]₁ vy priglasili [studentov t₁]? (Russian)
from which university you.PL invited students
 ‘Which university did you invite the students of?’
 b. *[Iz kakogo universiteta]₁ novosti vzvolnovali [gostej t₁]?
from which university news worried visitors
 (adapted from Belikova & Preminger 2004)

Surprisingly, both in (16b) and in (17b), an EXPERIENCER argument blocks extraction despite being internal.⁴

⁴ Johnson (1992), who is also quoted by Landau (2001), states that this is only true when the subject is non-agentive. English speakers that I have checked with do not share this judgment. The same is true of Russian speakers consulted by Belikova and Preminger (2004). Moreover, even if Johnson is co-

One could conceivably seek an explanation for this in terms of structural configuration. Object-Experiencer verbs are three-place predicates (their thematic roles are CAUSE, EXPERIENCER, and SUBJECT MATTER; see Pesetsky 1995, Reinhart 2001). As such, they can be argued to project a VP-shell structure (Larson 1988). As a result, the EXPERIENCER argument would be mapped to a specifier position.

As observed by Huang (1982) and Kayne (1984), most specifiers block extraction. This has been captured by various theoretical mechanisms, such as the Condition on Extraction Domains (CED; Huang 1982). Hence, one could argue that the reason for the islandhood of the EXPERIENCER argument in (16b) and (17b) is structural —namely, that the EXPERIENCER argument occupies a specifier position.

However, such an account is insufficient, as can be seen below:

- (18) a. Who₁ did you give [a picture of t₁] to John?
 b. Who₁ did you give a picture [to acquaintances of t₁]?
 (adapted from Landau, 1994)

For (18a-b), any phrase-structure which assumes Binary Branching (Kayne 1984) (including Larsonian VP-shells) will have at most one complement position in which an argument of *give* can be merged.⁵ Therefore, explaining the blocking of extraction in (16b) and (17b) in terms of structural configurations (i.e. specifier vs. complement) is at odds with the data in (18a-b), which constitutes an obvious exception to the generalization that the CED attempts to capture.

The conclusion is that neither the external vs. internal argument distinction, nor the specifier vs. complement distinction, are able to predict extractability in these cases (for a discussion of Dative Shift, see section 3.4).

2.3. Towards a Generalization

So far, I have shown that both the distribution of argument externality and the status of arguments in terms of islandhood defy explanation using currently available frameworks.

To reach a satisfactory account, let us start by taking another look at the empirical facts at hand. Given the data presented in sections 2.1-2.2, three groups of verbal arguments can be identified:

- (19) a. *A-arguments*: arguments which are always mapped externally (e.g. AGENT)
 b. *B-arguments*: arguments which are sometimes mapped externally and sometimes mapped internally (e.g. EXPERIENCER; see (2-3))

rect, this would still be a result that demands explanation, given that the Experiencer argument is internal in both cases.

⁵ To be exact, the structure will have at most one “pure complement” position. I use the term “pure complement” to denote a node from which a path to C⁰ exists, such that this path crosses only nodes of complementation. Obviously, even given Binary Branching, adjunction can introduce a complex constituent which contains other complement nodes. However, none of these will qualify as “pure complements”, given this definition.

- c. *C-arguments*: arguments which are always mapped internally (e.g. THEME)⁶

Consider the interaction between externality and accusative Case, for each of the groups defined above:

- *A-arguments* never check accusative Case
- *B-arguments* are mapped externally when they do not check accusative Case (recall (2-3), in 2.1.1)
- *C-arguments* are never mapped externally

This is normally taken to be a result of some principle along the following lines (which may or may not be derived from other properties of the verbal Case-checking system):

- (20) Only when an argument is internal, can it check accusative Case.

However, consider the possibility that cause and effect are actually reversed:

- (21) When a *B-argument* does not check accusative Case, it is external.

In fact, since *A-arguments* never check accusative Case, the generalization stated in (21) can be expanded somewhat:

- (22) When an *A/B-argument* does not check accusative Case, it is external.

Next, consider extraction, taking into account the data from section 2.2.

The Subject Condition (Chomsky 1986, Huang 1982, Kayne 1984) covers the blocking of extraction from two of the groups defined above:⁶

- from arguments that are always mapped externally (*A-arguments*)
- from arguments that can be external (*B-arguments*), when they are indeed external

However, *B-arguments* block extraction regardless of external/internal mapping (recall (16b) and (17b) in section 2.2, in which an EXPERIENCER argument blocks extraction even when mapped internally).

Therefore, splitting verbal arguments into these groups is advantageous in capturing the properties of two seemingly distinct phenomena — argument-mapping and extraction:

⁶ The actual picture regarding Theme arguments is a bit more complex. There is a class of verbs known as Emission Verbs or Theme Unergatives (Horvath & Siloni 2002, Levin & Rappaport-Hovav 1995, Reinhart 2000, 2002), which are one-place unergatives that select a Theme argument.

The existence of external Theme arguments poses a problem for the proposed framework. This can be handled by restricting the discussion to multi-place predicates (i.e. verbs with two or more arguments), or verbs derived from multi-place predicates (such as the unaccusative alternate of a transitive verb; see Reinhart 2000, 2002).

It is important to note, however, that the existence of such verbs is equally problematic for other frameworks, including those discussed in section 2.1.1.2, and would therefore require a similar caveat.

— *A/B-arguments*:

- block extraction (regardless of mapping)
- mapped externally when they fail to check accusative Case

— *C-arguments*:

- allow extraction⁷
- never mapped externally

Given an accurate definition for *A-arguments*, *B-arguments*, and *C-arguments*, a unified account of argument-mapping and extraction could be formulated.⁸

3. Proposal

3.1. Background: Feature Composition of Thematic Roles

Reinhart (2000) proposes decomposing thematic roles into features. Under this view, the standard thematic roles (AGENT, THEME, EXPERIENCER, etc.) are not primary entities of the grammar, but rather labels for feature clusters.

(23) *Feature composition of thematic roles* (Reinhart 2000):

- ±c: whether or not the argument in question is responsible for causing change (in the context of the given event)
- ±m: whether or not the mental state of the argument in question is relevant (to the given event)

Every thematic role is a cluster of these features. In a given cluster, each feature can be valued for /+ or /-, or unvalued (in which case both interpretations of the feature are possible).

The conventional thematic roles are composed as follows:

(24) *Thematic roles* (Reinhart 2000):

	+m	no /m	-m
+c	Agent	Cause	Instrument
no /c	Sentient	∅	Subject Matter
-c	Experiencer	Goal/Source	Theme

As shown by Reinhart (2000, 2001, 2002), this system proves advantageous in predicting various grammatical properties which otherwise defy explanation.

As an example, consider the case of unaccusative verbs. As argued by Reinhart (2000), precise definition of the set of unaccusative verbs is a desideratum, both in terms of theoretical completeness, and more importantly, in terms of learnability. Though the set of

⁷ Of course, this does not exhaust the cases covered by the Subject Condition. Even *C-arguments* may block extraction when they appear in subject position. As I will argue later, this is a separate issue, which has to do with the islandhood of moved constituents. See section 3.4.

⁸ To be exact, *C-arguments* can block extraction, but only in case they move to subject position. See note 7.

unaccusative verbs is obviously finite, it is quite large. Therefore, acquiring each member of the set separately seems extremely cumbersome. Furthermore, in English, there is neither morphological marking of unaccusative verbs, nor very substantial syntactic evidence to distinguish verbs in this set (25a) from one-place unergative verbs (25b):

- (25) a. She₁ moves t₁ gracefully. b. She dances gracefully. (Reinhart 2000)

Reinhart demonstrates that attempts to define the set of unaccusative verbs in terms of their aspectual properties (Borer 1994, Van Hout 1995, Van Valin 1990, *inter alia*) cannot be considered successful, while other prevailing accounts (Levin & Rappaport-Hovav 1994, 1995, Pesetsky 1995) also fail to account for the full range of facts (see Reinhart, 2000, for the precise argumentation in each case).

However, as Reinhart shows, the feature system in (23-24) does allow precise definition of the set of unaccusative verbs. The definition is given below:

- (26) A verb is unaccusative *iff* its concept includes a CAUSE ([+c]) role, and that role is reduced (not realized).

Under normal circumstances, (26) means that the verb has an alternate which has an additional CAUSE role, as in (27):

- (27) a. [The vase]_{THEME} broke. b. [The wind]_{CAUSE} broke [the vase]_{THEME}.

However, there may be instances in which a specific entry is non-existent in a given language (to be exact, it is frozen, existing in the lexicon but not usable in a syntactic derivation; Chierchia 1989). Consider *fall*, which is unaccusative but lacks an English counterpart which has an additional CAUSE ([+c]) role. It turns out that such a lexical entry does exist in Hebrew:

- (28) a. ha-cincenet₁ nafla t₁ (Hebrew)
 the-jar *fell*
 ‘The jar fell.’
 b. ha-ruax hipila et ha-cincenet
 the-wind ‘made fall’ ACC *the-jar*
 ‘The wind made the jar fall.’

As can be seen in (28b), Hebrew has a transitive alternate of *nafal* ‘fell’, which has an additional CAUSE ([+c]) role.

Now recall (25a-b), repeated below:

- (29) a. She₁ moves t₁ gracefully. b. She dances gracefully. (Reinhart 2000)

Given Reinhart’s system, the child knows that (29a) is unaccusative, since the concept *move* has a CAUSE ([+c]) role which is unrealized here. The concept *dance*, however, does not have a CAUSE ([+c]) role. Therefore (29b) must have an unergative derivation.

Notice that while *dance* does have a causative form, as in (30a), the additional role in this form is an AGENT ([+c +m]), and not a CAUSE ([+c]), as shown by the ungrammaticality of (30b):

- (30) a. John danced Mary around the room.
 b. *[The enthusiasm]/[Her enthusiasm] danced Mary around the room.

The feature system presented in (23) thus allows a precise definition of the set of unaccusative verbs. Many other cases where linguistic analysis can benefit from this system are cited by Reinhart (2000, 2001, 2002).

3.2. The Proposed System

To account for the generalizations reached in section 2.3, I propose the following system:

- (31) *Lexical Determination of Merger:*
- a. *uniform [-] clusters:* merged into syntax via *set-merge*, the structure-building operation standardly assumed for canonical complementation (i.e. complements of functional heads, canonical direct objects, etc.) (Chomsky 2004)
→ creating a domain which is accessible for extraction
 - b. *other clusters:* merged into syntax via *pair-merge*, the structure-building operation standardly assumed for adjuncts (Chomsky 2004)
→ creating an island for extraction

In other words, the feature-composition of the thematic role discharged on a given argument determines the type of syntactic merger which inserts that argument into the derivation.

THEME ([-c -m]), GOAL/SOURCE ([-c]) and SUBJECT MATTER ([-m]) roles are uniform [-] clusters, and therefore arguments receiving these roles will be merged via *set-merge*. All other arguments will be merged via *pair-merge*.⁹

Note that the fact that an argument is *pair-merged* does not mean it is late-merged (Fox & Nissenbaum 1999, Fox 2002). On the contrary, *pair-merge* is subject to the same cyclicity conditions as *set-merge*. The issue of late-merger is orthogonal to the *set-merge/pair-merge* dichotomy. See 3.6 for further discussion.

3.3. The Predictions

In this section, I will show that the proposal above correctly predicts the facts presented in sections 2.1-2.2, regarding argument-mapping and extraction —facts which previously defied explanation.

3.3.1. Extraction

Let us examine how the proposed system accounts for the extraction-related facts shown in section 2.2. Recall (16a-b), repeated below:

- (32) a. Who₁ did the counselor meet [teachers of t₁]?
b. *Who₁ did the situation worry [teachers of t₁]?

⁹ Ideally, this account would also explain the difference between A-arguments and B-arguments more clearly —namely, why A-arguments never check accusative Case (and are therefore always external). See Preminger (2005) for a detailed account.

The bracketed argument in (32a) receives the role of THEME ([-c -m]), which is a uniform [-] cluster. It is therefore merged via *set-merge*, predicting its accessibility to extraction. The bracketed argument in (32b), however, receives the role of EXPERIENCER ([-c +m]), which is not a uniform [-] cluster. It is therefore merged via *pair-merge*, predicting its islandhood. Compare this with (18a-b), repeated below:

- (33) a. Who₁ did you give [a picture of t₁] to John?
 b. Who₁ did you give a picture [to acquaintances of t₁]?

The bracketed arguments in (33a) and in (33b) receive the roles of THEME ([-c -m]) and GOAL ([-c]), respectively. Both are uniform [-] clusters. Therefore, both are merged via *set-merge*, predicting their accessibility to extraction (for a discussion of Dative Shift, see section 3.4).

The advantage of divorcing accessibility to extraction from the external/internal mapping of arguments, and from their structural configuration (i.e. specifier vs. complement), becomes clear. Both of the bracketed arguments in (32a-b) are internal, but their extraction-related properties differ. At most one of the two bracketed arguments in (33a-b) can be in complement position, but their extraction-related properties are the same. Therefore, CED-inspired accounts for extractability, which are based on either argument externality or the specifier/complement distinction (Huang 1982, Kayne 1984) cannot deal with the data presented here. Unlike such accounts, the proposed system correctly predicts this behavior.

Notice that this proposal essentially incorporates Kayne's (1994) intuition that specifiers are an instance of adjunction, since given the proposed system, most verbal arguments that occupy a specifier position will indeed be *pair-merged*. The two approaches diverge precisely on cases such as (33a), above, which represents felicitous extraction from a specifier position. Under this proposal, (33a-b) contain a specifier which is not *pair-merged*, and therefore does not pattern with adjuncts, in terms of islandhood.

We are now in a position to answer question (1c):

Question (1c): Why do certain internal arguments pattern with external arguments, in terms of syntactic behavior (i.e. accessibility to extraction)?

Answer: Such internal arguments behave this way because they are *pair-merged* (on par with external arguments —as discussed in 3.3.2, below).

3.3.2. Externality

Let us now turn to the facts relating to argument-mapping, as presented in section 2.1. Recall the argument groups identified in section 2.3:

— *A/B-arguments:*

- block extraction (regardless of mapping)
- mapped externally when they fail to check accusative Case

— *C-arguments:*

- allow extraction
- never mapped externally

The system proposed in 3.2, in addition to handling the extraction data presented in section 2.2, necessarily fixes the classification of arguments into these groups. For example, arguments receiving thematic roles which are uniform [-] clusters are *set-merged*, allowing extraction. Therefore, they must be *C-arguments*. Likewise, all other arguments (which are *pair-merged*, blocking extraction) must be *A/B-arguments*.

Since *A/B-arguments* and *C-arguments* have not only extraction-related properties, but also mapping-related properties, this classification results in precise predictions regarding how these arguments will be mapped:

- Arguments receiving roles which are uniform [-] clusters (i.e. *C-arguments*) must always be mapped internally.
- Arguments receiving other roles (i.e. *A/B-arguments*) must be mapped internally if and only if they check accusative Case.

One can now check if the mapping of these arguments to external/internal positions is in fact consistent with these predictions. Recall (2-3), repeated below:

- (34) a. It worried the children that John was smoking.
 b. hid'ig et ha-yeladim she-Dan me'ashen (Hebrew)
worried ACC the-children that-Dan smoking
 'It worried the children that Dan was smoking.'
- (35) a. The children worried that John was smoking.
 b. ha-yeladim da'agu she-Dan me'ashen (Hebrew)
the-children worried that-Dan smoking
 'The children worried that Dan was smoking.'

As shown in 2.1.1, the EXPERIENCER arguments in (35a-b) are external, while the derivations in (34a-b) do not contain external arguments. Given the thematic feature system in 3.1, the thematic roles involved in (34-35) are composed as follows:

- EXPERIENCER: [-c +m]
- SUBJECT MATTER: [-m]

The SUBJECT MATTER role is a uniform [-] cluster. The argument receiving this role is therefore a *C-argument*. As such, it should never surface as external, regardless of whether or not it checks accusative Case.

The EXPERIENCER role is a mixed cluster. The argument receiving this role is therefore an *A/B-argument*. Thus, it should be mapped externally precisely when it does not check accusative Case.

This is exactly the picture that emerges in (34-35). The SUBJECT MATTER argument is internal in both (34) and (35). As for the EXPERIENCER argument, its mapping is indeed dependent on accusative Case. The verb in (34) has accusative Case, which is checked by the EXPERIENCER argument (as can be seen overtly in (34b)). The EXPERIENCER argument is therefore mapped internally in (34). The verb in (35), on the other hand, does not have accusative Case. This allows the EXPERIENCER argument to be mapped externally (see 2.1.1 for the relevant diagnostics).

One can now answer the two remaining questions, (1a-b):

Question (1a): How is the external argument chosen?

Answer:

(36) *External Argument:*

A *pair-merged* argument that does not check accusative Case.

Question (1b): What is syntactically special about external arguments? Specifically, what is the syntactic difference between how the sole argument of an *unergative* verb is mapped and how the sole argument of an *unaccusative* verb is mapped?

Answer: The difference is in the type of operation that attaches the argument to the syntactic tree: *set-merge* vs. *pair-merge*.

The formulation of externality without reference to the specifier/complement distinction allows different mapping for external and internal arguments without stipulated X-bar structure or little-*v*, which as discussed in 2.1.2, is a desirable result.¹⁰

3.4. Movement and Islandhood: Completing the Picture

The brief discussion of ditransitive verbs in sections 2.2 and 3.3.2 did not deal with the phenomenon of Dative Shift. This section will deal with Dative Shift and related issues.

3.4.1. Dative Shift and Extraction: The Data

In the interest of perspicuity, I will adopt the terminology used by Larson (1988), i.e. *Dative Construction* to refer to the derivation containing overt dative marking (as in (37), below), and *Double Object Construction* to refer to the result of Dative Shift, where no overt dative marking is visible (as in (38), below).

Recall the data regarding extraction in the Dative Construction (18), as presented in 2.2 and repeated below:

- (37) a. Who₁ did you give [a picture of t₁] to John?
 b. Who₁ did you give a picture [to acquaintances of t₁]?
 (adapted from Landau 1994)

As shown above, in the Dative Construction, both the *THEME* and the *GOAL* arguments are possible domains for extraction. As argued in 2.2, this is important counter-evidence for the validity of a generalization on extraction which relies on the specifier/complement distinction, such as the CED (Huang 1982).

¹⁰ The insight that uniform [-] clusters form a natural class, and that this class behaves in a distinct fashion with respect to merger, is due to Reinhart (2002). In her system, however, belonging to this class of thematic roles has different consequences than in the system proposed here. Also, her system further sub-divides the other thematic roles, so the result is a system with three natural classes, and not two, as proposed here.

The facts regarding extraction in the Double Object Construction are different:

- (38) a. *Who₁ did you give [acquaintances of t₁] a picture?
 b. Who₁ did you give John's acquaintances [a picture of t₁?]
 (adapted from Landau 1994)

Somewhat surprisingly, Dative Shift appears to affect the islandhood of the GOAL argument (the bracketed argument in (38a)). I will account for this property in the following sub-section.

3.4.2. *The Interaction of Movement and Islandhood*

There are two issues, related to the current proposal and to islandhood effects, which have remained unexplained so far:

- Dative Shift extraction facts, as shown in 3.4.1, above
 - residual Subject Condition effects —as is well known, even arguments which allow extraction at their base positions, are islands when moved to subject position. Notice the difference between extracting from the bracketed argument in (39a) and in (39b), as shown in (40a) and (40b), respectively:
- (39) a. It seems strange to Mary [that John would enjoy rock music].
 b. [That John would enjoy rock music] seems strange to Mary.
- (40) a. [Which music]₁ does it seem strange to Mary [that John would enjoy t₁?]
 b. *[Which music]₁ does [that John would enjoy t₁] seem strange to Mary?

It is clear what these two issues have in common. If one accepts Larson's (1988) analysis of Dative Shift, the Double Object Construction (38) involves movement of the GOAL argument from its thematic position (on par with verbal passivization; *op. cit.*). Similarly, the difference between (40a) and (40b) is whether or not the bracketed argument is at its base position.

These facts are reminiscent of Wexler and Culicover's Freezing Principle:

- (41) *The Freezing Principle* (adapted from Wexler & Culicover 1977, 1980):
 A constituent which has undergone movement becomes an island.

More recent work has introduced the view that movement is simply *Internal Merge*, meaning the merger of a syntactic object that is already present in the derivation, into the derivation once more (Chomsky 2004, 2005). Given this, the Freezing Principle itself can be derived from the following restriction:

- (42) *Internal Merge* is always *pair-merge*.

Furthermore, as shown below, adopting (42) has the advantage of making the Subject Condition (Chomsky 1986, Huang 1982, Kayne 1984) derivable, instead of being a grammatical primitive.

3.5. Deriving the Subject Condition

As discussed in 2.1.2, the move to Bare Phrase Structure (Chomsky 1995a), means abandoning the primitive distinction between specifier and complement. Given this,

the Subject Condition (Chomsky 1986, Huang 1982, Kayne 1984) can no longer be taken to be an instantiation of the inherent properties of the specifier position.¹¹

In this respect, it is important to note that there is cross-linguistic variation on whether overt movement to subject position (Spec-T⁰) is obligatory.

Accounting for the Subject Condition therefore involves answering two separate questions:

- (43) a. What prevents extraction from an argument that has moved to TP?
 b. What prevents extraction from external arguments at their base positions?

Given the current proposal, an answer to (43b) is readily available. Recall the answer to question (1a) (how the external argument is chosen), namely the definition in (36), repeated below:

- (44) *External Argument*: A *pair-merged* argument that does not check accusative Case.

The fact that an external argument (if one exists) blocks extraction at its base position, is a result of the type of merger that attaches it to the syntactic tree —namely *pair-merge*, which creates a domain that is inaccessible to extraction.

The answer to (43a) can be found in the restriction (42) reached in section 3.4.2, and repeated below:

- (45) *Internal Merge* is always *pair-merge*.

Thus, if an argument has moved to TP, it should become an island by virtue of that movement.

One would therefore predict that if the subject is *set-merged* (as would be the case in unaccusative verbs and verbal passives, given the current proposal), and the language allows the subject to stay in-situ, extraction from it should be possible. This is indeed the case, as can be seen from the contrast in (46a-b), involving the Hebrew verbal passive *ne'emar* 'was said':

- (46) a. *ma₁ [she-Dan oxel t₁] ne'emar li? (Hebrew)
 what that-Dan eats said.PASV to.1sg
 b. ma₁ ne'emar li [she-Dan oxel t₁]?
 what said.PASV to.1sg that-Dan eats
 'What was it said to me that Dan eats?'

The cases in (46a-b) differ in whether or not the subject has moved to TP. The subject (in both cases) receives the role of THEME ([-c -m]), which is a uniform [-] cluster. Given the current proposal, this means it is inserted at its base position via *set-merge*. Hence, extraction from it is possible precisely when it has not been moved (46b).

Unlike Hebrew, English requires subjects to move to TP overtly. However, there are exceptions to this: since clausal arguments are exempt from the Case requirements that apply to DP's, they can remain in-situ, with an expletive pronoun in subject po-

¹¹ The relationship between argument-mapping and accusative Case, besides its evident empirical adequacy, may seem rather arbitrary at this point. This is hardly so, however; see Preminger (2005).

sition instead. Thus, the English counterparts of (46) are possible, and in fact behave the same way:

- (47) a. *What₁ [that John eats t₁] was said to me?
 b. What₁ was it said to me [that John eats t₁]?

Even more striking is the case of *ne*-cliticization in Italian. As can be seen in (48), Italian does not require overt movement of the subject to TP, on par with Hebrew:

- (48) a. Arriveranno molti esperti b. Telefoneranno molti esperti
will-arrive many experts *will-telephone many experts*
 ‘Many experts will arrive.’ ‘Many experts will telephone.’
 (Italian) (data from Burzio 1986)

Belletti and Rizzi (1981), Burzio (1986), and Cinque (1990) discuss the syntactic behavior of the *ne* clitic, which replaces the nominal complement of a quantifier in Italian. As discussed by Cinque (1990), *ne*-cliticization is possible only from “structural object” positions, including direct (but not oblique) objects, and post-verbal subjects of passives, unaccusatives, and impersonal *si* constructions.

Notice the contrast between the applicability of *ne*-cliticization to the post-verbal subject of an unaccusative verb (49a), and its inapplicability to the post-verbal subject of an unergative verb (49b):

- (49) a. Ne arriveranno molti (Italian)
NE(of-them) will-arrive many
 ‘Many of them will arrive.’
 b. *Ne telefoneranno molti
NE(of-them) will-telephone many (data from Burzio 1986)

Given the current proposal, the subject of (49a) receives the role of THEME ([-c -m]), which is a uniform [-] cluster, and is therefore attached to its base position via *set-merge*. The subject of (49b) receives the role of AGENT ([+c +m]), which is not a uniform [-] cluster, and is therefore attached to its base position via *pair-merge*.

Assuming *ne*-cliticization involves movement out of the quantified noun-phrase, this contrast accounts for the contrast in islandhood shown in (49).

The facts in (46-49) show the dependency of Subject Condition effects on the thematic role of the argument in question, therefore providing support for the dual nature of the Subject Condition, as described in (43a-b). Under the current proposal, this dependency on thematic roles is reduced to an explainable dependency on the type of merger involved: if an argument is not *pair-merged*, it will show Subject Condition effects only in case it is an overt subject—in other words, moved to TP—since *Internal Merge* is always *pair-merge* (45).

Furthermore, this dependency on thematic roles is in no way unique to subjects, as shown by the case of EXPERIENCER arguments which block extraction despite being internal (see sections 2.2, 3.2).

Thus, the Subject Condition can be seen as a conflation of two separate phenomena: the islandhood of moved constituents, and the islandhood of constituents which have been *pair-merged* at their base positions.

3.6. A New View on the Set-Merge/Pair-Merge Distinction

The discussion in 3.3-3.5 does leave one question open —that of the apparently divergent behavior of adjuncts and other *pair-merged* constituents. Recall that *pair-merge* was invoked by Chomsky (2004) to account for the exceptional behavior of adjuncts with respect to Condition C:

- (50) a. *He_i bought the book [that John_i wanted].
 b. [Which book [that John_i wanted]]₁ did he_i buy [which book [that John_i wanted]]₁?
- (51) a. *He_i liked the picture [of John_i].
 b. *[Which picture [of John_i]]₁ did he_i like [which picture [of John_i]]₁?
 (strikeout indicates unpronounced copies)

The interesting case is (50b): one would have expected the copy of [*which book that John wanted*], which is generated as an argument of *buy*, to cause a Condition C violation.

The generalization is as follows: phrases contained in an adjoined element (50) can only induce a Condition C violation at their “surface” (i.e. phonologically pronounced) position (as in (50a)), while phrases contained in non-adjointed elements (51) induce Condition C violations at their base and intermediate positions as well.

This was handled by Chomsky (2004) by asserting that *pair-merge* attaches constituents on what is essentially a “separate plain”, while *TRANSFER*, the narrow-syntactic preparation for Spell-Out, “flattens” them into *set-merged* structures. Presumably, this means that *pair-merged* structures are entirely exempt from Condition C effects, while the *set-merged* structures that result from their *TRANSFER* lose this property.

However, given the current proposal, arguments receiving thematic roles which are not uniform [-] clusters are also merged via *pair-merge*. Yet these arguments exhibit no such bleeding effects. Consider the following example:

- (52) a. *[Which sister [of Bill_i]]₁ did he_i think [t₁ kissed John]?

Here, [*which sister of Bill*] is originally merged as the AGENT of *kissed*. Since AGENT ([+c +m]) is not a uniform [-] cluster, it is merged via *pair-merge*. The constituent subsequently undergoes wh-movement to the matrix clause, but crucially, this does not exempt the base position from incurring a Condition C violation.

The effects are therefore restricted to actual adjuncts (i.e. unselected modifiers). This appears to contradict the idea that *pair-merge* is the operation responsible for the merger of both types of constituents.

However, recent work by Fox (2002) and Fox and Nissenbaum (1999) has shown that there is a completely different way to account for the Condition C facts in (50-51). Their proposal involves late-merger of adjuncts to (overtly or covertly) moved constituents.

Under Fox and Nissenbaum’s approach, unselected modifiers (i.e. adjuncts) can merge to a constituent at any point in the derivation. Specifically, they can merge to a higher copy which is a result of *Internal Merge* (i.e. movement). Thus, the derivation of (50b), above, would proceed as follows (abstracting away from irrelevant details):

- (53) a. he buy [which book]₁ b. [which book]₁ [he buy [which book]₁]
 c. [which book [that John wanted]] [he buy [which book]₁]

Under this analysis, there is never an instance of *John* within the c-command domain of the pronoun *he*, and therefore no Condition C violation occurs. The same cannot apply to (51), because the relation between *picture* and [*of John*] is thematic:

- (54) a. he like [which picture]₁ b. [which picture]₁ [he like [which picture]₁]
 c. [which picture [of John]] [he like [which picture]₁]

For [*of John*] to merge to *picture*, the latter must be an event-nominal, which has a thematic role of THEME to assign. However, if this is the case, then the lower copy of [*which picture*], where *picture* has no argument, represents a violation of the Theta Criterion. Therefore (54a-c) represent a derivation that will crash at LF.¹²

Thus, unlike Chomsky's *TRANSFER* approach, the late-merger approach captures the difference between arguments and adjuncts in terms of their inherent semantic differences: the former are subject to the Theta Criterion at LF, while the latter are not (Fox 2002). This allows adjuncts to be absent from the lower copies of a moved constituent, while barring the same from applying to arguments. The immediate result is the restriction of late-merger to adjuncts. Hence, even though adjuncts are attached to the syntactic tree via *pair-merge*, it is no longer necessary to postulate that some property of *pair-merge* is what makes them (partially) resistant to Condition C.

As a result, the grammaticality of (52), repeated below, no longer poses a problem to the proposed system:

- (55) *[Which sister [of Bill]]₁ did he_i think [t₁ kissed John]?

The argument [*which sister of Bill*] is *pair-merged* at its base position, but it is most certainly not an adjunct. It is an argument of *kissed*; likewise, [*of Bill*] is an argument of *sister*. Thus, neither can be late-merged to their respective heads (since this would result in a violation of the Theta Criterion at LF). They are therefore merged to those heads at their respective base positions, giving rise to a Condition C violation with respect to the pronoun *he*.

Perhaps the most important consequence of adopting the late-merger proposal, with respect to the current proposal, is the fact that it reduces the difference between *pair-merge* and *set-merge* to one property alone: the blocking of extraction. Since *set-merge* and *pair-merge* are theoretical primitives, it is independently desirable that the difference between them would be one single primitive property, rather than having a collection of differing properties.

4. Conclusion

I started by presenting unexplained data regarding *external arguments*, alongside unresolved theoretical issues pertaining to the notion of argument externality. In addition, I presented the case of certain verbal arguments, which behave with respect

¹² Notice that this is independent of the issue of little-v. As pointed out in 2.1.1.2, I adopt Horvath and Siloni's (2002) arguments for the rejection of the Little-v Hypothesis, as the latter is incompatible with the data presented in 2.1.1 (see note 3). However, even if one adopts the Little-v Hypothesis, it does not provide a clear advantage in explaining the blocking of extraction from subjects, unless an additional stipulation is made regarding the properties of the vP projection in general, and its specifier in particular.

to extraction in a manner that defies explanation in terms of the external vs. internal argument distinction, or the specifier vs. complement distinction.

Having taken these problems as a starting point, I showed that a surprising generalization emerges from the combination of argument-mapping and extraction from verbal arguments. From this generalization I then derived a system which links the type of merger used for a given argument to the thematic role assigned to that argument.

This system is able to account for the previously unexplained data, while also providing answers to the theoretical questions presented at the onset.

Finally, I addressed the issue of extraction from displaced arguments, proposing to subsume Wexler and Culicover's (1977, 1980) Freezing Principle under the restriction of *Internal Merge* to *pair-merge* (given the Copy Theory of Movement; Chomsky 1995b, 2004).

Given the proposal of Fox (2002) and Fox and Nissenbaum (1999), regarding a late-merger account for the bleeding of Condition C effects by adjuncts, the proposed system reduces the difference between the primitive notions *set-merge* and *pair-merge* to extractability alone. In addition to this desirable simplification, and the ability to account for previously unexplained data, the proposed system both clarifies the notion *external argument* and improves our understanding of the conditions on extraction in the verbal domain.

References

- Belikova, A. and Preminger, O., 2004, «Lexical causatives and argument mapping». Ms. Tel-Aviv University.
- Belletti, A. and Rizzi, L., 1981, «The syntax of *ne*: Some theoretical implications», *The Linguistic Review* 1: 117-145.
- Borer, H. and Grodzinsky, Y., 1986, «Syntactic vs. lexical cliticization: The case of Hebrew dative clitics». In *The syntax of pronominal clitics*, ed. H. Borer. San Francisco: Academic Press.
- , 1994, «The projection of arguments». In *Functional projections*, eds. E. Benedicto and J. Runner, University of Massachusetts Occasional Papers in Linguistics 17, 19-47. Amherst, MA: GLSA.
- Bowers, J., 1993, «The syntax of predication», *LI* 24:591-656.
- Burzio, L., 1986, *Italian syntax: A government-binding approach*: "Studies in natural language and linguistic theory". Dordrecht: Reidel.
- Chierchia, G., 1989, «A Semantics for unaccusatives and its syntactic consequences». In *The unaccusativity puzzle*, eds. A. Alexiadou, E. Anagnostopoulou and M. Everaert. Oxford: Oxford U. P.
- Chomsky, N., 1986, *Barriers*. Cambridge, MA: MIT Press.
- 1995a, «Bare phrase structure». In *Government and binding theory and the minimalist program*, ed. G. Webelhuth, 383-439. Oxford: Blackwell.
- 1995b, *The minimalist program*. Cambridge, MA: MIT Press.
- 2001, «Derivation by phase». In *Ken Hale: A life in language*, ed. M. Kenstowicz, 1-52. Cambridge, MA: MIT Press.
- 2004, «Beyond explanatory adequacy». In *Structures and beyond: The cartography of syntactic structures, Vol. 3*, ed. A. Belletti, 104-131. Oxford: Oxford U. P.

- 2005, «On Phases». Ms. MIT.
- Cinque, G., 1990, «Ergative adjectives and the lexicalist hypothesis», *NLLT* 8: 1-39.
- Collins, C., 1997, *Local economy*. Cambridge, MA: MIT Press.
- Fox, D., 2002, «Antecedent-contained deletion and the copy theory of movement», *LI* 33: 63-96.
- and Nissenbaum, J., 1999, «Extraposition and scope: a case for overt QR». In *Proceedings of 18th West Coast Conference on Formal Linguistics*, eds. S. Bird, A. Carnie, J. D. Haugen and P. Norquest, 132-144. Somerville, MA: Cascadilla Press.
- Heim, I. and Kratzer, A., 1998, *Semantics in generative grammar*. Oxford: Blackwell.
- Horvath, J. and Siloni, T., 2002, «Against the Little-v Hypothesis», *Rivista di Grammatica Generativa* 27: 107-122.
- Huang, Cheng-Teh J., 1982, *Logical relations in Chinese and the theory of grammar*, MIT: Ph.D. dissertation.
- Johnson, K., 1992, «Scope and binding theory: Comments on Zubizarreta». In *Syntax and Semantics, 26: Syntax and the lexicon*, eds. T. Stowell and W. Eric, 259-275. San Diego: Academic Press Inc.
- Kayne, R. S., 1984, *Connectedness and binary branching*. Dordrecht: Foris.
- 1994, *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kratzer, A., 1996, «Severing the external argument from its verb». In *Phrase structure and the lexicon*, eds. J. Rooryck and L. Zaring, 109-137. Dordrecht: Kluwer.
- Landau, I., 1994, *Dative shift and extended VP-shells*, Linguistics Department, Tel-Aviv University: M.A. thesis.
- 2001, *The locative syntax of experiencers*. Ms. Ben-Gurion University.
- Larson, R. K. 1988, «On the double object construction», *LI* 21: 589-632.
- Levin, B. and Rappaport-Hovav, M., 1994, «A preliminary analysis of causative verbs in English», *Lingua* 92: 35-77.
- 1995, *Unaccusativity*. Cambridge, MA: MIT Press.
- Pesetsky, D., 1995, *Zero syntax: Experiencers and cascades*. Cambridge, MA: MIT Press.
- Preminger, O., 2005, *Argument-Mapping and Extraction*, Linguistics Department, Tel-Aviv University: Master's thesis.
- Reinhart, T., 2000, «The theta system: Syntactic realization of verbal concepts», *OTS Working Papers in Linguistics* 00, 01/TL.
- 2001, «Experiencing derivations». In *Proceedings of Semantics and Linguistic Theory* 11, eds. R. Hastings, B. Jackson and Z. Zvolenszky. Ithaca, NY: CLC Publications.
- 2002, «The theta system: An overview», *Theoretical Linguistics* 28: 229-290.
- and Siloni, T., 2005, «The Lexicon-Syntax parameter: Reflexivization and other arity operations», *LI* 36: 389-436.
- Shlonsky, U., 1987, *Null and displaced subjects*, MIT: Ph.D. dissertation.
- Van Hout, A., 1995, *Event semantics of verb frame alternations: A case study of Dutch and its acquisition*, University of Tilburg: Ph.D. dissertation.
- Van Valin, R. D., 1990, «Semantic parameters of split intransitivity», *Lg* 66: 221-260.
- Wexler, K. and Culicover, P. W., 1977, «Some syntactic implications of a theory of language learnability». In *Formal Syntax*, eds. P. W. Culicover, T. Wasow and A. Akmajian, 7-60.
- 1980, *Formal principles of language acquisition*. Cambridge, MA: MIT Press.
- Williams, E., 1980, «Predication», *LI* 11: 203-238.

PRENOMINAL AND POSTNOMINAL DEMONSTRATIVES IN SPANISH: A [\pm DEICTIC] APPROACH

Inma Taboada
Georgetown University

Abstract¹

In this paper I propose an explanation for the difference in meaning and in structure between the prenominal and the postnominal demonstratives in Spanish, and its complimentary distribution with the article in prenominal position. The literature had traditionally considered that there was a difference only in structure, although Bernstein (1997) points out that the postnominal one does not have a deictic meaning. I propose that it is the feature [+deictic] which triggers the raising of the demonstrative to D^o, in order to check the [+Ref] feature present in this position. The postnominal demonstrative is marked [-deictic], what prevents it from moving, and forces the appearance of the expletive article in D^o.

0. Introduction

As traditional grammars (R.A.E. 1973) claim, we do not know the meaning of *este* 'this' or *ese* 'that', their semantic meaning is vacuous; but, what we know is that they are used to point at something, either physically, this is deictically, it points out something we see or we remember, or anaphorically/cataphorically, it makes reference to something that has already been mentioned/or is going to be mentioned in the discourse. This difference in the structure is due to the presence of the feature [+deictic] in the prenominal one. In section 1, I try to determine the categorial status of the demonstrative, since it cannot be an article, nor an adjective; in section 2, I explore the analyses other authors have proposed to explain the different structures we can find; in section 3, I present my analysis, where I propose the existence of the feature [\pm deictic], the trigger of the demonstrative movement in Spanish.

1. The Categorial Status of the Demonstrative

Demonstratives have traditionally been considered to share several properties with articles, if not to be the same type of element. Roca (1996b) mentions some of these common properties:

¹ I wish to express my gratitude to Héctor Campos for discussing the previous versions of this paper, and for all his support all these years. The remaining mistakes are all my own.

- (i) prenominal position
- (ii) complementary distribution always present in English, and in some Romance languages in prenominal position
- (iii) the definiteness value
- (iv) both can appear in argument position (Longobardi 1994)
- (v) coincidences in their semantic representation according to Diesing (1992), who makes a difference between indefinites and the other determiners.

Nevertheless, in recent years these similarities have been left aside, since it is important to make clear that determiners and demonstratives cannot be the same type of animal. Bernstein (1997) gives three reasons for this. Her first reason is that in some languages they can cooccur, as we can see in Spanish in example (6) and in Rumanian in example (9):

- (1) *The* students went to the bar
- (2) *These* students went to the bar
- (3) **The these* students went to the bar
- (4) *Los* estudiantes fueron al bar (Spanish)
- (5) *Esos* estudiantes fueron al bar
- (6) *Los* estudiantes *esos* fueron al bar
- (7) *Băiat-ul* frumos (Rumanian)
boy-the nice
- (8) *Acest* (frumos) *băiat* (frumos)
this (nice) boy (nice)
- (9) *Băiat-ul acesta* frumos
boy-the this nice

Bernstein (1997) gives examples of Hungarian and Javanese, languages where these elements can cooccur even in prenominal position, as we can see in examples (10-11):

- (10) *ez a* haz (Hungarian)
this the house
- (11) *ika n* anak (Javanese)
this the child

The second reason Bernstein (1997) proposes is that the demonstrative may stay alone, becoming a pronoun, but not the definite article, although the indefinite one can stay alone in some languages, as we can see in examples (12-14):

- (12) This is the one I want
- (13) *Éste* es el que yo quiero (Spanish)
this is the that I want
- (14) *Este* e o que eu quero (Galician)

And Bernstein's last reason is that in many languages the demonstrative is adjectival in nature, exhibiting a full range of adjectival inflection and often occupying the position typical of adjectives. Nevertheless, there is a lot of discussion about the categorial status of the demonstrative. Panagiotidis (2000) claims that the categorial status of the demonstrative cannot be the same as that of the adjective. He finds diachronic and synchronic evidence. As Greenberg (1978, 1991) had already stated, the shift from demonstrative to article is well attested in quite a few languages: demonstrative > definite article > non generic article > noun marker. According to Panagiotidis (2000), if demonstratives were adjectives or adjective-like elements, perhaps possessives could have undergone the same categorial shift into Determiner heads. Synchronically, if, as Brugè (1996) claims, demonstratives carry an interpretable [Referential] feature, they cannot possibly be adjectives; therefore, Panagiotidis assumes that demonstratives are DPs cross-linguistically, as demonstrative heads are Determiners.

In the literature, demonstratives had usually been considered to be generated in a specifier position, [Spec, AgrP] for Giusti (1997, 2002), being this a high position above the Functional Projections containing the Adjectival Phrases; [Spec, AgrP] for Brugè (1996, 2002), being this a low FP, right above the NP. However, Roca (1996a) proposes that the demonstrative is a head projecting its own phrase, the Demonstrative Phrase (DemP), at least in the case of Spanish, although we might find some languages in which it can be a specifier. Roca follows Cornilescu's (1992) work, who in her study of the different determiners in Rumanian concludes that there must exist a DetP below the DP, headed by a definite article; thus, if the D° position is occupied by the definite article, Det° might be filled by different determiners, such as demonstratives, indefinite, or quantifiers.

2. Prenominal vs. Postnominal Demonstratives

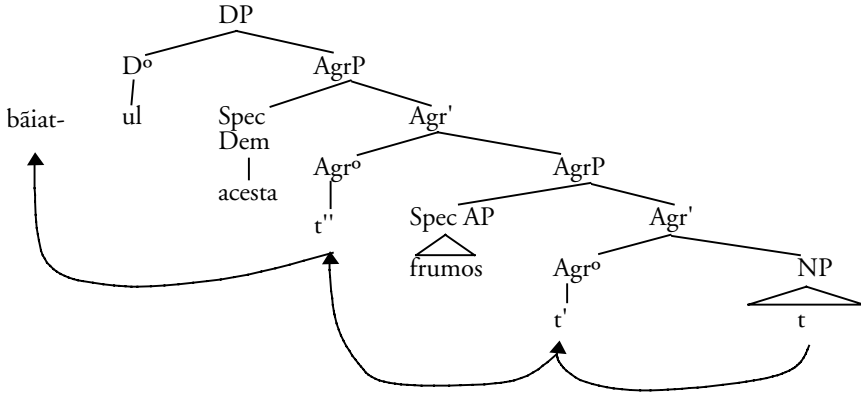
As we have seen in examples (6) and (9), in both Spanish and Rumanian, demonstratives can cooccur with a definite article which show that they cannot occupy the same position.

- (15) a. băiat-ul acest frumos b. acest (frumos) băiat (frumos)
 boy-the this nice this (nice) boy (nice)

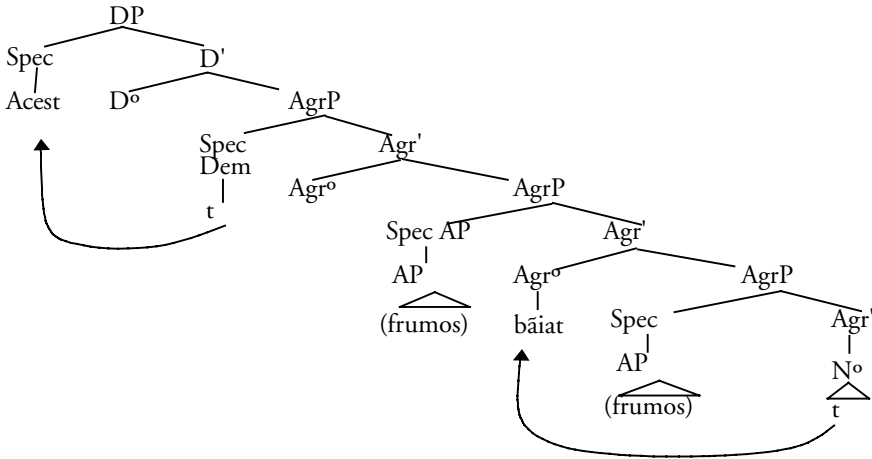
Giusti (1997) proposes two different structures for DPs, depending if they are headed by an article, or by a demonstrative, as we can see in (16) and (17):

As we can see in (16), N-movement crosses over both the demonstrative and the adjective, and once the noun has raised to D° and the article is present, the demonstrative does not need to move. However, in (17) we see that the demonstrative raises to [Spec, DP], and once this position is filled by an element able to check the [Ref] feature present in D°, there is no need for the article to appear.

(16) Headed by a determiner:



(17) Headed by a demonstrative:



2.1. Complementary distribution in prenominal position

A question widely discussed in the literature is why in some languages an element in [Spec, DP], the demonstrative is in complementary distribution with an article in D°. Both Giusti (1997, 2002) and Brugè (1996, 2002) point out that this complementary distribution is similar to the doubly-filled COMP filter' (Chomsky and Lasnik 1977), which is not universal. As Giusti (1997, 2002) and Brugè (1996, 2002) point out, two assumptions must be taken into account at this point. First, a functional projection is instantiated in order to realize some feature ϕ , and this feature must be visible in order to be properly interpreted at LF; and second, the relevant relation

for the satisfaction of the visibility condition imposed on functional features is universally the Spec-Head relation: if the specifier position is empty or does not have strong specification for the relevant feature, the head must be filled. Otherwise, the head can be abstract. Variation across languages can be reduced to variation across (inflectional) morphological systems. If an element in Spec makes the relevant features (morphologically) visible, the corresponding head in agreement with it will be empty. If the relevant features are not morphologically visible, or if the specifier position is empty, the relevant head must be inserted in order for the projection to be properly interpreted at LF. Languages vary with respect to the level at which the demonstrative moves to [Spec, DP], its final destination.

However, Bernstein (1997) proposes what I consider to be a more elegant solution to the question, not based on a stipulation, as the Doubly-Filled DP Filter was. Bernstein assumes that the X^0 corresponding to the demonstrative head in [Spec, AgrP] raises and substitutes into the D^0 position. This claim automatically accounts for the absence of a prenominal demonstrative cooccurring with a definite article in Romance and Germanic languages.

2.2. Different analyses

Roca (1996a; 1996b) claims that the Spanish demonstrative is a functional head projecting an XP (his DemP) situated below the DP. Nonetheless, the demonstrative is not syntactically homogenous crosslinguistically. We can find different base positions in different languages.

Both Giusti (2002) and Campos (2005) propose that the appearance of the article in prenominal position in languages such as Rumanian or Arvantovlaxica are last resort operations. Giusti (2002) claims that in Rumanian the determiner bears a set of ϕ -features, which includes Case, Gender and Number, the strong features, which must be checked (Chomsky 1995b, 1998, 1999). They can be checked in two different ways, moving an element to [Spec, DP] or, by default, by spelling out the definite article, which can be considered the morphological realization of these ϕ -features:

- (18) a. [_{DP} [+ ϕ F*] [_{FP} frumos F [_{NP} băiat]]]
 b. [_{DP} [+ ϕ F*] [_{FP} frumos băiat F [_{NP} băiat]]]
 c. [_{DP} băiat-ul [_{FP} frumos băiat F [_{NP} băiat]]]

In example (18) we see that the adjective is generated in the specifier position of an intermediate functional category FP, between the DP and the NP. The Noun raises to check agreement features on the adjective in (18b). However, since nothing has checked [+ ϕ F*], the definite article must spell-out. Since the article is an enclitic element, the noun has to raise for phonological reasons (Dobrovie-Sorin 1987).

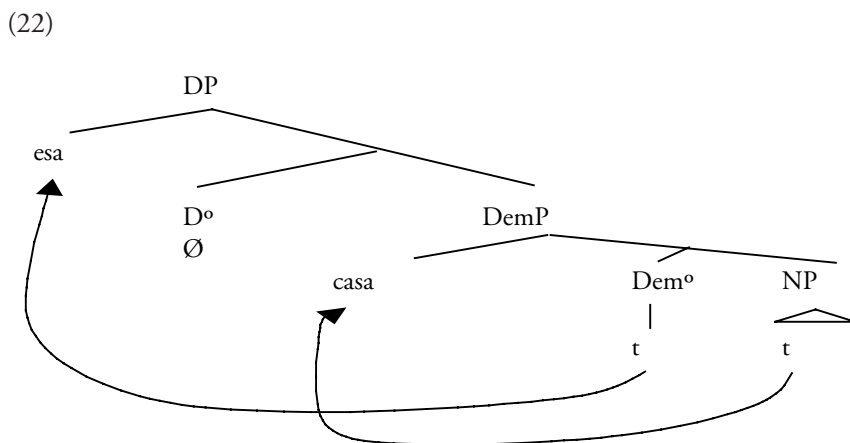
As we have already said, in Spanish and Rumanian we can find two different positions for the demonstrative. However, we find that in Spanish there is a clear difference in meaning: the prenominal demonstrative possesses two different features [+ deictic, \pm anaphoric], while the postnominal one is [- deictic, + anaphoric], as we can see in examples (19) and (20):

- (19) —¿Qué has leído?
 what have-you read
 —Este libro (while pointing at it)
 this book
 —*El libro este (while pointing at it)
 the book this
- (20) ¿dónde está la casa esa de la que hablas?
 where is the house that of the that you-talk
 ‘Where’s the house you’re talking about?’

As we can see in example (19), the postnominal use of the demonstrative is banned if we are physically pointing at the object we are talking about. As Bernstein (1997) points out, the postnominal construction in Spanish cannot be deictic, since we always need an adverb in order to express the deictic meaning:

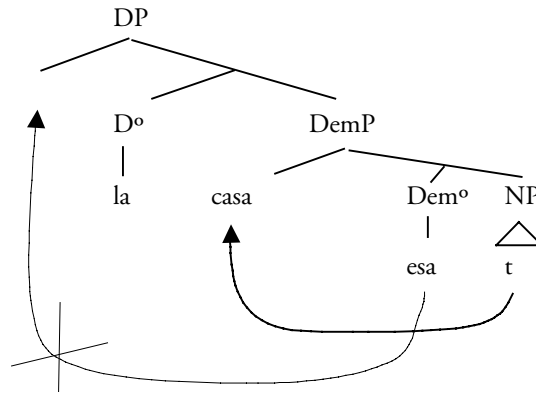
- (21) —¿Qué has leído? —El libro ese de ahí
 what have-you read the book that of there
 ‘What have you read?’ ‘That book there’

Roca (1996) proposes the same structure for both the prenominal and the postnominal order in (22) and (23):



As we can see in (22), we obtain the prenominal order by moving the noun *casa* ‘house’ to [Spec, DemP], and Dem° to [Spec, DP], while we find in (23) that the movement of the Dem° to [Spec, DP] is blocked because D° is filled by the article. Therefore, according to Roca (1996b) the prenominal order crucially depends on the movement of Dem° to D°.

(23)

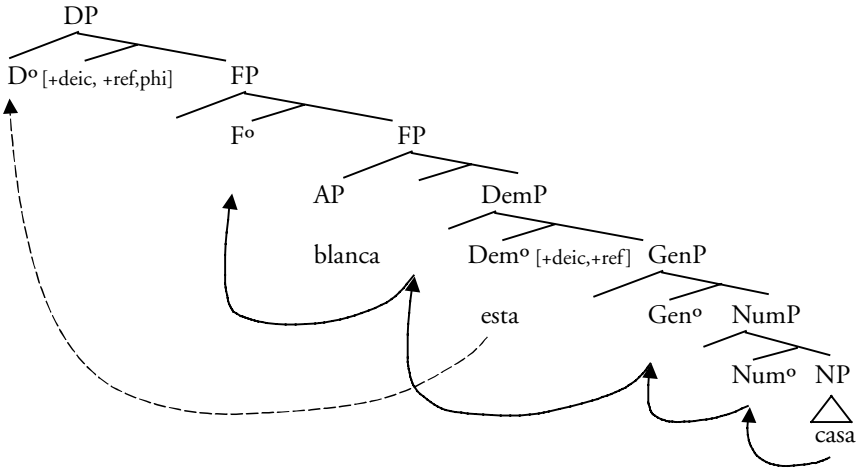


Brugè (1996) determines that the demonstrative must always raise to [Spec, DP] at some point through the derivation, optionally before Spell-Out, but obligatorily at LF. For her, demonstratives generate in a unique position in all languages, but they differ as to their power to allow, oblige or prevent the movement of the demonstrative to [Spec, DP] before Spell-Out. A good question at this point could be why do we need the article if the demonstrative always moves to [Spec, DP] at LF? Brugè says that if the demonstrative does not move to [Spec, DP] before Spell-Out, the definite article must be realized in D° in order to show at PF that this position contains some feature [+Ref] which prevents the DP from being interpreted as existential.

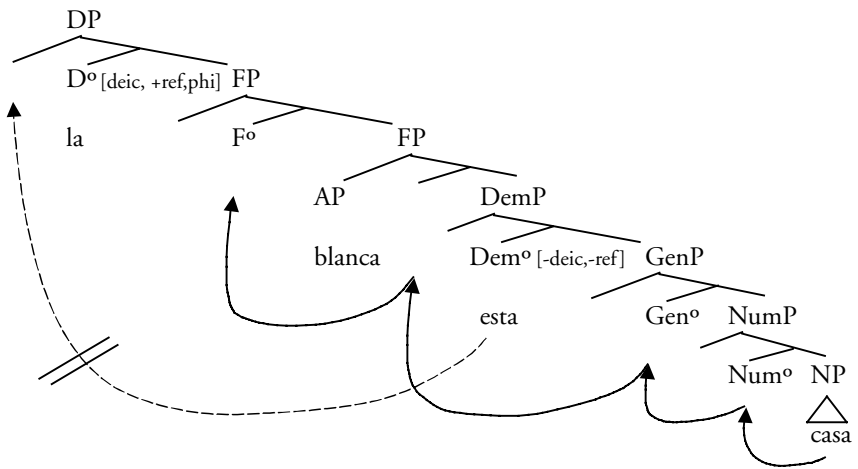
3. My proposal

Nonetheless, and according to the minimalist framework and Campos's (2005) analysis for AV, we can determine that, in Spanish, the prenominal demonstrative, the [+ deictic] one, can check the strong [+φ*] features in D°, as well as its [+ referential] feature, and for that reason the demonstrative has to raise to that position, to check all the features present in D°. These two features, [deictic] and [Referential] must be somehow related, since we see that only the [+ deictic] demonstrative, this is the prenominal one, can check this [Ref] feature in D°. On the other hand, the postnominal demonstrative, the [- deictic] one, cannot check the [+ Referential] feature in D°; hence, the expletive article must appear in a last resort operation, just to check the strong [+φ*] features, and the [Ref] feature of D°. The resulting structures we find are the one in (24) for prenominal demonstratives, and the one in (25) for postnominal ones.

(24) Prenominal:



(25) Postnominal:



In both examples we see that the noun moves head to head through all the FPs present in the structure in order to check its own ϕ -features, and the ones belonging to the adjective, as shown by (Cinque 1994). As we have seen, in (24), the prenominal demonstrative has the feature [+deictic], therefore, it can check the [Ref] feature in D° by entering into a Spec-head relation. However, in (25) we can see that the demonstrative does not have a deictic feature, therefore it cannot check the deictic/referential feature in the DP, triggering the appearance of the expletive article, in order to check the referential feature of the DP, and the ϕ -features of the whole DP.

4. Other constructions that support this analysis

Two other constructions that show this [\pm deictic] approach may be right, are the postnominal demonstrative that does not require an article, and the postnominal structure with the place adverb. Let us focus first in the former one. We can find examples such as the following one:

- (26) Bush ha decidido atacar otro país, *decisión esta* aplaudida por Blair
 Bush has decided attack another country decision this applauded by Blair
 ‘Bush has decided to attack another country, and this decision has been applauded by Blair’

This example does not pose a problem, since, following Longobardi (1994), bare NPs can appear in non-argument positions, and the structure N° -demonstrative can never appear in argument position, they are always appositions; the feature [Ref] is not present in D° , thus, although the [$-$ deictic] demonstrative cannot check it, the expletive article does not need to be present.

The second construction, the one with the place adverb present can also be explained by this approach. Bernstein (1997) claims that *ese de ahí* ‘that of there’ must form a constituent. And this can explain the fact that the demonstrative always has to agree with the adverb, as we can see in examples (27) and (28):

- (27) *esta de aquí/esa de ahí/aquella de allí* (28) **esta de ahí/ *aquella de aquí*
 this of here/that of there/that of there

Then, if we have examples (29) and (30):

- (29) *la casa esa de ahí* (30) *esa casa de ahí*
 the house that of there that house of there

either the article or the demonstrative must appear in prenominal position, since the [Ref] feature must be checked. Since it is the adverb the one that possesses the [$+$ deictic] feature, and the demonstrative has inherited it, we can say that this feature percolates to the whole $DemP$, and now we can choose the mechanism to check the [Ref] of DP: either the demonstrative can raise, or the expletive article can appear.

5. Conclusion

As we have seen, the presence of the [$+$ deictic] feature in Dem° , triggers the movement of the demonstrative to D° , in order to check the [$+$ Ref] feature in this position. These two features, [\pm deictic] and [\pm Referential], must be somehow related, since the presence of the [\pm deictic] one can check the [$+$ Ref] one, and it is decisive for the appearance of the expletive article or the movement of the demonstrative, and we have also seen that this analysis is able to explain other constructions containing a demonstrative.

References

- Bernstein, J., 1997, «Demonstratives and Reinforcers in Romance and Germanic Languages», *Lingua* 102: 87-113.
- Brugè, L., 1996, «Demonstrative Movement in Spanish: A Comparative Approach», *University of Venice Working Papers in Linguistics* 6: 1-53.
- , 2002, «The Position of Demonstratives in the Extended Nominal Projection». *Functional Structure in DP and IP: The Cartography of Syntactic Structures*. G. Cinque. Oxford, Oxford U. P.: 15-53.
- Campos, H., 2005, «Noun Modification, Pseudo-Articles, and Last Resort Operations in Arvantovlaxika and in Romanian», *Lingua* 115(3): 311-347.
- Chomsky, N., 1995, *The Minimalist Program*. Cambridge, MIT Press.
- 1998, «Minimalist Inquiries: The Framework», *MIT Occasional Papers in Linguistics*(15): 1-56. [Reprinted 2000 in R. Martin, D. Michaels and J. Uriagereka (Eds.), *Step by Step: Minimalist Essays in Honor of Howard Lasnik*, 89-155. MIT Press, Cambridge, Mass.].
- 1999, «Derivation by Phase». *MIT Occasional Papers in Linguistics*. Cambridge, MIT Press. 18: [Revised version appeared as Chomsky (2001).].
- and H. Lasnik, 1977, «Filters and Control», *LI* 8: 425-504.
- Cinque, G., 1994, «On Evidence for Partial N-Movement in the Romance DP». *Paths Towards Universal Grammar*. J. K. G. Cinque, J.-Y. Pollock, L. Rizzi and R. Zanuttini. Washington, DC, Georgetown U. P.: 85-110.
- Cornilescu, A., 1992, «Remarks on the Determiner System of Rumanian: the Demonstratives *al* and *cel*», *Probus* 4: 189-260.
- Diesing, M., 1992, «Bare Plural Subjects and the Derivation of Logical Representations», *LI* 23(3): 353-380.
- Dobrovie-Sorin, C., 1987, «A propos de la structure du groupe nominal en Roumain», *Rivista de Grammatica Generativa* 12: 123-152.
- Giusti, G., 1997, «The Categorical Status of Determiners». *The New Comparative Syntax*. L. Haegeman. London, Longman: 95-123.
- 2002, «The Functional Structure of Noun Phrases: A Bare Phrase Structure Approach». *Functional Structure in DP and IP: The Cartography of Syntactic Structures*. G. Cinque. Oxford, Oxford U. P.: 54-90.
- Greenberg, J. H., 1978, «How does a Language Acquire Gender Markers?» *Universals of Human Language*. C. F. J. Greenberg, E. Moravcsik. Stanford KY, Stanford U. P. 3: *Word Structure*: 47-82.
- 1991, «The Last Stages of Grammatical Elements: Contractive and Expansive Desemantization». *Approaches to Grammaticalization*. B. H. E. Traugott. Amsterdam/Philadelphia, John Benjamins Publishing House. 2: 301-314.
- Longobardi, G., 1994, «Reference and Proper Names: A Theory of N-Movement in Syntax and Logical Form», *LI* 25(4): 609-665.
- Panagiotidis, P., 2000, «Demonstrative Determiners and Operators: The Case of Greek», *Lingua* 110: 717-82.
- R.A.E., 1973, *Esbozo de una nueva gramática de la lengua española*. Madrid, Espasa-Calpe.
- Roca, F., 1996, *La determinación y la modificación nominal en español*. Universidad Central de Barcelona. Barcelona.
- 1996a, «Artículo versus demostrativo en español». VI Colloquium on Generative Grammar, Valencia.
- 1996b, *La determinación y la modificación nominal en español*. Universidad Central de Barcelona. Barcelona.

A LOOK AT SECOND LANGUAGE LEARNERS' TASK MOTIVATION

Íñigo A. Yanguas

Georgetown University

Abstract

In this study, part of a larger project, I explore a) the possible relationship between task motivation as operationalized in Dörnyei's Process model of motivation and linguistic variables in a written production task, and b) the differences in performance between two task motivation groups. This model has proven valid with oral argumentative tasks and now its main tenets are tested using a semi-guided writing activity. 65 students of Spanish at Georgetown University took part in this project. In this preliminary analysis, correlations between task motivation and linguistic variables were carried out to ascertain the possible relationships. Also, Independent samples *t*-test analysis served as a tool to establish possible significant differences among two task motivation groups. Results show that task motivation indeed holds a significant positive linear relationship with the linguistic variables investigated. Results also show the high motivation group significantly outperforms the low motivation group.

1. Introduction

The purpose of the present study is to assess the impact of task motivation on the quality and quantity of second language (L2) writing produced by college-level foreign language students in a semi-guided writing task. Given that Dörnyei's Process Model of motivation (Dörnyei 2000, Dörnyei and Ottó 1998) has been successfully used to account for L2 performance in the context of oral argumentative tasks, the next logical step is to test the model with tasks that promote the learning of other L2 skills, in this case, written skills.

To this end, a semi-guided L2 writing task was designed and the Process model was followed. In the manner of Dörnyei and associates (Dörnyei 2002, Dörnyei and Kormos 2000, Kormos and Dörnyei 2004) a number of motivational variables were tapped into through a questionnaire. As far as syntactic complexity measures are concerned, five variables were employed adapting the ones used in the oral argumentative tasks. L2 writing literature is rich in studies using syntactic complexity measures in very different contexts, which often makes it difficult to draw general conclusions. In their research syntheses, Wolfe-Quintero et al (1998) and Ortega (2003) successfully attempted to uncover the commonalities and interpret results across studies. Hence, these two studies serve as a source for the measures utilized in this study.

In this study I report results of correlation and independent samples *t*-test analyses for the main motivational variable in the model, task motivation, which lies at the heart of this Process model. This is a situated model of motivation in which tasks are the unit on which the whole theoretical framework is based. In my view, if results for task motivation are not significant, the foundations of the model are greatly affected; therefore, the preliminary analysis of task motivation by itself is necessary. If need be, further analysis of the data gathered for this study will reveal the relationships between the other motivational variables tapped into and the linguistic variables as well as differences in performance by the two different proficiency groups present in the pool of participants.

2. Motivation and second language acquisition (SLA)

The impact of motivation on the second or foreign language (Oxford 1996) acquisition process has been the subject of investigation for almost forty years. Despite clear differences in their approach as well as in their operationalization of the construct (for reviews, see Dörnyei 1998, 2001, Gardner & MacIntyre 1993), results of these studies have consistently shown a strong correlation between motivation and language learning success.

However, it has been argued that this research has remained isolated from conventional applied linguistics research due to the macro perspective it tends to adopt (Kormos and Dörnyei 2004).

Dörnyei & Ottó (1998), as part of a research project that aimed at motivating foreign language learners in the classroom, found that the models proposed until then lacked what they considered paramount to the investigation of motivation in language learning. First, they lacked a sufficiently detailed description of all the motivational influences in the classroom. Second, they focused on the how and the why of certain courses of action “while playing down the importance of motivational sources of executing goal-directed behaviour” (Dörnyei and Ottó 1998: 43). Finally, they did not consider motivation in its temporal dimension, which is so important in the Process Model of motivation. Here, motivation is seen as ever changing, dynamically evolving towards the completion of some goal.

To contribute to the investigation of these aspects in L2 motivation, Dörnyei's Process Model of motivation (Dörnyei 2000; Dörnyei and Ottó 1998) emerges, based on the “Action Control Theory”, from mainstream motivational literature. It is an attempt to capture more specific aspects of the learning situation, in which the tasks involved have special relevance, as well as to look at the “dynamic motivational processes that take place during task completion” (Dörnyei 2002: 139) This is no place to describe their Action Theory in detail (see Dörnyei and Ottó 1998 for summary of main tenets); it will suffice to say that the Process Model of motivation has its antecedents in the work of mainstream educational motivation scholars (Heckhausen 1991, Kuhl 1984). These authors make a difference between two kinds of processing: predecisional and postdecisional. The former makes reference to the cognitive processing involved in the setting of goals, whereas the latter involves those cognitive activities subsequent to goal setting. In this manner, those activities at the predecisional stage are motivational in nature while those at the postdecisional stage

are volitional (Pintrich and Schunk 1996). These two stages suggest a temporal continuum that allows for the sequences of events to be separated (Heckhausen 1991).

This kind of approach, therefore, looks for the changes in the motivational continuum by considering the different stages in the motivational process in ongoing social activities such as classroom learning (Dörnyei 2002).

The idea behind the model is that, first, motivation has to be generated (choice motivation); second, the generated motivation has to be actively maintained and protected while the action lasts (executive motivation); finally, in the phase following the completion of the action (motivational retrospection), retrospective evaluation of the action must be carried out (Dörnyei 2003).

Research using this Process Model of L2 motivation does not abound, since it is very recent; however, in a study carried out by MacIntyre and associates (MacIntyre et al., 2001) they test for overlap between motivation concepts coming from different models, one of which is Action Control theory proposed by Kuhl and associates. MacIntyre et al. (2001) run factor analyses to conclude that the factor Action Motivation can be separated as an independent factor, which would validate the Process Model of motivation.

Tasks are the basic unit of analysis on which this model is based. Not only is it a logical step, given the shift from a more general perspective on L2 motivation to a more situated, classroom based approach, but also it is an important link between the study of L2 motivation and instructed SLA where tasks have been analyzed both from a language processing perspective and from a methodological perspective (Dörnyei 2002).

Tasks were first emphasized in L2 motivation by Ushioda (1996) and Julkunen (2001) who also investigated the relationship between learning tasks and motivation. However, their approach was much more static because they failed to account for the fact that complex learning behaviors could last for a considerably long period of time (Kormos and Dörnyei 2004). These authors did not take into consideration the different phases involved in the motivational processes around the realization of learning tasks.

The next step is therefore to define *tasks* in the context of the Process Model of motivation: "tasks can be seen as primary instructional variables or building blocks of classroom learning" (Dörnyei 2002: 137). Dörnyei (2002) settles the matter: the importance of the tasks being well delimited is stressed, that is, being able to determine where the task starts and where it ends needs to be clear, since tasks, as conceived here, are "discrete units of situated learning behaviors" (Dörnyei 2002: 139).

In this fashion, Dörnyei & Kormos (2000), Dörnyei (2002) and Kormos & Dörnyei (2004) investigated the effects of motivation in oral argumentative tasks. They all use data from a British-Hungarian research project in which "the research objective was to examine how motivational factors affect the quality and the quantity of student performance in an L2 communicative task performed in dyads" (Kormos and Dörnyei 2004: 4). 44 Hungarian students (aged 16-17) learning English as a foreign language participated in this research project. In the first study, Dörnyei & Kormos (2000) look into the effects of some socio-affective variables on the way foreign language (L2) learners' engage in oral argumentative tasks. These variables included several aspects of L2 motivation and some factors that characterized the groups the

participants belonged to (such as group cohesiveness and intermember relations), as well as the learners' L2 proficiency and 'willingness to communicate' in their L1. As dependent variables, different measures of L2 output in two argumentative tasks were included. The results evidenced that it was a combination and interaction of variables that could be used to predict task performance. For example, linguistic confidence only affected task engagement among students with a positive attitude toward the task, whereas social factors affected task engagement in different ways depending on the task attitudes. This fact made the authors conclude that "task attitudes appear to function like a filter: if they are positive then the learner's performance follows 'regular' patterns.....however, if the filter is 'up', that is, if students assume negative attitudes towards the particular task examined, their performance becomes somewhat random" (Dörnyei and Kormos 2000: 295-96). Furthermore, when the language task was changed to their L1, the motivational pattern was different.

In the second study, Dörnyei (2002) investigates the relationship between motivational variables and the number of words and turns used in oral argumentative tasks. Correlation analyses were carried out to show that when "the relationship between motivation and concrete learning behavioral measures is assessed we can obtain considerably higher correlations than when motivation is related to global achievement measures" (Dörnyei 2002: 155). This would support the use of concrete learning tasks and the process oriented approach that accounts for fluctuations in motivation depending on several factors surrounding the concrete learning action taking place. In the third study, Kormos & Dörnyei (2004) not only investigated speech quantity, but they also included other linguistic measures such as number of words, number of turns, accuracy, complexity, lexical richness, number of arguments and counter-arguments. They run correlation analyses between these and the following motivational variables included in a questionnaire: Integrativeness, Incentive values of English proficiency, Attitudes towards the English course, Linguistic confidence, Language use anxiety, Task attitudes and Willingness to communicate (WTC). Given the large amount of variables involved, the results were complex: first, there were significant positive correlations between the quantity of speech and Course attitudes and Task attitudes, also between speech confidence and speech size. All in all, motivational variables explain a little more than one third of the variance (35-37%) in the quantity of speech produced. WTC was significantly correlated to the number of turns produced and Accuracy to Course Attitudes. Overall, motivational variables explain a low percentage (9-16%) of the variance in accuracy, complexity and lexical richness and a little less than one third of the variance (30%) in the number of turns produced.

Based on the findings reported in Dörnyei (2002), these authors expected to find stronger correlations between motivation and actual learning behavior. Kormos & Dörnyei (2004) explain this fact by looking at the possible intervening variables such as the participants' level of proficiency or the "diversity of students' attitude to the task that influenced their behavior to a considerable extent" (Kormos and Dörnyei 2004: 10). As in Dörnyei & Kormos (2000), attitudes towards the task seemed to be a crucial factor in predicting further behavior. Participants were then divided into 'high-task attitude' (the upper half of the sample) and 'low-task attitude' (the lower half of the sample) to run the same correlation analyses reported above in the two

samples separately. The authors found “high correlations between the composite of motivational variables and complexity and the number of arguments produced in the high-task attitude sub-sample” (Kormos and Dörnyei 2004: 12).

In general, Kormos & Dörnyei (2004) conclude that motivation influenced the quantity and not the quality of talk produced. When the whole sample was considered, it was the course attitudes that had a positive significant effect on accuracy, and when the sample was sub-divided, it was the attitudes towards the language course itself that had a positive effect even if they did not like the specific tasks.

To sum up, even if the results of these studies are complex and have to be interpreted in relation to each other, it can be concluded that motivation as operationalized and measured in the context of this model, has a positive influence, whether on quantity or quality, on the outcome of the oral argumentative tasks used in these studies. Also, these studies support the use of a process oriented approach that is able to account for motivation not being static and having different phases in the context of foreign language learning through tasks.

In the present study, however, it is another type of task that will be investigated: a semi-guided writing task. Following the aforementioned definition of task (Dörnyei 2000), there are myriads of activities taking place in the L2 classrooms that should be considered as such. Therefore, Dörnyei's process model of motivation should serve as a valid tool to explain, at least in part, as it did with oral argumentative tasks, L2 performance in a writing task. Furthermore, in order for the model to be validated outside the scope of oral tasks, a preliminary step has to be taken: the validation of the main concept in the model, which is, in my view, task motivation. This is what I intend to do in this paper, pending further analysis of data gathered.

3. Second language writing

Currently, L2 writing skills are consistently worked on in almost every college second language program. Different L2 writing curricula have been made available to instructors. It is therefore widely acknowledged that developing this skill is as important as developing the speaking, listening, or reading skills. As in any other task taking place in the second language classroom, it is also to be expected that learners' motivation plays a crucial role in the outcome of the activity.

As far as L2 writing research is concerned, the field has grown and developed in an unparallel manner for at least the last 40 years. There is a wealth of research traditions with their own agendas and ideas of what should be investigated. Further complicating matters, the advent of the World Wide Web and its related technologies adds an element to the L2 writing field that cannot be overlooked due to its relevance and influence in today's society (for an overview, see Matsuda et al. 2004).

One of the strands of research in the field investigates possible measures for L2 writing and their relationship to L2 proficiency and development. Here, the focus is on how to better measure the learner's written production to match it with certain proficiency levels and to delimit a path of L2 writing development. It is this part of the field that is of interest for my purposes in this study, since I will be using some of the measures typically used by researchers in this line of work to assess the learners' performance in an L2 writing task.

One of the most significant contributions to SLA, in general, and to L2 writing in particular, is the volume by Wolfe-Quintero et al (1998). This research synthesis reviews 39 L2 writing studies to address mainly two issues: 1) How do the measures utilized in these studies evaluate L2 writing? 2) Which are the best measures for L2 writing development? These authors focus their attention on measures of fluency, accuracy, and complexity. After reviewing each study and detailing the assessment measures utilized in these studies, the authors propose a number of measures that appear to hold higher validity rates. For fluency, words per T-unit, words per clause, and words per error-free T-unit were proposed. Complexity was further sub-divided into grammatical (clause per T-unit and dependent clause per clause) and lexical complexity (word types per total number of words and sophisticated word types per word types). Finally, two measures are proposed for accuracy: error-free T-units per T-unit, and errors per T-unit.

In light of the statistical analysis of the thirty-nine studies, L2 proficiency seems to significantly correlate with increases in syntactic complexity only when proficiency is defined by programme level.

In another research synthesis, Ortega (2003) analyses 27 studies that investigated L2 writing at the college level. Twenty-one of these studies were cross-sectional and six longitudinal. The author set out to explore three main issues: first, the impact of the instructional setting and proficiency criteria on the mean values and range of a given syntactic complexity measure across the twenty one cross-sectional studies; second, differences in performance by two different proficiency groups for a given syntactic complexity measure across studies; finally, the author was interested in evaluating the amount of change when gains in performance relative to length of observation are compared across longitudinal studies.

Ortega (2003) focuses on the six most common syntactic complexity measures across studies. She identifies three measures of length of production, one measure of amount of coordination, and two measures of amount of subordination. The analysis of the data in the twenty-seven studies showed significant results. First, ESL learners produced writing of more syntactic complexity than that produced by FL learners. Second, studies that set proficiency level based on holistic ratings showed a more homogeneous range of results across groups. Third, some critical magnitudes were established based on the results achieved for between-group differences. Finally, three months of instruction result in little difference in mean length of T-unit across ESL groups and even less across FL groups. After one year, changes may be greater (Ortega 2003: 512).

Taking into account the above review and to test the validity of the Process model of motivation in the context of L2 writing tasks, I am going to investigate the relationship between task motivation, as measured in this model, and five syntactic complexity measures in the participants' L2 writing: number of words (NW), number of t-units (NTU), proportion of error-free t-units (EFTU), number of words per t-unit (NWTU), and lexical variety (LV). NW and LV, fluency and complexity measures are used by Kormos and Dörnyei (Kormos and Dörnyei 2004) in their investigation of oral argumentative tasks. NTU, EFTU, and NWTU are common measures for fluency and accuracy in the L2 writing research field as pointed out by Wolfe-Quintero et al (1998). While there are several definitions of the term t-unit in the field of L2 writing,

the current study will define it as a minimal, independent, terminable clause, which has all modifying phases attached to it (Larsen-Freeman and Long 1991).

The Ubber Index was utilized to determine LV. This formula is used in Kormos and Dörnyei (2004):

$$\text{Ubber Index: } (\log \text{ tokens})^2 / (\log \text{ tokens} - \log \text{ types})$$

The following are the research questions for which I attempt to find an answer in this study:

- 1) Is there a significant correlation coefficient between the task motivation scores as measured in this study and any of the five syntactic complexity assessment measures under investigation?
- 2) Is there a significant difference in performance in any of the five syntactic complexity measures across groups?

4. Method

Participants: Sixty-five participants took part in this study and completed all its parts. Participants were both male and female and all were enrolled in the Spanish foreign language program at Georgetown University in Washington, D.C. as undergraduate students. Thirty-two attended second year Spanish language classes and thirty-three were enrolled in third year Spanish classes.

Materials and scoring procedures. The materials and scoring procedures used in the present study are detailed below:

Participants were handed a sheet with instructions to write a semi-guided writing activity in Spanish. Instructions are shown below:

Write a single, continuous short story in the past about what you think happened in the comic strips below. Not only narrate the events but also describe the characters and include dialogue for Mafalda and her mom. For example, write about what Mafalda and her mom are doing, when they are doing it, what they are wearing, what they are saying, and, in the case of Mafalda, what she is thinking. Make sure to follow the strips and provide the necessary dialogues for the dialogue bubbles but do not write on the strips. Use transition words to integrate everything into the narrative based on the combined events in the three comic strips. Your story should be approximately one page in length. TRY TO BE AS CREATIVE AS POSSIBLE! THANKS.

Language motivation questionnaire: 45 items on a 7 point Likert-type scale (from Strongly agree to Strongly disagree). This questionnaire was adapted from Dörnyei and Kormos (2004). These items tapped into Integrativeness, Incentive value of learning Spanish, Course attitudes, Linguistic self-confidence, Language use anxiety, task attitudes, and Willingness to communicate (WTC). In this preliminary analysis, I will only consider task motivation.

The scoring procedure for the motivation questionnaire is as follows: The scoring range went from 10 (strongly disagree) through 70 (strongly agree). This order was inverted for negatively formulated questions, that is, 10 (strongly agree) through 70 (strongly disagree).

Structure of the study and procedure. The present study involved 65 L2 learners of Spanish. Participants were given a series of comic strips without captions and were asked to write a story about what they thought had happened in the comic strips. Five syntactic complexity measures were used to evaluate their writing: number of words (NW), number of t-units (NTU), proportion of error-free t-units (EFTU), number of words per t-unit (NWTU), and lexical variety (LV). With previous permission granted by the instructors of the class, the study was carried out during one class time period. First, participants did the semi-guided writing task for which twenty five minutes were allotted. Then, they filled in the motivational questionnaire taking as long as they needed. Several motivational variables were tapped into through this questionnaire: Integrativeness, Incentive value of learning Spanish, Course attitudes, Linguistic self-confidence, Language use anxiety, task attitudes, and Willingness to communicate (WTC). However, in this preliminary analysis only task attitudes were assessed for the reasons mentioned above.

Analysis. In order to find an answer for research question (RQ) 1, I ran a correlation analysis between task motivation and the five linguistic variables investigated in the participants' L2 writing: NW, NTU, EFTU, NWTU, LV. To address RQ 2, participants were divided in High and Low task motivation groups based on their responses to the task motivation items in the questionnaire. Groups were established taking into consideration the lowest and highest task motivation score, the mid-score between them was the cut-off point. In this manner, 25 participants were assigned to the Low task motivation group, while the remaining 40 participants were assigned to the High task motivation group. Once both groups were established, Independent samples *t*-test were run to find out if there were significant differences between the two groups in any of the five linguistic variables under investigation.

5. Results

RQ1) Is there a significant correlation coefficient between the task motivation scores as measured in this study and any of the five syntactic complexity assessment measures under investigation?

For RQ1, correlation analyses were run to investigate if there were any significant relationships between any of the linguistic variables (NW, NTU, EFTU, NWTU, LV) and task motivation scores.

In figure 1 below, the correlation matrix for task attitudes and all the linguistic variables is shown:

In figure 1, the results of the correlation analyses for task motivation and each of the linguistic variables are presented. This matrix table shows that there are several significant correlation coefficients between the variables concerned. Task motivation holds a linear relationship with NW ($r = .631$, $p = .000$), NTU ($r = .566$, $p = .000$), and LV ($r = .471$, $p = .000$).

Therefore, the relationship between task motivation and these linguistic variables tends to be linear and positive. In simple words, the more task motivation, the more number of words, the more t-units, and the more lexical variety can be found in the participants' L2 writing.

Figure 1
Correlation matrix

		TaskMot	NW	NTU	EFTU	NWTU	LV
TaskMot	Pearson Correlation	1	.631(**)	.566(**)	.187	-.141	.471(**)
	Sig. (2-tailed)	—	.000	.000	.135	.262	.000
	N	65	65	65	65	65	65
NW	Pearson Correlation	.631(**)	1	.713(**)	.553(**)	.055	.886(**)
	Sig. (2-tailed)	.000	—	.000	.000	.661	.000
	N	65	65	65	65	65	65
NTU	Pearson Correlation	.566(**)	.713(**)	1	.467(**)	-.058	.612(**)
	Sig. (2-tailed)	.000	.000	—	.000	.644	.000
	N	65	65	65	65	65	65
EFTU	Pearson Correlation	.187	.553(**)	.467(**)	1	.183	.716(**)
	Sig. (2-tailed)	.135	.000	.000	—	.143	.000
	N	65	65	65	65	65	65
NWTU	Pearson Correlation	-.141	.055	-.058	.183	1	.102
	Sig. (2-tailed)	.262	.661	.644	.143	—	.418
	N	65	65	65	65	65	65
LV	Pearson Correlation	.471(**)	.886(**)	.612(**)	.716(**)	.102	1
	Sig. (2-tailed)	.000	.000	.000	.000	.418	—
	N	65	65	65	65	65	65

** Correlation is significant at the 0.01 level (2-tailed)

The answer to RQ 1 is therefore positive regarding NW, NTU, and LV.

RQ2) Is there a significant difference in performance in any of the five syntactic complexity measures between High and Low motivation groups?

In figures 2 and 3 below, the descriptive statistics and the independent samples tests are found. As figure 2 shows, the High task motivation group is made up of 40 participants, whereas the Low motivation group contains the remaining 25 participants. The mean scores for each group in every task are also shown.

The results for the independent samples *t*-test shown in figure 3 reveal that the High/Low task motivation groups performed significantly different in regards to NW ($t=-4.261$, $p=.000$), NTU ($t=-4.675$, $p=.000$), and LV ($t=-2.631$, $p=.011$). Therefore, these results confirm what seemed to be large differences in main scores between the groups, as shown in figure 2.

It is thus possible to answer RQ2 in positive terms regarding NW, NTU, and LV.

Figure 2
Descriptive statistics for Independent samples test

	tasklevel	N	Mean	Std. Deviation	Std. Error Mean
NW	1.00	25	166.8400	44.36936	8.87387
	2.00	40	228.1250	62.68447	9.91129
NTU	1.00	25	13.8000	4.11299	.82260
	2.00	40	18.2750	3.51544	.55584
EFTU	1.00	25	49.4268	20.82639	4.16528
	2.00	40	54.9808	21.89071	3.46123
NWTU	1.00	25	12.3828	2.33848	.46770
	2.00	40	14.3025	10.71565	1.69429
LV	1.00	25	317.3940	96.18602	19.23720
	2.00	40	388.9663	112.68773	17.81749

Figure 3
Independent samples test. High/Low motivation groups

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper	
NW	Equal variances assumed	1.69	.198	-4.26	63	.000	-61.2850	14.38251	-90.0261	-32.5438
	not assumed			-4.60	61.92	.000	-61.2850	13.30335	-87.8786	-34.6913
NTU	Equal variances assumed	.387	.536	-4.67	63	.000	-4.47500	.95716	-6.38774	-2.5622
	not assumed			-4.50	45.13	.000	-4.47500	.99279	-6.47442	-2.4755
EFTU	Equal variances assumed	.000	.997	-1.01	63	.315	-5.55395	5.47927	-16.50340	5.39550
	not assumed			-1.02	53.02	.310	-5.55395	5.41568	-16.4163	5.30840
NWTU	Equal variances assumed	.813	.371	-.880	63	.382	-1.91970	2.18077	-6.27762	2.43822
	not assumed			-1.09	44.74	.281	-1.91970	1.75766	-5.46036	1.62096
LV	Equal variances assumed	.624	.433	-2.63	63	.011	-71.5722	27.20396	-125.9350	-17.2094
	not assumed			-2.730	57.01	.008	-71.5722	26.22085	-124.0782	-19.0662

p < .05

6. Discussion

The present study investigated the relationship between task motivation, as defined in Dörnyei (2000), Dörnyei and Ottó (1998), and five linguistic variables found in participants L2 writing. A semi-guided writing task was designed to elicit the data that would be analyzed for NW, NTU, EFTU, NWTU, and LV. These syntactic complexity measures are commonly used in the L2 writing literature and have been argued to be valid measures for L2 writing development (Wolfe-Quintero et al. 1998).

Dörnyei and associates' Process model of motivation focuses on the tasks carried out in the L2 classroom. In this context, it could be said that tasks are the minimal units around which motivation evolves. Furthermore, Dörnyei (2000) and Dörnyei and Ottó (1998) argue that motivation to learn a second or foreign language is not static, but that it changes along a continuum. Therefore, motivation has different phases along which it may change.

Due to the micro perspective that it presents, this model underscores tasks attitudes and motivation as the level at which motivation should be investigated. However, only oral argumentative tasks have been tested so far (Dörnyei 2002; Dörnyei and Kormos 2000, Kormos and Dörnyei 2004) with complex but positive results for the motivational variables. In the present study, I have presented results for task motivation, since this variable could be considered as one of the foundations of the model. Dörnyei and Kormos (2000) discuss the important role of task motivation in terms of acting as a filter. In other words, if the learners have good attitudes toward the task, their performance seems to follow regular patterns. On the contrary, if the learners hold negative attitudes toward the task, their performance appears to be random. Along the same lines, Kormos and Dörnyei (2004: 10) subdivide participants into " 'high-task attitude' (the upper half of the sample) and 'low-task attitude' (the lower half of the sample)" to run correlation analysis between motivational and linguistic variables. Results showed linear relationships at different levels that made the authors conclude that task motivation or attitudes toward the task seemed to be a crucial factor in determining further performance.

Results in the present study go along with the previous findings. Participants' attitudes toward the task surface as a very important factor in their L2 writing performance. Task motivation is significantly correlated with NW, NTU, and LV. In a nutshell, the better the attitudes toward the task, the more words, the more *t*-units, and the more lexical variety produced. In addition, Kormos & Dörnyei (2004) concluded that, in general, motivation influenced the quantity of speech produced, rather than the quality. Results reported here deviate slightly from those previous findings, since LV is a quality measure. In the present investigation, both quantity and quality of L2 writing produced are shown to be significantly correlated with measures of task motivation.

Task attitudes have been shown to be linearly correlated with linguistic measures in the L2 writing produced. It is therefore reasonable to state that attitudes toward the task or task motivation, as has been termed in this study, are a very important factor that can affect the linguistic outcome. Results for the *t*-test analysis further confirm this fact, since there are significant differences in performance bet-

ween High and Low task motivation groups. As with the correlation analyses, NW, NTU, and LV are the measures in which those differences are significant. These measures assess fluency and complexity, or in Dörnyei's terms, quantity and quality of writing.

Kormos & Dörnyei (2004) subdivided their sample into High and Low task motivation groups and found that the number of arguments produced orally correlated significantly with a complex of motivational variables. Hence, results presented here do not contradict previous findings and task motivation, the only motivational variable investigated, holds a significant linear association with three linguistic variables.

Finally, no significant correlation or difference in performance between the two task motivation groups have been found involving the other two linguistic variables concerned, EFTU and NWTU. To account for this fact, participants' proficiency level should be taken into consideration, given its possible influence in the outcome of the activity.

7. Limitations and future research

Two main limitations need to be mentioned. First, we have shown several significant correlation coefficients between task motivation and linguistic variables. However, this does not imply causation. Correlations test for possible linear relationships amongst variables. Positive linear relationships have been proven to exist between the variables as defined in this study, i.e. when the scores for one variable increase, the scores for the other variable increase as well. Second, participants' proficiency level could have had a bearing on the results achieved, but this variable has not been investigated here.

Given that results for the main motivational variable in the model have been positive in the context of the Process model, all the motivational variables utilized by Dörnyei and Kormos should be addressed in future research. Furthermore, proficiency level should also be added to the equation, since it could be a possible explanation for some of the results achieved.

8. Conclusion

In the present study I have contributed to the growing L2 motivation literature. I have shown that the model proposed by Dörnyei and associates, thus far only tested with oral argumentative tasks, has the potential for accounting for performance in other types of tasks. I have investigated the main variable's relationship with five linguistic variables in an L2 writing task. Not only have three of the linguistic measures been proven to be significantly related to the participants' motivation regarding this task, but also High and Low task motivation groups have been shown to significantly differ in linguistic outcome.

As mentioned above, future studies should address the remaining motivational variables and their relationship with different aspects of L2 performance.

9. References

- Dörnyei, Z., 2000, «Motivation in action: Towards a process-oriented conceptualisation of student motivation», *British Journal of Educational Psychology* 70.519-38.
- 2002, *The motivational basis of language learning tasks. Individual differences in second language acquisition*, ed. by Peter Robinson, 137-58. Amsterdam: John Benjamins.
- 2003, *Attitudes, Orientations, and Motivations in Language Learning: Advances in Theory, Research, and Applications. Attitudes, Orientations, and Motivations in Language Learning*, ed. by Z. Dörnyei, 3-32. Oxford: Blackwell.
- and Kormos, J., 2000, «The role of individual and social variables in oral task performance», *Language teaching research* 4.275-300.
- and Ottó, I., 1998, «Motivation in Action: A process model of L2 motivation», *Working papers in Applied Linguistics*, 43-69. London: Thames Valley University.
- Heckhausen, H., 1991, *Motivation and action*. Berlin; New York: Springer-Verlag.
- Julkunen, K., 2001, *Situation- and task-specific motivation in foreign language learning. Motivation and Second Language Acquisition*, ed. by Z. Dörnyei and R Schmidt, 28-41. Manoa: University of Hawaii Press.
- Kormos, J. and Dörnyei, Z., 2004, «The interaction of linguistic and motivational variables in second language task performance», *Zeitschrift für Interkulturellen Fremdsprachenunterricht* [Online] [<http://www.ualberta.ca/~german/ejournal/kormos2.htm>], 9.19.
- Kuhl, J., 1984, «Volitional aspects of achievement motivation and learned helplessness: Toward a comprehensive theory of action control». vol. 12. *Progress in experimental personality research*, ed. by B. A. Maher. 99-170. New York: Academic Press.
- Larsen-Freeman, D. and Long, Michael H., 1991, *An introduction to second language acquisition research*. New York: Longman.
- MacIntyre, P. D.; MacMaster, K. and Baker, S. C., 2001, *The convergence of multiple models of motivation for second language learning: Gardner, Pintrich, Kuhl, and McCroskey. Motivation and Second Language Acquisition*, ed. by Z. Dörnyei and R. Schmidt, 461-92. Manoa: University of Hawai'i at Manoa, Second Language Teaching and Curriculum Center.
- Matsuda, P. K.; Canagarajah, A. S., Harklau, L., Hyland, K. and Warschauer, M., 2004, «Changing currents in second language writing research: A colloquium», *Journal of Second Language Writing* 12.151-79.
- Ortega, L., 2003, «Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing», *Applied Linguistics* 24.492-518.
- Pintrich, P. R. and Schunk, D. H., 1996, *Motivation in education: Theory, research, and applications*. Englewood Cliffs, NJ: MerrillPrentice-Hall.
- Ushioda, E., 1996, *Learner autonomy 5: the role of motivation*. Dublin: Authentik.
- Wolfe-Quintero, K.; Inagaki, S. and Kim, H.-Y., 1998, *Second language development in writing: Measures of fluency, accuracy, and complexity*. Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center.

«Julio Urkixo» Euskal Filologi Mintegiaren Aldizkariaren Gehigarriak
Anejos del Anuario del Seminario de Filología Vasca «Julio de Urquijo»
Publications of «Julio de Urquijo» Seminar of Basque Philology

- I. *El Seminario "Julio de Urquijo". Antecedentes y constitución*, 1955. Agotado.
- II. JOSÉ MARÍA LACARRA, *Vasconia medieval. Historia y filología*, 1957. Agotado.
- III. MANUEL AGUD - LUIS MICHELENA, *Landuccio, Dictionarium Linguae Cantabrigae (1562)*, 1958. Agotado.
- IV. LUIS MICHELENA, *Fonética histórica vasca*, 1990 (1961¹, 1977², 1985³). 18 €.
- V. NILS M. HOLMER, *El idioma vasco hablado. Un estudio de dialectología vasca*, 1991 (1964¹). 12 €.
- VI. LUIS VILLASANTE, *Pedro A. de Añibarro, Gramática vascongada*, 1970. 8 €.
- VII. CÁNDIDO IZAGUIRRE, *El vocabulario vasco de Aránzazu-Oñate y zonas colindantes*. (Índice inverso de Gerardo Markuleta), 1994 (1970¹). 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. 18 €.
- X. LUIS MICHELENA, *Sobre historia de la lengua vasca*, 1988. 2 vols. Agotado.
- 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. Agotado.
- XIV. JOSEBA A. LAKARRA (ed.), *Memoriae L. Mitxelena magistri sacrum*, 1991. 36 €.
- XV. RICARDO GÓMEZ - JOSEBA A. LAKARRA (arg.), *Euskalaritzaren historiaz I: XVI-XIX. mendak*, 1992. 18 €.
- XVI. BEÑAT OYHARÇABAL, *La pastorale souletine: édition critique de "Charlemagne"*, 1990. 18 €.
- XVII. RICARDO GÓMEZ - JOSEBA A. LAKARRA (arg.), *Euskalaritzaren historiaz II: XIX-XX. mendak*. Prestatzen.
- XVIII. JOSEBA A. LAKARRA, *Harrieten Gramatikako biztegiak (1741)*, 1994. 10 €.
- XIX. MANUEL AGUD - ANTONIO TOVAR, *Diccionario etimológico vasco, II. Arduin-Beuden*, 1990, 1993. Agotado.
- XX. LUIS MICHELENA, *Lenguas y protolenguas*, 1990 (1963, 1986). 8 €.
- XXI. ARENE GARAMENDI, *El teatro popular vasco. Semiótica de la representación*, 1991. 12 €.

- xxii. LASZLÓ K. MARÁ CZ, *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 (Barraso-Bazur)*, 1991. Agotado.
- xxv. JOSÉ M.^a SÁNCHEZ CARRIÓN, *Un futuro para nuestro pasado*, 1991. 15 €.
- xxvi. MANUEL AGUD - ANTONIO TOVAR, *Diccionario etimológico vasco, IV. Egiluma-Galanga*, 1991. Agotado.
- 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, 1991ko 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. 18 €.
- 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)*. En preparación.
- xl. AMAIA MENDIKOETXEA - MYRIAM URIBE-ETXEBARRIA (eds.), *Theoretical issues at the morphology-syntax interface*, 1997. 21 €.
- xli. BERNARD HURCH - MARÍA JOSÉ KEREJETA, *Hugo Schuchardt - Julio de Urquijo: Correspondencia (1906-1927)*, 1997. 21 €.
- xlii. JOSÉ I. HUALDE, *Euskararen azentuerak*, 1997. 15 €.
- xliiii. RUDOLF P. G. de RIJK, *De lingua Vasconum: Selected Writings*, 1998. 15 €.
- xliv. XABIER ARTIAGOITIA - PATXI GOENAGA - JOSEBA A. LAKARRA (arg./eds.), *Erramu Bona: Festschrift Rudolf P. G. de Rijk*, 2002. 30 €.
- xlv. JOSEBA A. LAKARRA, *Ikerketak euskararen historiaz eta euskal filologi az. Argitaratzeko*.

- XLVI. BEÑAT OYHARÇABAL (ed.), *Inquiries into the lexicon-syntax relations in Basque*, 2003. 18 €.
- XLVII. BLANCA URGELL, *Larramendiren "Hiztegi Hirukoitza"-ren Eranskina: saio bat hiztegi-gintzaren testukritikaz*. Argitaratzeko.
- XLVIII. ÍÑIGO RUIZ ARZALLUZ, *"Aitorkizunen" historia eta testua: Orixeren eskuizkributik Lekuonaren edizioa*, 2003. 21 €.
- XLIX. GOTZON AURREKOETXEA - XARLES VIDEGAIN (arg.), *Haur prodigoaren parabola Ipar Euskal Herriko 150 bertsiotan*, 2004. 21 €.
- L. JOSEBA A. LAKARRA, *Raíz y reconstrucción del protovasco*. En prensa.
- LI. XABIER ARTIAGOITIA - JOSEBA A. LAKARRA (arg.), *Gramatika jaietan Patxi Goenagaren Omenez*, 2008.

MONUMENTA LINGVAE VASCONUM
STUDIA ET INSTRUMENTA

- I. BLANCA URGELL, *Larramendiren "Hiztegi Hirukoitza"-ren Eranskina: saio bat hiztegi-gintzaren testukritikaz* (= Gehigarriak XLVII). Argitaratzeko.
- II. ÍÑIGO RUIZ ARZALLUZ, *"Aitorkizunen" historia eta testua: Orixeren eskuizkributik Lekuonaren edizioa*, 2003, (= Gehigarriak XLVIII). 21 €.
- III. OROITZ JAUREGI, *Correspondencia de Gerhard Bähr con R. M. Azkue, H. Schuchardt y J. Urquijo (1920-1944)*, (= ASJU xxxvi-2), 21 €.
- IV. CÉLINE MOUNOLE HIRIART-URRUTY, *C. H. de Belsunce Bizkondea Tableau analytique et grammatical de la langue basque (1858) azterketa eta edizioa* (= ASJU xxxvii-2).

BIBLIOGRAFIA-LABURDURA GOMENDATUAK
ABREVIATURAS BIBLIOGRÁFICAS RECOMENDADAS
RECOMMENDED BIBLIOGRAPHICAL ABBREVIATIONS

Hemen agertzen ez denerako, erabil bitez *Orotariko Euskal Hiztegia*-n agertzen direnak.

Para las obras no citadas abajo, se emplearán las abreviaturas del Diccionario General Vasco.

For any works which do not appear below, the abbreviations given in the *Diccionario General Vasco* should be used.

- AEF* = *Anuario de Eusko Folklore*, Vitoria-Gasteiz, 1921-1936; Donostia-San Sebastián, 1956-
- AION* = *Annali dell'Istituto Orientale di Napoli*, Napoli, 1979-
- ASJU* = *Anuario del Seminario de Filología Vasca "Julio de Urquijo"*. *International Journal of Basque Linguistics and Philology*, Donostia-San Sebastián, 1954-1955, 1967-
- Azk* = Resurrección M.^a de Azkue, *Diccionario vasco-español-francés*, Bilbao, 1905-1906 [1969², 1984³].
- Azk Morf* = Id., *Morfología vasca (Gramática básica dialectal del euskera)*, Bilbao, 1923-1925 [1969²].
- BAP* = *Boletín de la Real Sociedad Vascongada de Amigos del País*, Donostia-San Sebastián, 1945-
- BGS* = *Beitrag zur Geschichte der Sprachwissenschaft*, Münster, 1991-
- BISS* = *Boletín de la Institución "Sancho el Sabio"*, Vitoria-Gasteiz, 1957-81. Vide *Sancho el Sabio*.
- BMB* = *Bulletin du Musée Basque*, Baiona, 1924-43, 1964-
- BRAE* = *Boletín de la Real Academia Española*, Madrid, 1914-
- BRAH* = *Boletín de la Real Academia de la Historia*, Madrid, 1877-
- BSL* = *Bulletin de la Société de Linguistique de Paris*, Paris, 1884-
- BLS* = (*Proceedings of the*) *Berkeley Linguistics Society*, Univ. of California, Berkeley, 1975-
- CAJ* = *Central Asiatic Journal*, Wiesbaden, 1955-
- Campión* = Arturo Campión, *Gramática de los cuatro dialectos literarios de la lengua éuskara*, Iruñea/Pamplona, 1884 [1977²].
- CEEN* = *Cuadernos de Etnografía y Etnología de Navarra*, Pamplona, 1969-
- CIL* = *Corpus Inscriptionum Latinarum*, Berlin, 1863-
- CLAO* = *Cahiers de Linguistique - Asie Orientale*, Paris, 1971-
- CLS* = (*Proceedings of the*) *Chicago Linguistics Society*, Univ. of Chicago, 1965-

- Contr* = Ibon Sarasola, "Contribución al estudio y edición de textos antiguos vascos", *ASJU* 17 (1983): 69-212. L. Michelena & I. Sarasola, *Textos arcaicos vascos. Contribución...*, Anejos de *ASJU* 11, Donostia-San Sebastián, 1989.
- DCECH* = Juan Corominas & José Antonio Pascual, *Diccionario crítico etimológico castellano e hispánico*. Madrid, Gredos, 1980-1991.
- DELL* = Alfred Ernout & Antoine Meillet, *Dictionnaire étymologique de la langue latine. Histoire des mots*, Paris, 1932 [1939², 1951³, 1959⁴].
- DGV* = vide *OEH*.
- Diachronica* = *Diachronica. International Journal for Historical Linguistics*, Amsterdam-Philadelphia, 1984-
- DRA* = Manuel de la Sota, Pierre Lafitte, Lino de Akesolo et al., *Diccionario Retana de Autoridades de la Lengua Vasca*, Bilbao, 1976-1989.
- Euskera* = *Euskera. Euskaltzaindiaren lan eta agiriak*, Bilbao, 1920-1936, 1953-
- EAA* = *Estudios de Arqueología alavesa*, Vitoria-Gasteiz, 1966-
- EFDA* = Luis Michelena, *Estudio sobre las fuentes del diccionario de Azkue*, Bilbao, 1970 [= Azk 1984].
- EFOu* = *Études finno-ougriennes*, Paris, 1964-
- EH* = Ibon Sarasola, *Euskal hiztegia*, Donostia-San Sebastián, 1996.
- EI* = Ana M.^a Echaide (arg.), *Erizkizundi irukoitza*, Bilbao, 1984.
- EJ* = *Eusko Jakintza*, Baiona, 1947-1957.
- ELH* = *Enciclopedia Lingüística Hispánica*, Madrid, 1959-
- FEW* = W. von Wartburg, *Französisches Etymologisches Wörterbuch*, Bonn, 1928-
- FHV* = Luis Michelena, *Fonética histórica vasca*, Anejos de *ASJU* 4, Donostia-San Sebastián, 1961, 1977² [1985, 1990].
- FLV* = *Fontes Linguae Vasconum. Studia et documenta*, Iruñea/Pamplona, 1969-
- FL* = *Folia Linguistica. Acta Societatis Linguisticae Europaeae*, Berlin-New York, 1967-
- FLH* = *Folia Linguistica Historica. Acta Societatis Linguisticae Europaeae*, Berlin-New York, 1980-
- GH* = *Gure Herria*, Baiona, 1921-
- HEL* = *Histoire, Epistémologie, Langage*, Paris, 1979-
- HL* = *Historiographia Linguistica: International Journal for the History of the Language Sciences*, John Benjamins, 1974-
- HLEH* = Ibon Sarasola, *Hauta-lanerako euskal hiztegia*, Donostia-San Sebastián, 1984-1995. Vide *EH*.
- HLV* = Luis Michelena, *Historia de la literatura vasca*, Madrid, 1960 [1988].
- HLV* = Luis Villasante, *Historia de la literatura vasca*, Bilbao, 1961, 1979².
- HomUrq* = *Homenaje a don Julio de Urquijo e Ybarra*, Donostia-San Sebastián, 1949-1951.

- HSLV = Ibon Sarasola, *Historia social de la literatura vasca*, Madrid, 1976 [1982].
- I EW = Julius Pokorny, *Indogermanisches Etymologisches Wörterbuch*, Berna, 1951-1969.
- IF = *Indogermanische Forschungen*, Berlin, 1892-
- IJAL = *International Journal of American Linguistics*, Chicago, 1917-
- IL = *Indian Linguistics. Journal of the Society of India*, Pune (India), 1931-
- IMU = *Italia medioevale e umanistica*, Padova, 1958-
- Incipit = *Incipit. Seminario de edición y crítica textual*, Buenos Aires, 1981-
- JALL = *Journal of African Languages and Linguistics*, Berlin-New York, 1979-
- JEAL = *Journal of East Asian Linguistics*, Berlin, etc., 1992-
- JWAL = *Journal of West African Languages*, Dallas, 1964-
- Lexicographica = *Lexicographica. Internationales Jahrbuch für Lexikographie*, Tübingen, 1985-
- Lg = *Language*, Baltimore, 1924-
- Lb = P. Lhande, *Dictionnaire Basque-Français*, Paris, 1926.
- LH = Luis Michelena, *Lengua e historia*, Madrid, Paraninfo, 1985.
- LI = *Linguistic Inquiry*, Cambridge (Mass.), 1971-
- MDEV = Manuel Agud & Antonio Tovar, *Materiales para un diccionario etimológico vasco*, Anejos de ASJU 13, Donostia-San Sebastián, 1989-
- Memoriae = Joseba A. Lakarra (ed.), *Memoriae L. Mitxelena Magistri Sacrum*, Anejos de ASJU 14, Donostia-San Sebastián, 1991.
- NLLT = *Natural Language and Linguistic Theory*, Dordrecht, 1983-
- NTS = *Norks Tidsskrift for Sprogvidenskap*, Oslo, 1928-
- OEH = Luis Michelena, *Diccionario General Vasco. Orotariko Euskal Hiztegia*, Donostia-San Sebastián, 1987-2005.
- OL = *Oceanic Linguistics*, Univ. of Hawaii, 1962-
- Phonology = *Phonology*, Cambridge, 1984-
- PT = Luis Michelena, *Palabras y textos*, Bilbao, UPV/EHU, 1987.
- PV = *Príncipe de Viana*, Pamplona, 1940-
- RDTP = *Revista de Dialectología y Tradiciones Populares*, Madrid, 1944-
- REW = W. Meyer-Lübke, *Romanisches etymologisches Wörterbuch*, Heidelberg, 1930³.
- RFE = *Revista de Filología Española*, Madrid, 1914-
- RIEV = *Revista Internacional de los Estudios Vascos*, Paris-San Sebastián, 1907-1936, 1983-
- RLPhC = *Revue de Linguistique et Philologie Comparée*, Paris, 1867-1916.
- RPh = *Romance Philology*, Berkeley (CA), 1947-
- SAL = *Studies in African Linguistics*, Bloomington (Indiana) / Columbus (Ohio), 1970-

- Sancho el Sabio* = *Sancho el Sabio: revista de cultura e investigación vasca = euskal kultura eta ikerketa aldizkaria*, Vitoria-Gasteiz, 1991-
- SHLV* = Luis Michelena, *Sobre historia de la lengua vasca*, Anejos de *ASJU* 10, J. A. Lakarra (ed.), Donostia-San Sebastián, 1988.
- Symbolae* = José Luis Melena (ed.), *Symbolae Ludovico Mitxelena Septuagenario Oblatae*, Vitoria-Gasteiz, Instituto de Ciencias de la Antigüedad-Antzinate-Zientzien Institutua, 1985.
- Syntax* = *Syntax, A Journal of Theoretical, Experimental and Interdisciplinary Research*, Oxford, etc., 1998-
- TAV* = Luis Michelena, *Textos arcaicos vascos*, Madrid, Minotauro, 1964 [= Luis Michelena - Ibon Sarasola, *Textos arcaicos vascos. Contribución*, Anejos de *ASJU* 11, Donostia-San Sebastián, 1989].
- TPhS* = *Transactions of the Philological Society*, London, 1842-
- UAJ* = *Ural-Altäische Jahrbücher*, Wiesbaden, 1981-
- Vinson = Julien Vinson, *Essai d'une bibliographie de la langue basque*, Paris, 1891-1898 [vide Vinson-Urquijo].
- Vinson-Urquijo = Julien Vinson, *Essai... con las anotaciones del ejemplar de Julio de Urquijo*, Anejos de *ASJU* 9, Donostia-San Sebastián, 1984.
- ZRPb* = *Zeitschrift für romanische Philologie*, Halle, 1877-

EGILEENTZAKO OHARRAK

ASJU-n euskaraz edo nazioarteko zientzi elkarteetan ohiko diren hizkuntzetako batean idatziriko euskal linguistika eta filologiazko lanak argitaratzen dira, baita eremu ezberdin edo zabalago bati atxikiak izan arren, euskalaritzarako interesgarri izan daitezkeenak ere. Originalak helbide honetara bidali behar dira: Joseba A. Lakarra, Hizkuntzalaritza eta Euskal Ikasketak Saila, Filologia eta Geografi-Historia Fakultatea, Unibertsitateko ibilbidea 5, 01006 Gasteiz (joseba.lakarra@ehu.es).

ASJU-ra igorritako artikulua gutxienez bi aztertzailek irakurriko dituzte, haien iruzkinak kontuan izan atera edo ez erabakitzeko; erabakia ahalik eta lasterren gaztigituko zaie egileei. Artikulua onartzekotan, oztopo, akats edo aldabeharren zerrenda ere emango zaie. Egileek lehendabiziko inprenta probak jasoko dituzte (eta originalarekin batera itzuli beharko dituzte); eskuratzen dituztenetik astebeteko epea izango dute zuzentzeko. Argitaratzailearen baimenik gabe ezingo dute garrantziko aldaketa, gehiketa edo kenketarik egin. Egileei *ASJU*-ko zenbakiaren ale bana eta lanaren 25 separata emango zaizkie (10, liburu iruzkinak badiira); gehiago nahi izanez gero, kostu prezioan agin ditzakete aurretiaz.

Ez da inongo murrizketarik originalen luzeraz, baina ez lukete izan behar berez behar baino gehiagokoa; lanek zehatzak eta argiak beharko dute izan. Berariazko abegia egingo zaie ohar laburrei, batez ere dagoeneko argitaratu beste lanen bat kritikatzin edo garatzen dutenean.

Originalen hasieran egilearen/egileen helbidea, telefona eta helbide elektronikoa ezarriko dira; biko espazioan, orrialde bakarrean, eta zein-nahi argitasun edo zuzenketarako albo guztietan zuriune zabalekin idatzirik aurkeztuko dira lanak. Orrialdeak eta oin-oharrak segidako zerrendan zenbatuko dira. Lana euskarri elektronikoa (programa erabilienetako batean) eta paperean (3 kopia) bidaliko da. Horrekin batera 10-20 lerroko laburpena ere erantsiko da. Aurkeztu baino lehen zuzen bedi ahalik eta hobekienik originala, inprenta hutsak gutxitzeko; orobat, argazki, lauki, mapa, grafiko, taula, irudi, etab. eman ez gero, izan bitez kalitate handienekoak gardentasunik gal ez dezaten. Hauek guztiak zenbatuko dira eta ezagutzeko oinperpau laburra erantsiko zaie, testuan ere non jarri nahi diren argiro markaturik. Adibideak zenbatu egingo dira: (1), (2)a, (2)b, etab.; testuan aipatzerakoan egin bedi era honetan: (2a), (2b), (2a, b), (4d-h), etab. Inprentan ohiko ez den zein-nahi zeinu, letra edo diakritikoren azalpen argia ezarriko da lehendabiziko agerrialdian testu aldameneko zuriunean.

Testua honako arauok beteaz aurkeztuko da: Aipu luzeak ahapaldi berezian joango dira, sangratuta, hasiera eta amaiera kakotxik gabe, letra borobilean; aipu laburrak ere borobilean, testuan bertan eta kakotx bikoitzen artean (“ ” edo « »). Kakotx bakunak (‘ ’) adierak edo hitz solteen itzulpenak emateko baliatuko dira. Metalinguistikoki erabilitako edota artikulua idazteko erabili den hizkuntzaz beste bateko hitzak letra etzanean ezarriko dira.

Liburu eta aldizkariaren izenei letra etzana dagokie, eta kakotxak artikuluei. Aldizkariaren zenbaki, urte eta orrialdeak eta liburuen argitaletxe eta edizio (ez inprimatze) tokia emango dira. Hala dagokionean, berriinprimatzea, berrargitalpena edo itzulpena den zehaztuko da. Aipuetarako erabil bedi urte-egile sistema, ahal den neurrian, eta urte bereko egile baten lan bat baino gehiago aipatu bada, a, b... hurrenkeran bereiziko dira: adib. (Vinson 1897a: 35-38), (ikus Lacombe 1924, Azkue 1923-25, Unhlenbeck 1947b). Amaierako bibliografiarik ez bada, eman bitez bibliografia zehaztasunak oro soilik lehen agerrialdian, eta ondokoetan egilearen deitura eta lanaren izenburu laburtua bakarrik, *op. cit.* eta *ibidem* direlakoak saihestuaz: adib. Guerra, *Cantares*, 22-24. Bibliografia ere biko espazioan idatziko da, eta honako formatu honi atxikiko zaio:

Mitxelena, K., 1950b, “La aspiración intervocálica”, *BAP* 6, 443-449. Berrarg. bere *Sobre historia de la lengua vasca*, *ASJU*-ren Gehigarriak 10, Donostia 1988, I, 191-202.

—, 1981a, “Lengua común y dialectos vascos”, *ASJU* 15, 291-313. Berrarg. bere *Palabras y Textos*, UPV/EHU, Vitoria-Gasteiz 1987, 35-55.

Ortiz de Urbina, J., 1989, *Some parameters in the grammar of Basque*, Foris, Dordrecht.

Rijk, R. P. de, 1985, “Un verbe méconnu”, in J. L. Melena (arg.), *Symbolae Ludovico Mixelena Septuagenario Oblatae*, UPV/EHU, Vitoria-Gasteiz, II, 921-935.

Sarasola, I., 1986, “Larramendiren eraginaz eta”, *ASJU* 20: 1, 203-216.

Bibliografi laburdurarako erabil bedi ale honetan bertan erantsi den laburdura gomendatuen zerrenda. Beharrezkoa balitz, egileak besterik ere erabili ahalko luke, beti ere esangura lehendabiziko agerrialdian azalduz.

INFORMATION FOR AUTHORS

Papers on Basque linguistics and philology, and more general fields related or of interest to Basque studies are accepted, provided they are written in the languages most used by the international scientific community. Submissions should be sent to: Joseba A. Lakarra, Department of Linguistics and Basque Studies, Faculty of Philology and Geography and History, Unibertsitate Etorbidea/Paseo de la Universidad 5, 01006 Vitoria-Gasteiz (joseba.lakarra@ehu.es).

Papers received by *ASJU* are submitted to at least two reviewers; the decision on publication is communicated to the author(s) within as short a time as possible. Should a paper be accepted, a list of objections or changes deemed necessary will be sent to the author(s). When the authors receive the first proofs of their work, these should be returned to the editor together with the original within one week. No changes, additions or deletions may be made without the permission of the editor. Authors receive a copy of the *ASJU* volume in which their article appears and 25 offprints of their papers (10 in the case of reviews). Further additional offprints may be ordered at cost price.

There is no restriction on the maximum length of submissions, but they should be no longer than is necessary; authors must be concise and clear. Preference will be given to short notes, especially when they offer critique or elaborate on previously published papers.

The originals, which should include the address, telephone number(s) and e-mail of the author(s) in the first lines, must be typed and double-spaced throughout on single-sided sheets; this also applies to the notes. Wide margins for possible corrections or clarifications are required. Pages are to be numbered serially, as are the notes. Manuscripts must be submitted in digital format (in one of the commonly used program formats) together with three printed copies, and must include an abstract of 10-20 lines. It is recommended that the paper be carefully corrected before presentation to avoid possible errors, and that photographs, pictures, maps, graphs, tables, figures, etc. be of the best possible quality to avoid loss of detail in reproduction. These graphics should all be numbered and have a short footnote or key for identification; their approximate position in the text should also be indicated. Examples should be labelled consecutively by numbers enclosed in brackets: (1), (2)a, (2)b, etc.; these bracketed numbers should be used in the body of the text when referring to examples, like so: (2a), (2b), (2a, b), (4d-h), etc. A clear description of any unusual symbols, characters or diacritics should be provided in the margin on their first occurrence.

The text must be formatted as follows: long quotations must be indented, without inverted commas at the beginning and end of the text, in plain type; short quotations, also in plain type, must be enclosed by double inverted commas (“ ” or « »). Single quotation marks (‘ ’) are to be used to denote translations of isolated terms. Terms used metalinguistically or in a language different to that of the text should be in italics.

The titles of books and journals should be in italics and those of papers between inverted commas. The issue, year and page numbers of journals should be given, and for books, the publisher's name and place of edition; where relevant, state whether the quotation is from a reprint, reedition or translation. Where possible use the author-year system for quotation, e.g. (Lafitte 1976a: 35-38), (see Schuchardt 1900, Azkue 1923-25, 1935). Otherwise, the complete bibliographical information should be given only on the first occurrence, limiting any subsequent references to the surname of the author and the abbreviated title (avoiding notations such as *op cit.* and *ibidem*), e.g. Altuna, *Versificación*, pp. 43-57. The bibliography must also be double-spaced, with the following format:

- Mitxelena, K., 1950b, “La aspiración intervocálica”, *BAP* 6, 443-449. Reed. in *Sobre historia de la lengua vasca*, Supplements of *ASJU* 10, Donostia 1988, I, 191-202.
—, 1981a, “Lengua común y dialectos vascos”, *ASJU* 15, 291-313. Reed. in *Palabras y Textos*, UPVI/EHU, Vitoria-Gasteiz 1987, 35-55.
Ortiz de Urbina, J., 1989, *Some parameters in the grammar of Basque*, Foris, Dordrecht.
Rijk, R. P. de, 1985, “Un verbe méconnu”, in J. L. Melena (ed.), *Symbolae Ludovico Mitzelena Septuagenario Oblatae*, UPV/EHU, Vitoria-Gasteiz, 921-935.
Sarasola, I., 1986, “Larramendiren eraginaz eta”, *ASJU* 20: 1, 203-216.

For abbreviations of secondary sources the “Abbreviation Index” published in this issue should be used. If necessary, other abbreviations may be used, and these should be made explicit on their first appearance in the text.

INFORMACIÓN PARA LOS AUTORES

ASJU publica artículos sobre lingüística y filología vasca y campos más generales relacionados con o de interés para la vascofilia, escritos en euskera o en cualquiera de los idiomas utilizados por la comunidad científica internacional. Los originales se enviarán a: Joseba A. Lakarra, Depto. de Lingüística y Estudios Vascos, Facultad de Filología y Geografía e Historia, Paseo de la Universidad 5, 01006 Vitoria-Gasteiz (joseba.lakarra@ehu.es).

Los artículos recibidos en *ASJU* son examinados al menos por dos revisores, cuyos informes condicionan su aceptación; la decisión será comunicada a los autores lo antes posible. En caso de aceptarse el artículo, se enviará al autor una lista de objeciones o cambios necesarios. Los autores recibirán las primeras pruebas de su trabajo (que deberán devolver junto con el original), y dispondrán para corregirlas de un plazo no superior a una semana desde su recepción. No podrán hacerse cambios, adiciones o supresiones importantes sin permiso del editor. Se entregará a cada autor un ejemplar del *ASJU* y 25 separatas de su artículo (10 en el caso de las reseñas), si bien es posible encargar con antelación otras adicionales, que le serán facturadas a precio de coste.

No existe restricción alguna sobre la longitud máxima de los originales, pero no deben ser más largos de lo necesario; los autores han de ser concisos y claros. Se dará preferencia a notas breves, particularmente cuando critiquen o desarrollen artículos publicados con anterioridad.

Los originales, al comienzo de los cuales se hará constar la dirección, el teléfono y la dirección electrónica del(os) autor(es), se presentarán escritos a doble espacio y por una sola cara, con márgenes amplios para posibles correcciones y aclaraciones. Las páginas irán numeradas correlativamente, así como las notas. Los manuscritos se enviarán en soporte electrónico (en alguno de los programas más usados) y en papel por triplicado, e irán acompañados de un resumen de entre 10 a 20 líneas. Se recomienda que el original sea corregido antes de su presentación para minimizar las erratas, y que fotografías, cuadros, mapas, gráficos, tablas, figuras, etc., sean de la mejor calidad posible para evitar pérdidas de detalle en la reproducción; todos ellos irán numerados y llevarán un breve pie o leyenda para su identificación; se indicará asimismo el lugar aproximado de colocación en el texto. Los ejemplos irán numerados: (1), (2)a, (2)b, etc.; al referirse a ellos en el texto se usará el formato (2a), (2b), (2a, b), (4d-h), etc. Se aclarará al margen en su primera aparición en el texto cualquier símbolo, carácter o marca diacrítica inusual.

El texto se presentará de acuerdo con las siguientes normas:

1) Las citas largas irán en texto sangrado, sin comillas al principio y final, en letra redonda; las citas breves, también en redonda, irán entre comillas dobles (“ ” o « »). Se usarán los ápicos (‘ ’) para denotar acepciones o traducciones de términos aislados. Los términos utilizados metalingüísticamente o en una lengua distinta a aquella en la cual está redactado el texto irán en cursiva.

2) Los títulos de libros y revistas irán en cursiva, y los de los artículos entre comillas. Se indicará el n.º, año y páginas correspondientes de las publicaciones periódicas, y editorial y lugar de edición de los libros; en su caso se indicará si se cita de una reimposición, reedición o traducción. Útese en la medida de lo posible el sistema autor-año para las citas, p. ej. (Lafitte 1967a: 35-38), (véase Schuchardt 1900, Azkue 1923-25, 1935). En su defecto, se darán datos bibliográficos completos sólo en la primera ocasión, limitándose en las siguientes a señalar apellido del autor y título abreviado, evitando *op. cit.* e *ibidem*: Altuna, *Versificación*, pp. 43-57. La bibliografía irá también a doble espacio, ajustándose al siguiente formato:

Mitxelena, K., 1950b, “La aspiración intervocálica”, *BAP* 6, 443-449. Reproducido en su *Sobre historia de la lengua vasca*, Anejos del *ASJU* 10, Donostia 1988, I, 191-202.

—, 1981a, “Lengua común y dialectos vascos”, *ASJU* 15, 291-313. Reproducido en su *Palabras y Textos*, UPV/EHU, Vitoria-Gasteiz 1987, pp. 35-55.

Ortiz de Urbina, J., 1989, *Some parameters in the grammar of Basque*, Foris, Dordrecht.

Rijk, R. P. de, 1985, “Un verbe méconnu”, in J. L. Melena (ed.), *Symbolae Ludovico Mitxelena Septuagenario Oblatae*, UPV/EHU, Vitoria-Gasteiz, II, 921-935.

Sarasola, I., 1986, “Larramendiren eraginaz eta”, *ASJU* 20: 1, 203-216.

3) Para las abreviaturas de fuentes primarias o secundarias se recurrirá al índice de abreviaturas recomendadas publicado en este mismo número. En caso necesario el autor podrá utilizar otras, cuyo valor explicará en la primera aparición.

ANUARIO DEL SEMINARIO DE FILOLOGÍA VASCA
"JULIO DE URQUIJO"

International Journal of Basque Linguistics and Philology
ASJU, xli-2, 2007

Preface	V
Introduction	VII
ASIER ALCÁZAR, Consumer Eroski parallel corpus	1
CARLOS DE CUBA, Negative polarity licensing, factivity and the CP field	11
RICARDO ETXEPARE, Aspects of quotative constructions in Iberian Spanish	25
KATA GÁBOR, ENIKŐ HÉJA, Complements and adjuncts in machine translation	59
ÁNGEL J. GALLEGO, Phase theory, case and relative clauses	71
VALÉRIE GAUTIER, How strong islands are derived from the way a top-down derivation is linearized	95
ION GIURGEA, Adjectives and proper nouns in Romance and English .	107
ENIKŐ HÉJA, KATA GÁBOR, Extracting information from participial structures	121
MAIDER HUARTE, The acquisition of Basque ergative case: an experimental study	131
DANIEL HUBER, On the interaction of velars and labials	145
ARITZ IRURTZUN, The structure of pair-list answers	163
ARITZ IRURTZUN, ÁNGEL J. GALLEGO, Consequences of pair-merge (at the interfaces)	179
ALEXANDRU MARDALE, Case marking and prepositional marking. Some remarks concerning <i>de</i> -phrases in Romanian	201
AYA MELTZER, Adjectival passives and adjectival decausatives in Hebrew	209
SUSANNA PADROSA TRIAS, Argument structure and morphology: the case of <i>en</i> -prefixation revisited	225
DORIS PENKA, A crosslinguistic perspective on n-words	267
DIEGO PESCARINI, Types of syncretism in the clitic systems of Romance	285
OMER PREMINGER, Argument-mapping and extraction	301
INMA TABOADA, Prenominal and postnominal demonstratives in Spanish: a [± deictic] approach	323
ÍÑIGO A. YANGUAS, A look at second language learners' task motivation	333