Università
Ca'Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School

Dottorato di ricerca
in Scienze del Linguaggio
Ciclo 27
Anno di discussione 2015

Nominal modification in Italian Sign Language (LIS)

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: L-LIN/01
Tesi di Dottorato di Lara Mantovan, matricola 824989

Coordinatore del Dottorato

Prof. Alessandra Giorgi

Tutore della Dottoranda

Prof.ssa Anna Cardinaletti

Co-tutore della Dottoranda

Prof. Carlo Geraci

E Pluribus Unum

(uncertain origin, attributed to

Virgilio, Moretum, v. 103)

Acknowledgments

Three years ago I began this journey with many question marks. I jumped on the *research apprenticeship* train without knowing exactly where it would take me. I had the good fortune to travel with people who kept me company, showed me the way, supported me and encouraged me to carry on. Now, I am almost at the end of this journey. Being so close to finishing this thesis, I would like to look back – at least for a while – to remember and thank all these people.

First and foremost, I would like to thank my dissertation advisors, Carlo Geraci and Anna Cardinaletti. Without their constant guidance, my train would probably have lacked a track on which to run. I thank them both for having supported me with great professionalism throughout the process, and for having provided me with a wealth of knowledge and skills.

I thank Carlo Geraci because he was the first person who talked to me about the chance to start this new experience and because he supervised me from Paris during all the stages of this work. Despite geographical distance, I could always count on him. In his weekly work schedule he always found the time for a Skype meeting in which we could take stock of the situation and discuss new directions. With unceasing dedication, he was always ready to address my doubts and correct my mistakes. Not only did he supervise me during all stages of this dissertation, but he also imparted both rigorous scientific methodology and technical expertise. I thank him also because he ensured that I spent a four-month research period with him and the Sign Language Linguistics group at Institut Jean-Nicod in Paris. This valuable experience allowed me to immerse myself in an international research context and appreciate the importance of working in a research team.

I thank Anna Cardinaletti because she has been a constant reference point in Venice. Despite her various academic commitments, she was always ready to meet with me and to dedicate her time to me. With patience and understanding, she was willing to take

into consideration my ideas – sometimes expressed in a confused way – and she helped me develop convincing arguments to support them. In these three years, she encouraged me to participate in various kinds of events including conferences and workshops and it is precisely on these occasions that I was able to benefit from the discussions with researchers coming from all over the world.

I am also very grateful to Carlo Cecchetto, Roland Pfau, Natasha Abner, and Caterina Donati for having perused and evaluated this work. Their sharp observations and useful suggestions guided me during the whole revision process and helped me improve both the content and the general form of this thesis.

Special thanks go to all my LIS informants without whom an important part of this work could not have been developed: Mirko Santoro, Gabriele Caia, Rosella Ottolini, and Fabio Scarpa. In particular, I thank them for their great helpfulness, willingness to collaborate, and friendship.

I am also very grateful to all the linguistics professors and researchers working in Ca' Bembo (Venezia) for their support and suggestions. My sincere thanks go to Chiara Branchini, Carmela Bertone, Michele Brunelli, Giuliana Giusti, and Guglielmo Cinque for their genuine interest in my work and valuable discussion. For my Parisian studies, I am grateful to Philippe Schlenker, Vincent Homer, and Emmanuel Chemla. I wish to thank Luca Des Dorides for having provided me with historical details concerning the Deaf community.

I would like to express my gratitude to the audiences who attended the conferences at which I presented some parts of my research work (FEAST in Warsaw and Venice, TISLR in London, SignNonmanuals in Klagenfurt, LREC in Reykjavik, NELS in Boston). They contributed to improve the quality of this thesis with their comments and questions. I also thank my friends at COST for having welcomed me in their working group and for having assigned me a fulfilling task.

I could not forget to thank my PhD colleagues at Ca' Bembo with whom I have built a cooperative and supportive relationship: Aquiles, Lorenzo, Luciana, Cho, Giorgia, Chi Fung, Camelia, Laura, Margherita, and Ilaria. Much to my surprise, when I was still at the beginning of this journey, they elected me as their representative. I hope that in this three-year mandate I have lived up to their early confidence. I also would like to thank

the staff working in Ca' Bembo, in particular Virginia, Anna, and Irma.

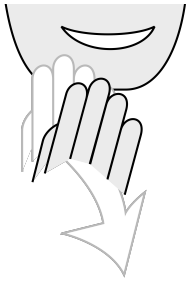
I am thankful to my friends for their moral support and true friendship: Tiziana, Amy, Valentina, Maria, Lisa, Enrico, Giorgia, and Linda. I was blessed with hugs and pats on the back when I needed them. I thank Carlo for having taught me that sometimes when the opportunity arises to board a train, you must allow ambition to overcome trepidation, and find the courage to get in.

From the bottom of my heart, I wish to thank my partner Matteo for his love, encouragement, and patience. I thank him for having tolerated my mood swings, for having pushed me to overcome obstacles, and for having made me take some breaks every now and then.

I am profoundly indebted to my family for all their love and support. I thank all those who, motivated by unconditional love, have always allowed me to look to the future with confidence and serenity. I am especially grateful to my mum for having instilled in me the tenacity to face obstacles and new challenges, as well as the humility to realize that there is always more to achieve. Without her example of humanity, I could not have become who I am.

To conclude, I wish to thank all the readers who are going to devote some time to read the pages of this thesis and retrace my journey as an *apprentice researcher*.

THANK YOU



GRAZIE

Ringraziamenti

Tre anni fa ho cominciato questo viaggio con mille punti interrogativi. Sono salita sul treno degli *apprendisti ricercatori* senza sapere dove esattamente mi avrebbe portato. Ho avuto la fortuna di intraprendere questo viaggio insieme a persone che mi hanno accompagnato, indirizzato, supportato e incoraggiato a proseguire. Ora sono quasi giunta alla fine di questo viaggio. Avendo ormai terminato la tesi di dottorato, vorrei voltarmi almeno per un attimo per ricordare e ringraziare tutte queste persone.

Il primo doveroso ringraziamento va ai miei due relatori di tesi, Carlo Geraci e Anna Cardinaletti. Senza la loro guida costante, probabilmente, il mio treno non avrebbe saputo quale direzione prendere. Li ringrazio per avermi seguito con grande professionalità durante tutto il programma dottorale e per aver arricchito il mio bagaglio di conoscenze e competenze.

Ringrazio Carlo Geraci per avermi per primo menzionato la possibilità di cominciare questa nuova esperienza e per avermi seguito da Parigi in tutte le tappe di questo lavoro. Nonostante la distanza geografica, ho sempre potuto contare su di lui. Nel suo piano di lavoro settimanale non è mai mancato il tempo per un appuntamento via Skype per fare il punto della situazione e darmi nuove indicazioni. Con instancabile dedizione è sempre stato pronto a rispondere alle mie perplessità e a correggere le mie imprecisioni. Non solo mi ha guidato attraverso le varie fasi del mio lavoro, ma mi ha anche saputo trasmettere una rigorosa metodologia scientifica e competenze di tipo tecnico. Lo ringrazio anche per aver fatto sì che per quattro mesi potessi svolgere attività di ricerca con lui e con il gruppo di Sign Language Linguistics presso l'Institut Jean-Nicod a Parigi. Questa tappa importante mi ha permesso di immergermi in un ambiente di lavoro internazionale e di apprezzare l'importanza di lavorare all'interno di un gruppo di ricerca.

Ringrazio Anna Cardinaletti per esser stata un costante punto di riferimento a Venezia. Nonostante i vari impegni accademici, si è sempre dimostrata disponibile ad

incontrarmi e a dedicarmi il suo tempo. Con pazienza e comprensione ha saputo accogliere le mie idee talvolta confuse aiutandomi a sviluppare argomentazioni convincenti per supportarle. In questi tre anni mi ha incoraggiato a partecipare a iniziative di vario tipo tra cui convegni e seminari e proprio in queste occasioni ho beneficiato del confronto con ricercatori di tutto il mondo.

Sono molto grata anche a Carlo Cecchetto, Roland Pfau, Natasha Abner e Caterina Donati per aver letto e valutato questo lavoro. Le loro attente osservazioni e i loro utili suggerimenti mi hanno guidato nella fase di revisione e mi hanno aiutato a migliorare il contenuto e la forma di questa tesi.

Un ringraziamento speciale è doveroso nei confronti degli informanti LIS senza i quali una parte importante di questo lavoro non avrebbe potuto essere sviluppata: Mirko Santoro, Gabriele Caia, Rosella Ottolini e Fabio Scarpa. Li ringrazio in particolare per la loro grande disponibilità, voglia di collaborare e amicizia.

Vorrei ringraziare anche tutti i docenti e ricercatori dell'area linguistica di Ca' Bembo (Venezia) per il loro supporto e per i loro consigli. Un ringraziamento particolare va a Chiara Branchini, Carmela Bertone, Michele Brunelli, Giuliana Giusti e Guglielmo Cinque per il loro sincero interesse nel mio lavoro e per i preziosi scambi di idee. Per i miei studi a Parigi, sono grata a Philippe Schlenker, Vincent Homer ed Emmanuel Chemla. Desidero ringraziare Luca Des Dorides per avermi fornito informazioni di carattere storico riguardanti la comunità Sorda.

Vorrei esprimere la mia gratitudine anche nei confronti di tutti coloro che hanno partecipato alle conferenze nelle quali ho presentato parti del mio lavoro di ricerca (FEAST a Varsavia e a Venezia, TISLR a Londra, SignNonmanuals a Klagenfurt, LREC a Reykjavik, NELS a Boston). Con i loro commenti e con le loro domande hanno contribuito a migliorare la qualità di questa tesi. Ringrazio anche gli amici del COST per avermi accolto nel loro gruppo di lavoro e per avermi affidato un incarico gratificante.

Non potrei dimenticare di ringraziare i miei colleghi di Ca' Foscari con i quali ho costruito un rapporto collaborativo e solidale: Aquiles, Lorenzo, Luciana, Cho, Chi Fung, Camelia, Giorgia, Laura, Margherita, Ilaria. Con mia grande sorpresa mi hanno voluta come loro rappresentante quando ancora ero all'inizio di questa avventura. Spero

con il mio impegno di aver meritato la loro fiducia. Un sentito ringraziamento va a tutto il personale che lavora a Ca' Bembo, in particolare a Virginia, Anna e Irma.

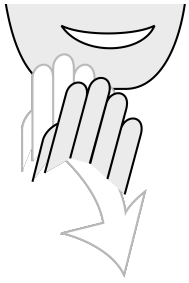
Ringrazio affettuosamente i miei amici per il loro supporto morale: Tiziana, Amy, Valentina, Mirko, Maria, Lisa, Enrico, Giorgia e Linda. Sono stata fortunata a ricevere da loro un abbraccio o una pacca sulla spalla quando ne avevo bisogno. Ringrazio anche Carlo per avermi insegnato che a volte quando un treno ti passa vicino devi lasciare che l'ambizione prevalga sulle preoccupazioni e avere il coraggio di salirci.

Un grazie dal profondo del mio cuore spetta al mio compagno Matteo per il suo amore, per l'incoraggiamento e la pazienza. Lo ringrazio per aver sopportato i miei sbalzi d'umore, per avermi incoraggiato a superare le difficoltà e anche per avermi fatto staccare la spina di tanto in tanto.

Sono profondamente riconoscente alla mia famiglia per tutto il loro amore e sostegno. Ringrazio tutti coloro che, mossi da affetto incondizionato nei miei confronti, mi hanno sempre permesso di guardare avanti con fiducia e serenità. Sono grata specialmente a mia mamma per avermi trasmesso la tenacia per affrontare ostacoli e nuove sfide e anche l'umiltà di non sentirsi mai arrivati. Senza il suo esempio di umanità non potrei essere quella che sono.

Infine, desidero ringraziare di cuore tutti i lettori che vorranno dedicare il loro tempo per avventurarsi tra le pagine di questo lavoro e ripercorrere questo mio viaggio come *apprendista ricercatrice*.

THANK YOU



GRAZIE

Table of contents

Annotation conventions

CITY	ID-Gloss identifying a sign
CITY+++	Reduplicated sign
M-A-R-I-A	Fingerspelling
LOOK-FOR	Two or more English words corresponding to one single sign
$SIGN_i$	Coreferential index
IX	Pointing sign (generic)
IX-1	Pointing sign functioning as personal pronoun
IX_DEM	Pointing sign functioning as demonstrative
IX_POSS	Pointing sign functioning as possessive
IX_LOC	Pointing sign functioning as locative
CL	Classifier sign
$_1CRITICIZE_2$	Agreement verb
$\overline{\text{SIGN}}^{\text{indef}}$	Non-manual marking
SIGN, SIGN	Prosodic pause
(you)	Understood item
?	Dubious acceptability judgment
*	Judgment of unacceptability

List of abbreviations

ABSL	Al-Sayyid Bedouin Sign Language
ASL	American Sign Language
HKSL	Hong Kong Sign Language
LIBRAS	Brazilian Sign Language
LIS	Italian Sign Language
LSA	Argentine Sign Language
LSF	French Sign Language
NSL	Nicaraguan Sign Language
PJM	Polish Sign Language
TSL	Taiwan Sign Language
AgrP	Agreement Phrase
AP	Adjective Phrase
CP	Complementizer Phrase
DemP	Demonstrative Phrase
DP	Determiner Phrase
MP	Measure Phrase
NegP	Negation Phrase
NMM	Non-manual marker
NP	Noun Phrase
NumP/CardP	Cardinal numeral Phrase
QP	Quantifier Phrase
QuantAP	Quantitative Adjective Phrase
TP	Tense Phrase
WALS	World Atlas of Language Structures

Introduction

The languages of the world differ from each other in various respects (i.e. phonological inventory, morphological strategies, syntactic constructions, etc.). Despite the considerable variety of different linguistic features, there are principles that are claimed to hold for all languages and to constitute a common linguistic basis. These properties shared by all languages are known as language universals. In the literature of linguistics, these universal features have been looked for and accounted for by looking mainly at spoken languages. Only scarce attention has been payed (until very recently) to sign languages.

From a theoretical perspective, an interesting research ground where language universals can be tested is represented by sign languages which are full-fledged languages expressed and perceived in the gestural-visual modality.[1] On the one hand, sign languages represent an intriguing empirical domain where the tenability of the universals claimed to be valid for spoken languages can be examined. If universals really hold for all human languages, then it should be possible to detect them both in spoken and signed languages regardless of differences due to modality. On the other hand, sing languages also represent a challenging research domain. This is motivated by the fact that they usually display a considerable degree of intralinguistic variation (a.o. Lucas, Bayley & Valli, 2001). In other words, different forms conveying the same linguistic features may coexist within one and the same language.

Internal variation in sign languages may depend on several sociolinguistic factors. One of the main driving factors in the observed variation is represented by the fact that sign languages are used in small communities whose users are embedded in a hearing majority that uses spoken languages. The linguistic implication of this social environment is that sign languages undergo the constant pressure of the dominant spoken languages that are often regarded as more prestigious. The effects of this

1 For a collection of contributions addressing a wide range of topics in sign language linguistics, the reader is referred to Pfau, Steinbach, and Woll, eds. (2012).

linguistic influence can be detected in several aspects such as the use of fingerspelling, the production of mouthings, and also word order. In some countries, sign languages have not yet been formally recognized by the official authorities. The lack of concern shown by governments may result in the lack of public television's programs in sign language, scarce financial resources supporting bilingual programs (i.e. education programs designed to support the acquisition of the local sign language and spoken language), and other activities accessible to Deaf[2] people. Another factor that contributes to the internal variation of sign languages consists in atypical patterns of language acquisition and transmission. Given that only 5-10% of Deaf children are born to Deaf parents, the majority of signers do not get exposed to signs until late childhood or early adulthood. Moreover, the paucity of bilingual programs involuntary contributes to delay the age of first exposure to sign languages. Signing communities include signers with different levels of language competence, thus resulting internally variegated. Consequently, the fact that intralinguistic variation is attested is not surprising at all. Furthermore, sign languages are languages transmitted through face-to-face interaction since they do not have a widely recognized written form. Although several writing systems have been proposed so far, none of these systems have been adopted by signing communities. Because of the absence of an official written form, the diffusion of standardized forms is somehow more complicated.[3]

All these sociolinguistic factors contribute to the intralinguistic variation attested in sign languages and to the consequent coexistence of different forms. Given this scenario, the main challenge consists in understanding whether this language-internal variation observes, or rather violates, the universals of language. In principle, the linguistic forms violating the universal principles of language structure should not be selected by signers. This is the theoretical research question that this thesis will address.

In syntax, internal variation emerges as word order flexibility. From a theoretical point of view, this aspect of sign languages is sometimes problematic to capture and to explain. The main risk is that order flexibility attested in sign languages may obscure

2 In this thesis, I adopt the standard convention of using capitalized 'Deaf' to indicate deaf people who identify themselves members of a signing community.

3 Notice that internal variation is not attested to the same extent in all sign languages. This is due to the fact that each of them is used in different sociopolitical contexts. Besides, independently of these sociolinguistic factors characterizing sign languages, internal variation is also attested in spoken languages.

underlying universal patterns leading to the conclusion that sign languages are unclassifiable or, even worse, do not adhere to language universals. Thus, it would be advisable to check whether sign order flexibility is a random phenomenon, or is rather constrained by independent motivations.

This dissertation aims at stimulating the debate on the applicability of language universals to sign languages by considering this issue from various theoretical perspectives. This will be done by looking at Italian Sign Language (hereafter LIS) and specifically at its nominal domain. The first empirical research question is: to what extent is Universal 20 (Greenberg, 1963) in its recent formal revisitation (Cinque, 2005) also valid for LIS? Examining the distribution of nominal modifiers with respect to the noun, I will quest for the typological pattern characterizing LIS nominal expressions.

The choice of studying the nominal domain is motivated by the fact that it is not entirely clear to what extent this research area is affected by language-internal variation. Although some recent studies have already shown a considerable degree of variation in the prosody, lexicon, and syntax of wh-questions and other linguistic domains of LIS (Cardinaletti, Cecchetto & Donati, eds., 2011; Conte, Santoro, Geraci & Cardinaletti, 2010; Geraci, Battaglia, Cardinaletti, Cecchetto, Donati, Giudice & Mereghetti, 2011; Branchini, Cardinaletti, Cecchetto, Donati & Geraci, 2013), no systematic investigation has taken into consideration variation issues in the syntax of nominal expressions. On the contrary, the existing literature (Branchini, 2007; Bertone, 2007, 2009; Brunelli, 2011) offers a stable picture in which the noun and its modifiers seem to be distributed according to rather unvarying patterns. Given the expectations on word order flexibility inspired by a considerable portion of literature on sign language linguistics, the nominal domain appears as an intriguing, yet controversial, research area. The second empirical research question that will be addressed concerns the range of variation characterizing the syntax of nominal expressions in LIS. In particular, if variation is observed, I intend to find out the driving linguistic and sociolinguistic factors. From a generative perspective, I will also discuss how these observations can be accounted for in a formal proposal. Besides, making a connection between the theoretical and empirical research questions, I will check whether possible different word order patterns in LIS still adhere to universal generalizations fulfilling their predictions.

Along with these research goals, this dissertation will show how different types of data and different methodological strategies can be combined together in order to obtain deeper insights on the issues under investigation. This innovative approach will be adopted in the study of nominal modification and specifically in the analysis on the distribution of cardinal numerals, namely a class of modifiers showing a particularly intricate picture.

Conceptually, this dissertation is divided into three parts: a theoretical, a methodological, and an empirical one. The first part corresponding to the first two chapters presents the main theoretical ingredients needed to set the ground for the presentation of the research work developed in the following chapters. The second part corresponding to the third chapter illustrates the methodological issues that have been considered during the various stages of this work. Finally, the third part including the last two chapters presents the empirical studies that have been conducted on the syntax of nominal expressions in LIS.

The remainder of the thesis is organized as follows.

Chapter 1 offers a general introduction on three theoretical frameworks, namely linguistic typology, generative linguistics, and sociolinguistics. These provide the foundational theories upon which this research work is based. To illustrate each framework, I will discuss what are the main theoretical assumptions, how investigations are usually conducted, and what has been discussed with respect to word order issues so far. Importantly, this chapter emphasizes the contribution that the study on sign languages may give to the three disciplines.

Chapter 2 gives a brief overview of the most significant literature dealing with nominal modification. In particular, the explicatory adequacy of the formal derivations and typological developments connected to Greenberg's (1963) Universal 20 will be critically discussed and evaluated. Then, the fine-grained structure of nominal expressions will be illustrated showing that it appears more complex than what emerged from early studies in this research domain. Finally, data coming from different sign languages will show the considerable degree of order flexibility characterizing the distribution of nominal modifiers in different sign languages. Contrary to this scenario, previous studies on the nominal domain in LIS have shown the existence of rigid

patterns without addressing intralinguistic variation in detail.

Chapter 3 presents the methodology that has been adopted for the empirical studies described in this thesis. Interestingly, different types of data (i.e. corpus data and elicited data) and different procedures have been employed in order to obtain a more complete view on the issues under investigation. Under this multi-strategy approach, the chapter will explain how qualitative and quantitative data have been used and combined together. From a methodological perspective, the main stages involved in the empirical studies conducted for this thesis will be described in detail, from data annotation to data analysis.

Chapter 4 introduces the reader to the two quantitative studies that have been conducted on the basis of corpus data. The first study will show how nominal modifiers are distributed with respect to the noun. In the second study, the duration of modifiers will be investigated. Crucially, the findings resulting from the two analyses will show that the intralinguistic variation characterizing nominal modification in LIS is influenced both by linguistic and socio-linguistic factors. Moreover, despite the apparent unrelatedness, the two studies will reveal interesting interconnections.

Chapter 5 is aimed at showing how it is possible to investigate the syntactic behavior characterizing the modifiers populating the nominal domain. To illustrate how this type of investigation can be conducted, an in-depth study on the syntax of cardinal numerals will be presented. This understudied research topic will show unexpected variation in word order. Considering different sources of data, namely corpus data and elicited data, will help clarify what triggers this variation. Also, a special construction containing cardinals, namely Measure Phrases, will be presented and discussed. Finally, a syntactic analysis explaining the data and accounting for word order issues involving cardinals will be provided.

1. Theoretical framework

1.1. Introduction

This thesis aims at investigating nominal modification on the basis of a multidisciplinary approach combining linguistic typology, generative linguistics, and sociolinguistics. The main advantage of considering various approaches to language facts is that the research topic can be observed from many different perspectives, and a comprehensive account can be offered. This chapter provides the basis for a better understanding of the theoretical and empirical work developed in the next chapters.

In this thesis, three ideal linguists work together on the same issue, namely the distribution of nominal modifiers in LIS. In particular, the three linguists are a typologist, a generativist, and a sociolinguist. The three of them follow their own scientific principles and procedures, but they occasionally interact with each other pursuing some interdisciplinary investigations.

The typologist observes data coming from a considerable number of languages and tries to classify them into different language types. Through crosslinguistic studies, typologists aim at identifying generalizations and universal principles governing human language. Our ideal typological linguist is interested in studying how LIS behaves with respect to nominal modification with the purpose of enlarging the empirical base for his or her observations.

The primary objective of the generativist is to observe language data and infer the speakers'[1] knowledge of the rules governing that language. On the basis of the relevant nominal constructions in LIS, our ideal generative linguist is interested in making the underlying mechanisms explicit. Afterwards, s/he needs to represent the inferred structural configurations into a model for descriptive purposes.

What the sociolinguist does is explore the way society affects language. The idea is that

1 Under the label 'speakers', both speakers and signers are meant, unless differently specified.

language cannot be abstracted from its social context. Language variation is one of the main research topics in this field and is generally conceived as "two ways of saying the same thing" (Labov, 1972: 271). So, our ideal sociolinguist is interested in exploring how nominal modification in LIS varies according to both linguistic and extra-linguistic factors (i.e. factors such as age, gender, social class, etc.).

This chapter is not intended to be a comprehensive review of these three theoretical frameworks because each of them is connected to a vast bibliography. For reasons of time and space, I present the most relevant works that will be useful to contextualize the ideas and topics of the following chapters. For each framework, I provide a quick summary of the main tenets and explain why sign languages are relevant for it. Section 1.2 deals with linguistic typology. In Section 1.3 generative linguistics is introduced. Section 1.4 explains how sociolinguistics accounts for language variation.

The typologist, the generativist, and the sociolinguist usually work independently. What happens if they work jointly on a common linguistic issue will be shown in the following chapters. Language-internal variation in the nominal domain in LIS is the issue that needs to be explained. Variation may be motivated by sociolinguistic factors, stylistic choices, or other factors. Crucially, what needs to be evaluated is whether the attested word order permutations respect or violate what the typologist observes and the generativist explains.

1.2. Linguistic typology

In this section, the typological framework is presented. I explain what typologists are mostly interested in and how they conduct their research. In particular, for the purposes of this thesis, this section is focused on the notion of universal. Also, it is discussed how word order issues are questioned from a typological perspective.

1.2.1. General framework

Generally speaking, a typology is considered a classification of phenomena according to their characteristics. This type of research, aiming at taxonomies, is quite widespread

crossdisciplinarily since it finds application in several domains from hard sciences to humanities (a.o. anthropology, archaeology, theology, architecture).

In linguistics, typological studies aim at classifying languages according to common formal features. Typologists usually identify clusters of linguistic properties and then verify to what extent languages vary according to these properties. They conduct crosslinguistic investigations to have a better understanding of the diversity of the languages of the world. Typologists are not only interested in diversity, but also in commonalities. Indeed, they try to detect those properties that hold for all languages, and hence can be regarded as universal properties. The empirical basis for typological studies is typically represented by a large set of languages, even "exotic" and understudied languages (Daniel, 2011).

Another goal of typological studies is to understand the direction of the diachronic evolution of languages. Observing the properties characterizing the various language families, it is possible to identify language changes that happened in the past or even to predict the changes that might happen in the future. For example, Haspelmath (1990) focuses on the passive construction in a crosslinguistic study and identifies a possible evolutionary path that leads from passive to ergative case-marking.[2]

The founder of modern typology is Joseph Greenberg. In his seminal work (1963), he points out that language variation is not arbitrary, as previously assumed. He asserts that it is rather constrained by universal principles, although the nature of the term 'universal' may be quite different from that used by generativists.

However, typological studies are not limited at classifying languages according to their properties. Another goal of these studies is to understand the potentialities, preferences, and also limits of human language. Indeed, the comparative study on a sample of languages provides insights not only on the range of variation but also on the constraints regulating such variation.

Some empirical questions tackled by typologists concern the possible realizations of

2 According to the passive-to-ergative hypothesis, in some languages passive is more and more generalized and gradually develops as an unmarked construction. The next evolutionary step is a reanalysis of the argument structure: the patient of the passive construction is marked by absolutive case and the agent of the passive construction is marked by ergative case. The diachronic change is claimed to be unidirectional since there is no evidence supporting the shift from ergative to passive (Haspelmath, 1990).

given linguistic phenomena, namely the attested forms. With respect to the unattested forms, special attention should be paid in order to evaluate whether these forms are really ungrammatical constructions or rather options that accidentally have not been produced. “Being unattested clearly does not imply being impossible, and when something is possible this does not mean that it has to be attested” (Cysouw, 2010: 256).

As previously mentioned, typological linguistics clusters languages into types according to shared properties. The generalizations that potentially hold for all world’s languages are called universals.

From a structural point of view, universals can be divided into three categories: absolute universals, statistical regularities, and implicational universals (Baker & McCloskey, 2007). An example of each is provided in (1).

(1) a. Absolute universals: all languages have demonstratives

b. Statistical regularities: in the vast majority of languages, the subject precedes the object

c. Implicational universals: if a language has fricative phonemes, then it will also have stop phonemes

Absolute universals hold for every language and are built on the format “every language/no language has property X”. By definition, they cannot be falsified by any exceptions. This category of universals does not contain many elements and usually describes very general linguistic properties. Statistical regularities are not defined as strict universals because they describe general trends and admit exceptions. These tendencies apply to the format “languages with property X are common/uncommon”. In implicational universals (or restricted universals), the presence of a feature implies the presence of another feature. This type of universals, originally discussed for syntax by Greenberg (1963), follows the format “if a language has property X, then it also has property Y”. So, implicational universals capture the correlation between two distinct properties.

It is important to note that implicational universals can be either bidirectional or

unidirectional. They are bidirectional if they function in both directions, thus showing the following configuration: “if a language has property X, then it also has property Y and if it has property Y, then it also has property X” [X → Y] & [Y → X]. They are unidirectional if they function in one direction only [X → Y]. Both cases are exemplified in (2).

(2) a. Unidirectional implicational universals: if a language has nasal vowels, then that language has also non-nasal vowels. The contrary (if a language has non-nasal vowels, then that language also has nasal vowels) is not true because, for instance, Italian does not have nasal vowels

b. Bidirectional implicational universals: if a language has the order Verb>Object, then it tends to have the order Prep>Noun, and if a language has the order Prep>Noun, then it tends to have the order Verb>Object

The study on the universals characterizing human language is common in both Greenberg’s typological tradition and Chomsky’s 'Generative Grammar' (Croft, 1990). Both approaches focus on what is possible in human language. They assume that it is possible to abstract a number of linguistic generalizations from data. These abstractions are not language-specific.

Croft (1990) shows that, although the two approaches share some assumptions, they also diverge. The universal principles investigated by generative linguists (cf. Section 1.3.1) reflect a more rationalist approach to language. What 'universal' refers to is the common biological endowment that enables every human being exposed to a limited but sufficiently rich input to acquire a full-fledged linguistic system ('poverty of the stimulus argument', see Chomsky, 1980). The study of a single language may potentially provide some insights on the constraints applied by this universal linguistic competence. On the other hand, the typological perspective on 'universal' refers to abstract patterns that spread across languages. Hence, the study of a single language or even of a small sample of languages would not suffice to grasp these universal properties. According to typologists, language universals can be extrapolated only if a sufficiently large and widely distributed sample of languages is investigated. The

research methodology used by typological linguists to capture generalizations and to abstract universals from data is basically empirical and inductive.

According to Cinque (2007), the two fields are not radically different in nature and do not exclude each other. Focusing on a case study on relative clauses, he demonstrates how a formal approach can contribute to typology and vice versa. In the same line of reasoning, this dissertation shows how phenomena ascribed to crosslinguistic variation or even intralinguistic variation can be captured and explained by abstract formalizations giving support to the idea that linguistic typology and formal linguistic theory are not irreconcilable approaches.

1.2.2. Working in the framework

When conducting their investigations, typologists usually make use of large databases. Early works by Greenberg (1963) included 30 languages carefully chosen (every continent was represented), so to have a representative sample. Traditionally, sample collection involves an accurate selection of the languages on the basis of their geographical origin and genetic relationship. So, in order to obtain a geographically and genetically balanced sample, typologists tend to exclude languages sharing similar or identical linguistic properties. Dryer's (2011) approach is however different in that he does not exclude any language. This author clusters languages into genera (classes of languages that are historically related to each other) and finally considers five macro-areas roughly corresponding to the five continents (Africa, Australia and Papua New Guinea, Eurasia, North America and South America, Southeast Asia, and Oceania).

Probably, the currently best known database is the World Atlas of Language Structures (WALS). This is a comprehensive typological database, available both in print and online versions (Dryer & Haspelmath, eds., 2013). Its content comes from the contributions of 55 typologists. WALS is structured in 144 chapters concerning different language aspects (phonology, morphology, nominal categories, nominal syntax, verbal categories, word order, simple and complex clauses, lexicon) and a few extra chapters. It contains data from 2600 languages, although the core sample consists of approximately 200 languages.

To have a better grasp on how typologists usually conduct their investigations, let us

consider, for instance, a research topic of potential interest, namely polar questions. First of all, they need to know how different languages express this particular type of syntactic construction. What they do is conduct comparative research on the basis of a more or less large sample of languages. To conduct crosslinguistic surveys, typologists may consult existing reference grammars, WALS resources, or available corpora. They may also get in contact with native-speaker consultants and elicit data from them. A quicker method for data collection consists in distributing to consultants or other researchers questionnaires specifically conceived for the research topic under investigation. Turning back to the example on polar questions, typologists analyze the data at their disposal and try to identify the possible strategies that different languages employ to mark this particular construction. The resulting typology would include at least the types listed in (3) (data taken from WALS, see Dryer, 2013b).

(3) a. Interrogative particle: 585 languages (e.g. Turkish, Albanian)

b. Interrogative intonation: 173 languages (e.g. Italian, Kawaiisu)

c. Interrogative affix on the verb: 164 languages (e.g. Korean, Pawaian)

d. Interrogative word order: 13 languages (e.g. English, Warekena)

WALS allows users to visualize typological classifications on a map that can be graphically tailored according to the researcher's preferences. Figure 1 shows where the four strategies listed in (3) are spread among the languages of the world.

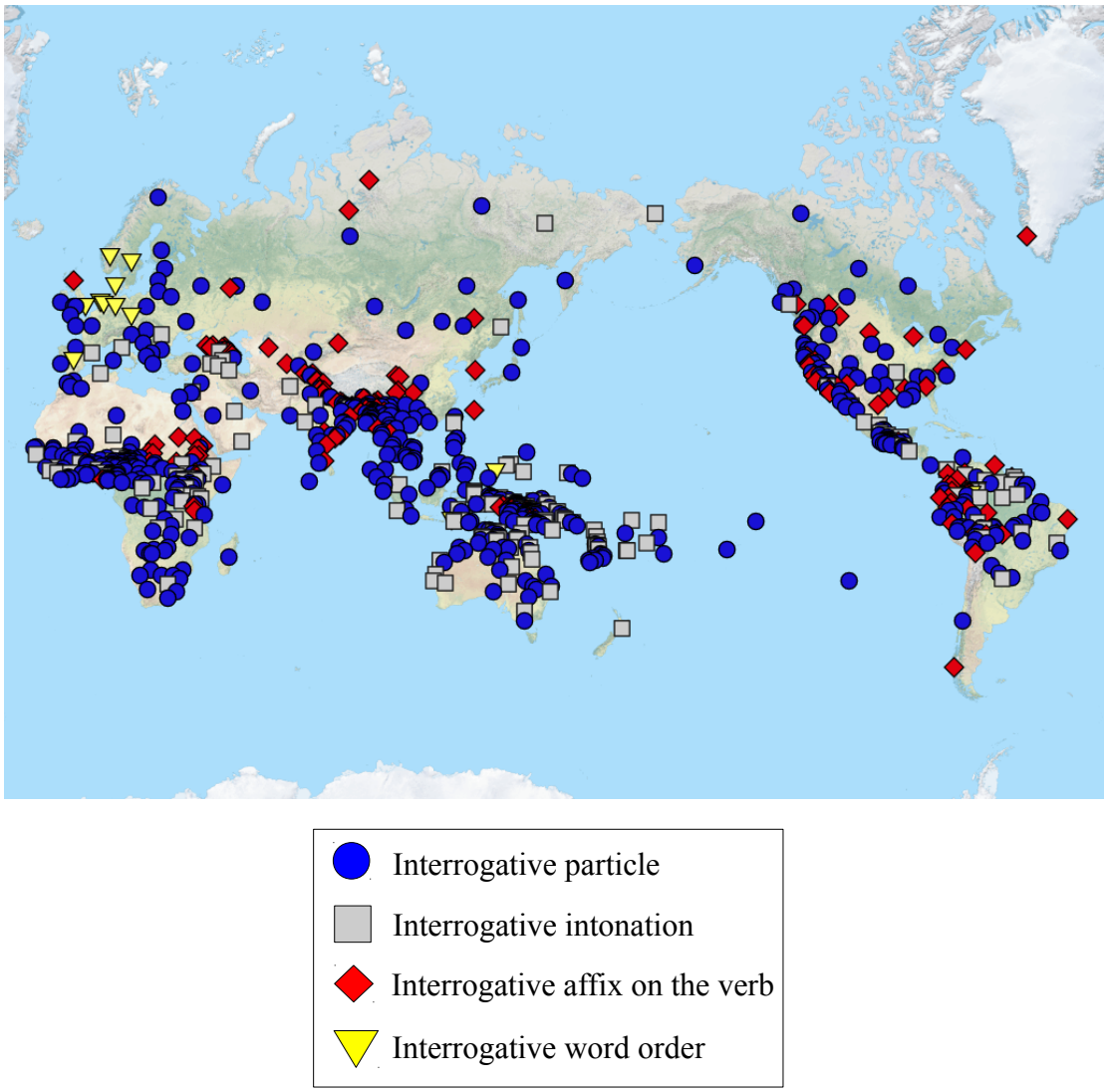


Figure 1: Geographical distribution of the strategies marking polar questions

This visual approach helps understand how linguistic features are geographically distributed and identify similarities in geographical contiguous clusters.

1.2.3. Major findings in word order

One of the main research areas for typologists is word order. By applying the typically typologist comparative methodology, word order studies focus on the patterns emerging at various levels (clausal, nominal domain, etc.) with the purpose of appreciating the similarities and differences among world's languages.

A popular research topic among word order typologists is the basic word order, namely the order in which subject (S), object (O), and verb (V) occur in an unmarked declarative sentence. The six logically possible combinations are: SOV (English equivalent: 'my cousins geography teach'), SVO (English equivalent: 'my cousins teach geography'), VSO (English equivalent: 'teach my cousins geography'), VOS (English equivalent: 'teach geography my cousins'), OVS (English equivalent: 'geography teach my cousins'), OSV (English equivalent: 'geography my cousins teach'). According to Dryer (2013a), these six combinations are all attested, as shown in (4). However, VOS, OVS, and OSV orders are extremely rare options since they are found in only 4% of Dryer's sample of languages. Interestingly, in the three most common orders, the subject precedes the object (cf. statistical regularity in (1)b) and in the three least common orders the object precedes the subject.

(4) a. SOV (Japanese, Tamil, Georgian, etc.): 565 languages (= 48%)

b. SVO (English, Macedonian, Zulu, etc.): 488 languages (= 41%)

c. VSO (Arabic, Welsh, Maori, etc.): 95 languages (= 8%)

d. VOS (Kiribati, Nias, Malagasy, etc.): 25 languages (= 2%)

e. OVS (Cubeo, Selknam, Urarina, etc.): 11 languages (= 1%)

f. OSV (Kxoe, Nadëb, Tobati and Wik Ngathana): 4 languages (= < 1%)

Before Greenberg's paper published in 1963, the various linguistic phenomena were mostly investigated as independent aspects. After Greenberg's (1963) influential work, the one that started up modern typology, a new approach arose: linguistic phenomena are interconnected. Therefore, knowing the basic word order of the sentence it is possible to make some predictions on the patterns characterizing other domains. Greenberg (1963) discovered many correlations between the basic word order at clause level and the relationship between heads and modifiers (e.g. noun and adposition, noun and genitive, noun and adjective). One of these correlations is contained in Greenberg's Universal 16 and is reported in (5).

(5) “In languages with dominant order VSO, an inflected auxiliary always precedes the main verb. In languages with dominant order SOV, an inflected auxiliary always follows the main verb” (Universal 16, Greenberg 1963: 85)

The pattern illustrated in (5) is confirmed by the two pieces of evidence below. Examples in (6) are taken from a VSO language like Arabic (Cheng & Corver, 2013: 316-319), and examples in (7) are taken from an SOV language like Korean (Bybee & Fleischman, 1995: 170).

(6) a. Basic word order in Arabic: VSO

qaraʔ-at al-banaat-u al-kitaab-a

read-PERF-3.FEM the-girls-NOM the-book-ACC

'the girls read the book'

b. Auxiliary/main verb in Arabic: Aux + V

kaan-at talʕabu al-banaat-u

be-PERF-3.FEM play-IMPERF-3.FEM the-girls-NOM

'the girls were playing'

(7) a. Basic word order in Korean: SOV

Younghi-ka Seoul-ul ttena-ss-e

Younghi-SUBJ Seoul-OBJ leave-PAST-SE

'Younghi left Seoul'

b. Auxiliary/main verb in Korean: V + Aux

Younghi-ka Seoul-ey ka-ya/to tway-e

Younghi-SUBJ Seoul-to go-CONN AUX-SE

'Younghi must/can go to Seoul'

Observing the various correlations between basic word order and head-modifier order, a dichotomy emerges. This consists of two ideal cases, namely head-initial languages and head-final languages. As illustrated in (8), these two poles show opposite word order patterns (Croft, 2003: 72, see also Cinque, 2013b). The crucial generalization is that in head-initial languages modifiers follow the respective heads and that in head-final languages modifiers precede the respective heads.

(8) a. Head-initial languages: verb>object, verb>subject, auxiliary>verb, verb>adverb, subordinator>verb, verb>purposive, qualifier>sentence, preposition>noun, noun>genitive, noun>relative clause, noun>adjective, noun>demonstrative, noun>numeral, adjective>adverb

b. Head-final languages: object>verb, subject>verb, verb>auxiliary, adverb>verb, verb>subordinator, purposive>verb, sentence>qualifier, noun>postposition, genitive>noun, relative clause>noun, adjective>noun, demonstrative>noun, numeral>noun, adverb>adjective

According to the principle of crosscategorial harmony (Hawkins, 1983), languages tend to show the same pattern across different syntactic domains. Nonetheless, the languages that consistently conform to either one of the two ideal cases outlined in (8) are not many. Apparently, even the most polarized languages have cases that contradict the ideal paradigm. The majority of languages departs from this consistency to different degrees. English is one of these: even if it is closer to head-initial languages in some aspects, it is rather a mixed language. The patterns in (9) show this aspect on English word order.

(9) a. Head-initial patterns: verb>object (e.g. sing a song), preposition>noun (e.g. on the table)

b. Head-final patterns: adjective>noun (e.g. red car), demonstrative>noun (e.g. this girl)

1.2.4. The relevance of sign languages for linguistic typology

Typological studies exhibit an impulse toward the examination of different languages, even exotic ones. However, as discussed in Zeshan (2006), sign languages have been barely considered in typological works.

The study of languages expressed in a different modality, namely the visual-gestural modality, allows typologists to take an even broader view on the diversity of human languages. Interestingly, some properties characterizing sign languages may emerge as typologically unusual. For example, Brazilian Sign Language (LIBRAS) employs a wh-sign (WHAT/WHO) that can mean either 'what' or 'who' depending on the context, as shown in (10) and in (11), respectively (Zeshan, ed., 2006: 273-274). Notice that the straight line above the glosses indicates the presence of a specific prosodic contour mainly due to grammatical facial expressions.

(10) WHAT/WHO J BUY
wh

'What did John buy?'

(11) BALL WHAT/WHO-IX GET
top wh

'As for the ball, who gets it?'

The fact of having one interrogative element covering the meaning of both 'who' and 'what' is unexpected from the typological point of view. According to Zeshan, "these two interrogatives are often regarded as the most basic ones in a paradigm of question words, so it is unusual for a language not to distinguish between the two" (Zeshan, ed., 2006: 22).

The study of the properties characterizing sign languages is important in that it contributes to enrich the empirical basis for typologists' generalizations. By definition, universals hold for all languages: this means that they are valid across languages

regardless of their modality. This assumption implies that universals should be verified not only with respect to spoken languages but also to sign languages. So, only inter-modal variation can reveal if a universal classification really holds for all languages, be they spoken or signed.

Crossmodal comparisons are particularly fruitful in the syntactic domain, rather than in the phonological or morphological domain. The reason is that syntactic phenomena offer a high level of abstraction that help circumvent the constraints imposed by modality.

Many studies (a.o. Meier, Cormier & Quinto-Pozos, eds., 2002; Aronoff, Meir & Sandler, 2005; Barberà, 2012) have shown the effects of modality on the use of iconicity, space, simultaneous morphology, etc. Such characteristics are often assumed to form the 'sign language type', a substratum shared by all world's sign languages. It would be interesting to investigate whether there are sign languages showing data that diverge from the assumed common basis. The study of genetically unrelated sign languages from a typological perspective may contribute either to reinforce or to deconstruct the hypothesis of common modularity-driven properties.

1.3. Generative linguistics

In this section, the basic facts of the generativist framework are presented. The generativist approach to the study of language facts is addressed from both a theoretical and methodological perspective. With respect to linearization issues, I discuss Kayne's (1994) Antisymmetry and Cinque and Rizzi's (2008) Cartographic Project. I do not discuss the minimalist approach (Chomsky, 1995) in detail, even though in the next chapter I will briefly refer to Abels and Neeleman's (2009) work on Universal 20, which has been developed in the light of minimalist principles.

1.3.1. General framework

The human language faculty allows speakers to produce and understand an infinite number of sentences. New sentences that have not been perceived before could be

understood or uttered at any time. The sentence in (12) has not been heard nor read presumably by any English speaker, however it could be well understood by any of them.

(12) In one of my past lives, I dreamt I was a furry dinosaur with five small horns on each paw.

This ability, commonly referred to as creativity, is an argument used to demonstrate that language can be acquired neither by imitating other people's utterances nor by formal teaching.

Another property of human language that shows its unlimited potentiality is recursion. Indeed, any sentence could be implemented with new linguistic material. As demonstrated in (13), a sentence can be embedded in another sentence, and this procedure could be iterated an infinite number of times. Therefore, there is potentially no limit to sentence creation.

(13) a. Television is addictive

b. The New York Times reported that television is addictive

c. My sister told my mum that the New York Times reported that television is addictive

d. I remember distinctly that my sister told my mum that the New York Times reported that television is addictive

e. My husband doesn't believe that I remember distinctly that my sister told my mum that the New York Times reported that television is addictive

f. I suspect that my husband doesn't believe that I remember distinctly that my sister told my mum that the New York Times reported that television is addictive

The idea that human language is characterized by a pervasive 'open-endedness' has been proposed by Chomsky (1957) and it is a central idea in generative linguistics. Crucially, the possibility to produce and comprehend an infinite number of sentences and the potential production of sentences of any length do not result in an anarchic system where words can be combined at random. Rather, sentences can be generated only by obeying a set of grammatical rules. These rules are subconsciously used in everyday life and in most cases it is hard to explicitly retrieve them in our minds.

Chomsky was the first who gave the name to the implicit knowledge of a language that is shared homogeneously among a linguistic community. He called it 'competence' (or 'I-language', internalized language) and contrasted it with the concept of 'performance' (or 'E-language', externalized language). The definitions of these two concepts are explained in Chomsky's quotation in (14).

(14) "We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of the language) and performance, the actual use of language in concrete situations" (Chomsky, 1965: 4)

The description of the finite set of rules characterizing a speaker's competence is what generative linguistics aims at. So, generative linguists analyze linguistic expressions in order to abstract the underlying rules.

The nature and the acquisition of these rules have been discussed within the Principles and Parameters framework (Chomsky, 1981, 1986). Looking at the rules governing language, an important distinction emerges. Some rules hold for all the languages of the world indistinctly ('principles'); some of them are language-specific and vary crosslinguistically ('parameters'). In other words, principles explain general ideas on the structure and organization of human language. Parameters explain how individual languages vary from each other.

The distinction between principles and parameters has some important consequences with respect to language acquisition. Principles do not need to be acquired, they are universal and are part of our innate linguistic endowment ('Universal Grammar'). One

well-known principle, shared by all human beings with typical language acquisition, is the principle of structure dependency. Its definition is reported in (15). It refers to the fact that the knowledge of a language is built on structural relations rather than on the linear ordering of words.

(15) Structure dependency

> "All known formal operations in the grammar of English, or of any other language, are structure-dependent" (Chomsky 1971: 30)

Parameters are influenced by the immediate linguistic environment and are fixed in the early stages of language acquisition. One well-known parameter is the one regarding null subjects. It is defined in (16).

(16) Pro-drop (or null subject)

> The pro-drop parameter "determines whether the subject of a clause can be suppressed" (Chomsky 1988: 64)

So, if a child is acquiring English, s/he will set the [-] value since English is a language that does not allow for null subjects. Conversely, a child acquiring Italian (i.e. a language with null pronouns) will set the [+] value.

In the Principles-and-Parameters framework, the architecture of language is generally represented through the so-called Y-model. The levels of representation included in this model are illustrated in Figure 2.[3]

3 Minimalism (Chomsky, 1995) has challenged the Y-model by removing both D-structure and S-structure and substituting them with two structural operations, namely Merge and Move. In this theoretical framework, the levels of representation get reduced to three: Numeration (the set of lexical items), Phonological Form, and Logical Form.

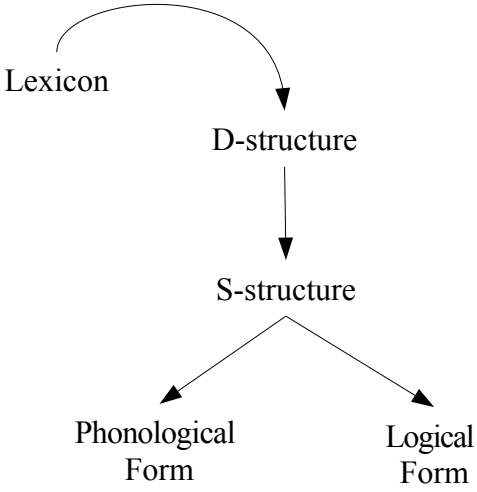


Figure 2: Y-model

The levels composing the Y-model are exemplified by providing the derivation of the sentence in (17).

(17) Mary has met the teacher

The input for this syntactic model consists of the lexical items which provide the system with information about their syntactic category. Lexical entries include lexical elements (e.g. nouns, verbs, etc.), and functional elements (e.g. determiners, auxiliaries, etc.). Each lexical entry is stored in the lexicon with its categorial specification and its selectional properties. To exemplify this notion, the speaker's mental lexicon includes the items listed in (18).

(18) Lexical entries

the	→ determiner
Mary, teacher	→ noun
meet	→ verb
have	→ auxiliary

These atoms are then merged in the deep structure ('D-structure') in which both argument structure and thematic relations are coded. For example, the verb 'meet' is a two-place predicate because it selects two arguments, as shown in (19).

(19) Argument structure in D-structure

'meet': verb; 1 2

Noun Phrase Noun Phrase

These two arguments are assigned two different theta-roles: the verb 'meet' assigns the role of agent to the argument in subject position and the role of patient to the argument in object position. In order to fulfill the selection requirements, elements get combined together through the operation of Merge. First, the verb ('meet') merges with its object ('the teacher') creating the first part of the verb phrase in (20)a. Then, the verb phrase merges with its subject ('Mary') creating the entire verb phrase as in (20)b.[4]

(20) a. {{met} {the teacher}}

b. {{{met} {the teacher}} {Mary}}

Before continuing the derivation, let me focus on how the structural configuration in which words are merged is captured, namely the X-bar theory (for a detailed overview see Fukui, 2001). This model was originally proposed by Chomsky (1970) and has been further refined by Jackendoff (1977) and many others. According to this proposal, X is a variable that can represent various syntactic elements (lexical and functional elements). One of the core ideas of X-bar theory is that all linguistic categories conform to one and the same syntactic structure. This is generally referred to as 'Uniformity Condition'. In other words, nouns, verbs, prepositions, and adjectives share the same deep phrase structure. The X-bar structure is endocentric because it is organized around a head. Specifically, the head of the phrase projects its category to the two upper levels

4 For the sake of brevity, I am not considering vP and DP. For the purposes of the argumentation, these notions are not relevant here.

('Projection Principle'), namely to the intermediate projection (X') and to the maximal projection (XP). The phrase structure can be enriched by two other elements. A complement (ZP) can combine with the head X° to form the X' level. A specifier (YP) can combine with the intermediate projection X' to form the XP level. The X-bar format is represented in Figure 3.

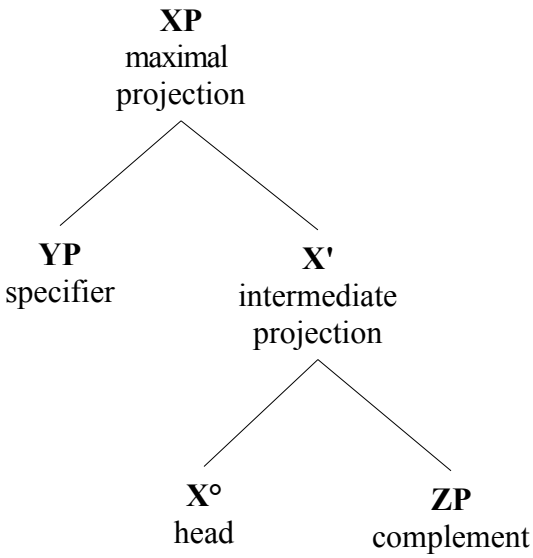


Figure 3: X-bar format

The main advantage associated with X-bar theory is that it accounts for the structure of different categories by capitalizing on one and only one template. The idea that a general schema can be applied to all categories is simple and economic. As a result, the X-bar structure is assumed to be the base structure of all categories. Another common assumption is that each node can have at most two daughters. This means that each node must be binary branching (Kayne, 1984). Following the implementation of the antisymmetric framework (Kayne, 1994), the structure in Figure 3 rigidly applies to all languages also producing a fixed linear order in which specifiers precede heads and heads precede complements.

To illustrate, the verb phrase in (20)b is represented by the structure in Figure 4.[5] The

5 At this point of the argumentation, tense morphology and the internal structure within the nominal domain are not relevant issues. Aspects relative to the DP structure will be discussed in detail in Chapter 2.

linear order of the relative VP is given in (21).

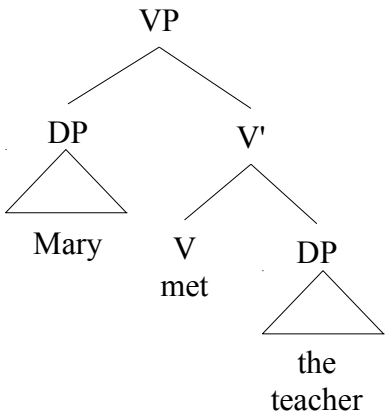


Figure 4: Merge of the verb and its arguments in D-structure

(21) [$_{VP}$ Mary met the teacher]

Continuing the derivation in (20)b, the auxiliary ('has') merges in the structure producing (22). The structural representation is given in Figure 5, and the resulting linear order is given in (23).

(22) {{has} {{Mary} {{met} {the teacher}}}}

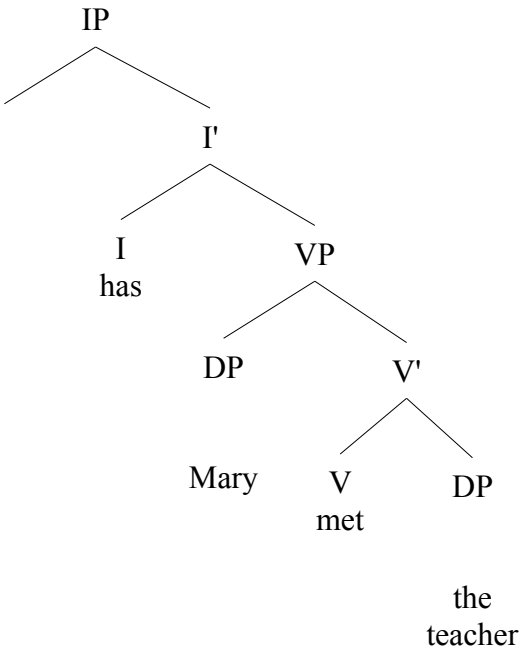


Figure 5: Merge of the auxiliary in D-structure

(23) [$_{IP}$ has [$_{VP}$ Mary met the teacher]]

The structure in Figure 5 is the last step of the derivation of the D-structure.

Then, the D-structure undergoes the application of transformation rules that turns it into the surface structure ('S-structure'). Such transformation component is represented by Move-α where α stands for any category. Syntactic movement happens when elements need to move toward a different syntactic position in order to check linguistic features. This operation produces traces (t) or copies marking the original position of moved elements. For instance, subjects move to the Inflectional projection in order to check their Nominative Case features.[6]

Turning to the example in (22) and (23), the subject of the sentence ('Mary') moves to the specifier position of IP in order to check its Nominative Case feature. Notice that the theory only allows maximal projections to move to specifier positions. This operation consists in re-merging the same element in the structure, as in (24). This is shown in

6 According to the Case Filter, the structure must assign a case to all DPs. Nominative Case is assigned by the finite inflection, so the subject of finite clauses must raise to the specifier position of IP in order to be assigned Nominative Case and fulfill Case requirements.

Figure 6. The resulting linear order is given in (25).

(24) {{Mary}{{has} {{Mary} {{met} {the teacher}}}}}

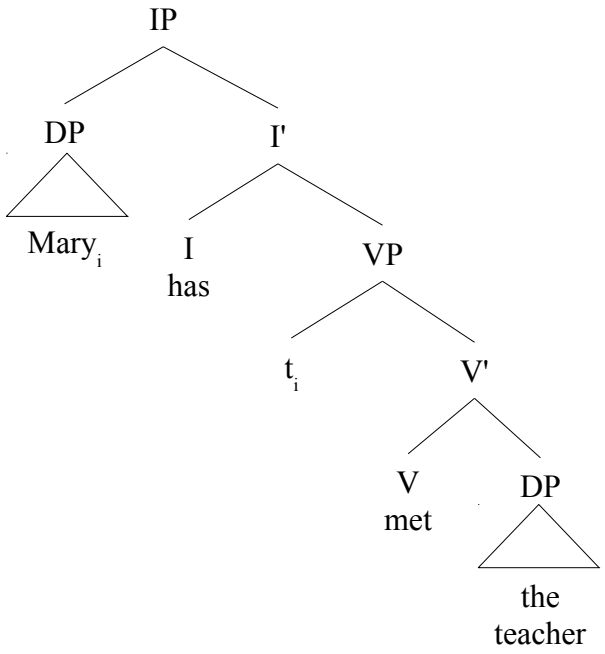


Figure 6: Syntactic movement in S-structure

(25) [$_{IP}$ Mary has [$_{VP}$ <Mary> met the teacher]]

This operation is the last step of the derivation and the syntactic structure is sent to Spell-Out. Then, S-structure is mapped into two interfaces: Phonological Form (PF) which is responsible for the articulatory-perceptual outcome, and Logical Form (LF) which is responsible for the semantic interpretation.[7]

1.3.2. Working in the framework

As discussed in Section 1.3.1, the native speaker's competence (i.e. his/her internal

7 Notice that in more recent generative approaches (Chomsky, 2000; Halle & Marantz, 1993), the lexicon is not considered as a primitive syntactic element, but rather as an aspect of the syntax-phonology interface. Syntax is regarded as the mechanism responsible for the generation of both sentences and (complex) words. The syntactic computation is provided with semantic, and syntactic bundles of features. Syntax operates on the basis of these abstract features.

grammar) is a finite system of knowledge that allows him or her to interpret and produce an infinite number of sentences.

I-language also allows speakers to establish whether a given sentence is well-formed or ill-formed. So, asking someone whether or not s/he would accept a given sentence implies having access to the informant's competence, that otherwise would be inaccessible. This is the reason why generativists frequently collect evidence by asking native speakers for their intuitions ('grammaticality judgments'). For example, a researcher investigating the syntax of negation in English might ask some native-speaker informants whether the sentence in (26) is well-formed or not.

(26) She not is alone

In this case, the sentence would be judged as unacceptable, and hence marked with an asterisk ('*'). If it was a dubious case (neither completely acceptable nor completely unacceptable) it would have been marked by a question mark ('?'). Different levels of acceptability allow a better understanding of the cluster of rules defining competence. Therefore, it is essential for generativists to develop grammars which can account for the speakers' judgments.

Notice that syntactically well-formed sentences are not necessarily acceptable from a semantic point of view. The famous example in (27) shows this contrast: the sentence is structurally well-formed because each element is placed correctly, but it is semantically ill-formed because it is a nonsensical sentence. A different case is shown in (28), which presents a sentence that is both syntactically and semantically ill-formed.

(27) Colorless green ideas sleep furiously (Chomsky, 1957)

(28) Furiously sleep ideas green colorless (Chomsky, 1957)

The main advantage of using grammaticality judgements is that production errors characterizing linguistic performance (slips of the tongue, reformulations, stutters, etc.)

are avoided. Therefore, generativists can fully concentrate on the informant's competence.

As Schütze (2014) observes, there are multifarious judgment tasks. Some options are described in (29).

(29) a. Forced choice task: e.g. "read these two sentences and decide which one is the most (or least) acceptable"

b. Yes/No task: e.g. "is this sentence acceptable? Yes or no?"

c. Likert scale task: e.g. "read this sentence and rate it according to the scale 1-7, with 1 meaning completely unacceptable and 7 meaning definitely acceptable"

In some cases, linguists might employ fillers. These items are not connected with the research questions and are elicited in order to make informants unaware of the researcher's targets.

1.3.3. Major findings in word order

Word order issues can be observed by considering the syntactic structure. Differences in terms of word order can be captured by the reorganization of the structural hierarchy.

Adopting Kayne's (1994) antisymmetric approach, there is one and only one type of structure underlying all languages. On the basis of a number of asymmetries, he argues that the basic order within the X-bar template is Specifier>Head>Complement. At the clausal level, this universal hierarchy is mapped onto the linear order SVO. In order to obtain different linear orders (e.g. SOV, VSO), movement operations allow elements to move along the structure and to re-merge accordingly. For example, the order VSO has an underlying SVO order and is derived by moving the verb across the subject.

A formal approach that developed on the basis of antisymmetric assumptions is the Cartographic Project (Cinque & Rizzi, 2008). It challenged the Principles and Parameter's perspective according to which the clause is structured as in Figure 7

(Chomsky, 1986).

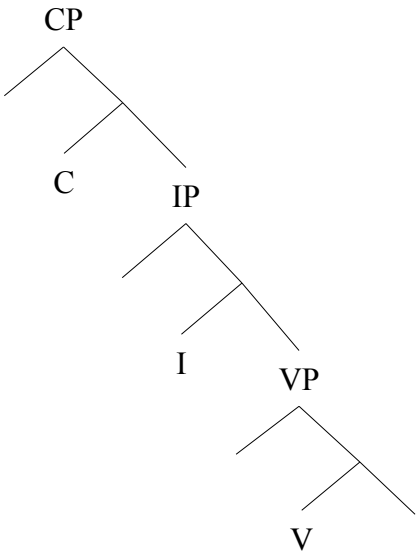


Figure 7: Clause structure according to the Principles-and-Parameters framework

Cartographic studies have been motivated by the fact that the previous syntactic theory was not articulated enough to account for fine-grained analyses. The need for more structural space led the syntacticians to postulate the existence of new functional projections.

The scholar who inspired this new research direction is Pollock (1989). The empirical basis for his argumentation consists in the different distribution of verbs in French and English. In order to show this difference, Pollock employs two diagnostic tools, namely negation and frequency adverbs. The distribution of negation with respect to finite lexical verbs shows the following paradigm: the order in English is Neg>V^{fin} and the order in French is V^{fin}>Neg. This is shown in the examples in (30).

(30)	a. * Mary goes not to Spain	English: * V^{fin}>Neg
	b. Mary does not go to Spain	English: Neg>V^{fin}
	c. Marie ne va pas en Espagne	French: V^{fin}>Neg
	d. * Marie ne pas va en Espagne	French: * Neg>V^{fin}

This could be accounted for by saying that finite lexical verbs in English stay in their original position in V, while in French they move to I. However, the paradigm is more complex than that since in English finite and infinite verbs are consistent, whereas in French infinite verbs behave differently from finite verbs. Compare (30)c with (31)d.

(31) a. * They told me to go not to Spain English: * V^{inf}>Neg

b. They told me not to go to Spain English: Neg>V^{inf}

c. * Ils m'ont dit de ne aller pas en Espagne French: * V^{inf}>Neg

d. Ils m'ont dit de ne pas aller en Espagne French: Neg>V^{inf}

The example in (31)d shows that the lexical infinite verb in French, unlike its finite counterpart, does not move across the negation. If the co-occurrence of finite verbs and frequency adverbs is examined, the resulting pattern reflects the one found with negation: finite verbs in English do not move across the adverb (Adv>V^{fin}), whereas finite verbs in French do (V^{fin}>Adv). This patten is shown in (32).

(32) a. * Mary eats often eggs English: * V^{fin}>Adv

b. Mary often eats eggs English: Adv>V^{fin}

c. Marie mange souvent d'œufs French: V^{fin}>Adv

d. * Marie souvent mange d'œufs French: * Adv>V^{fin}

Given the fact that negation and adverbs show the same distribution with respect to finite verbs, one could conclude that these two elements compete for one and the same syntactic position. The crucial data for the analysis are presented in (33). The French examples show that two word order options are available: infinite verbs can either precede the negation and the adverb or occur in between. In particular, the sentence in (33)c proves that the negation and the adverb must occupy two different positions.

(33) a. * They told me not to eat often eggs — English: * Neg>V^{inf}>Adv

b. They told me not to often eat eggs — English: Neg>Adv>V^{inf}

c. Il m'ont dit de ne pas manger souvent d'œufs — French: Neg>V^{inf}>Adv

d. Il m'ont dit de ne pas souvent manger d'œufs — French: Neg>Adv>V^{inf}

In the light of these data, the structure in Figure 7 is problematic. Assuming that the negation and the adverb are adjoint to VP, there is no available head position between these two elements that can accommodate for the raising of the infinite verb in French. The puzzle is represented in (34).

(34) $[_{IP}\ [_{I'}\ [_{VP}$ pas ? $[_{VP}$ souvent $[_{VP}$ t]]]]]

|_____|

According to Pollock, the only way to account for these data is to expand the structure of the clause providing verb movement with new landing sites. Specifically, he proposes that IP should be divided into two functional projections: one projection hosting agreement morphemes (Agreement Phrase or AgrP) and one projection hosting tense morphemes (Tense Phrase or TP). This proposal is well-known under the name of Split-INFL Hypothesis. Moreover, Pollock (1989) proposes that these two functional projections are separated by an intermediate projection hosting the negation, namely NegP. Pollock's (1989) proposal is linearly represented in (35).

(35) $[_{TP}\ [_{T'}\ [_{NegP}\ [_{Neg'}\ [_{AgrP}\ [_{Agr'}\ [_{VP}]]]]]]]$

Similarly to Pollock, other linguists propose to expand the hierarchy by splitting syntactic projections into more elementary components (a.o the VP-shell hypothesis discussed in Larson, 1988; the Split-DP hypothesis discussed in Giusti, 1996; the Split-

CP hypothesis discussed in Rizzi, 1997).

These studies on functional elements highlight one major problem connected with the structure in Figure 7. This kind of representation does not offer enough space to accommodate everything. At the beginning of the Nineties, a group of linguists started to discuss the need for a richer representation, and that determined the origin of the Cartographic Project (Cinque & Rizzi, 2008). The main aim of the Cartographic studies is to realize exhaustive structural maps that are detailed enough to account for the complexity of human language within an antisymmetric framework.

One of the assumptions on which cartography develops is the 'Uniformity Principle' (Chomsky, 2001). This means that all attested linguistic forms are based on a uniform hierarchical macro-structure. All languages are claimed to share the same inventory of projections, and their relative orders is claimed to be universal. What determines crosslinguistic variation is the way constituents move and the overt or covert realization of the syntactic elements. This proposal is supported by the fact that a universal structure would guide children toward language acquisition.

Another assumption followed by cartography concerns the one-to-one relation between morphosyntactic features and their structural counterparts. In other words, each feature should be represented in the syntactic hierarchy by a unique head (Cinque & Rizzi, 2008). In order to capture the relative order of features and hence of the distinct functional projections, minimal pairs are compared with each other. Cinque's (1999) work on the functional elements populating the former IP layer examines the position of different types of adverbs by observing their positions within minimal pairs. In particular, the examples in (36) concern adverbs occurring in the higher portion of the structure (adapted from Cinque, 1999: 33).

(36) a. Honestly I am unfortunately unable to help you

a'. * Unfortunately I am honestly unable to help you

b. Fortunately, he had evidently had his on opinion of the matter

b'. * Evidently, he had fortunately had his on opinion of the matter

c. Clearly John probably will quickly learn French perfectly

c'. * Probably John clearly will quickly learn French perfectly

The order resulting from (36)a and (36)a' is $Mood_{speech\ act}$(honestly) > $Mood_{evaluative}$(unfortunately). The resulting order from (36)b and (36)b' is $Mood_{evaluative}$(fortunately) > $Mood_{evidential}$(evidently). The resulting order from (36)c and (36)c' is $Mood_{evidential}$ (clearly) > $Mood_{epistemic}$(probably). By transitivity, the overall order is represented in (37).

(37) honestly > (un)fortunately > evidently/clearly > probably

(38) presents the full cartographic representation provided by Cinque with respect to the array of functional projections hosting adverbs (Cinque, 1999: 106).

(38) [*frankly* $Mood_{speech\ act}$ [*fortunately* $Mood_{evaluative}$ [*allegedly* $Mood_{evidential}$ [*probably* $Mod_{epistemic}$ [*once* T(Past) [*then* T(Future) [*perhaps* $Mood_{irrealis}$ [*necessarily* $Mod_{necessity}$ [*possibly* $Mod_{possibility}$ [*usually* $Asp_{habitual}$ [*again* $Asp_{repetitive(I)}$ [*often* $Asp_{frequentative(I)}$ [*intentionally* $Mod_{volitional}$ [*quickly* $Asp_{celerative(I)}$ [*already* T(Anterior) [*no longer* $Asp_{terminative}$ [*still* $Asp_{continuative}$ [*always* $Asp_{perfect}$ [*just* $Asp_{retrospective}$ [*soon* $Asp_{proximative}$ [*briefly* $Asp_{durative}$ [*characteristically* $Asp_{generic/progressive}$ [*almost* $Asp_{prospective}$ [*completely* $Asp_{SgCompletive(I)}$ [*tutto* $Asp_{PlCompletive}$ [*well* Voice [*fast/early* $Asp_{celerative(II)}$ [*again* $Asp_{repetitive(II)}$ [often $Asp_{frequentative(II)}$ [*completely* $Asp_{SgCompletive(II)}$

Methodologically, in order to map features into a detailed hierarchy of projections, cartographic linguists need to work in a comparative way. Once they have discovered a number of possible linguistic realizations, they need to collect data coming from different languages and compare them so that they can abstract from data the common denominators, namely the common underlying structure valid for all languages. On the basis of this universal hierarchy, the existence of different word orders follows from the application of different movement operations.

1.3.4. The relevance of sign languages for generative linguistics

The study of sign languages appears as a fascinating research area for generativists.

First of all, it provides new insights into the architecture of the language faculty. Apparently, spoken languages and sign languages are very different from each other because they capitalize on two different communicative modalities (auditory-oral vs. visual-manual). Nonetheless, if language has to be considered as a biological endowment defining all human beings (a sort of 'mental organ'), then spoken languages and sign languages should be based on the same core properties. The general prediction is that language universals are such that they can be found in any language regardless of the modality in which languages are transmitted. So, looking at sign languages is important in order to confirm the universal validity of some linguistic features.

For example, in (39), verbal morphology can be observed both in a spoken and in a sign language, namely in Ubykh (a Northwest Caucasian language; the example is taken from Charachidzé, 1989: 395) and LIS. Despite modality differences, in both languages the verb agrees with both the subject and the object showing an instance of polypersonal agreement.

(39) a. wə-Ø-byά-n Ubykh

2sg-3sg-see-PRS

'She sees you'

b. $_3$CRITICIZE$_1$ LIS

'She criticizes me'

Assuming that a common architecture is shared by spoken and sign languages means that the core properties are linguistically coded in both modalities. However, we should not necessarily expect that the linguistic features licensed by Universal Grammar can be exhaustively studied by looking at the spoken modality only. Specifically, there are

some aspects that can be investigated only by looking at sign languages. For example, one of these is the use of space to express referential features. So, studying sign languages may help have a better grasp on the potentialities and also the limits of Universal Grammar.

The possible effects of language modality on the structure of grammar is not the only benefit that sign language linguistics can offer to Universal Grammar. In some areas of the world, signed linguistic systems have emerged spontaneously only in recent years. Nicaraguan Sign Language (NSL) and Al-Sayyid Bedouin Sign Language (ABSL) are two examples of early languages and provide evidence for studying language genesis (for more details on emerging sign languages the reader is referred to Meir, Sandler, Padden & Aronoff, 2010). These languages not only emerged out of nothing, but also are completely unrelated to any well-developed language. As a matter of fact, the users of these early sign languages do not have any first language. This scenario characterized by the emergence of a new language with no substratum interferences is completely uncommon among spoken languages and can help understand what Universal Grammar exactly allows and disallows.

Similar circumstances can be observed when studying the case of home-signers. These are deaf individuals born to hearing non-signing parents. They experience linguistic isolation because their hearing deficit prevents them from acquiring a spoken language and the lack of contact with signers prevents them from acquiring a sign language (Goldin-Meadow, 2012). Due to this exceptional situation, they spontaneously develop a gestural system that may reflect some restrictions imposed by Universal Grammar. In case they get in contact with signers they get exposed to a full-fledged linguistic system. The accuracy with which they can acquire it is indirectly proportional to the period of time they experienced language deprivation. Such situation may provide new insights on the critical period for language acquisition.

Another atypical circumstance that can be observed when studying sign languages is the fact that the majority of signers do not get exposed to signs until late childhood or early adulthood.[8] In their own families they are often first generation signers. They often have access to impoverished stimuli (e.g. hearing parents using a pidgin sign language) and

8 This aspect has been considered during the collection of corpus data used for this dissertation (cf. Section 3.2.1.).

non-conventional linguistic sources (e.g. teachers using manual codes that follow the word order of spoken languages). Given this scenario, signers tend to introduce some adjustments, and this determines a process of constant re-creolization in each generation (Fischer, 1978).

1.4. Sociolinguistics

In this section, the sociolinguistic framework is presented. First, I discuss the main research topics that sociolinguists are interested in with a special focus on language variation. As for methodology, I go through the main procedural steps of a quantitative study from the identification of relevant sociolinguistic variables to the final statistical computation. In the section concerning word order, I present two case studies showing instances of diachronic and diatopic variation, respectively.

1.4.1. General framework

The basic claim underlying the sociolinguistic view of language facts is that human languages do not function as monolithic systems. In other words, a language used by a certain community is not completely uniform among the community itself. If this was not the case, all the speakers of that language would express themselves identically, and no variation in space and time would occur.

A discipline emerged in the 1960s, sociolinguistics, proved not only that languages actually vary, but also that this variation can be systematically investigated. One of the tenets followed by sociolinguists is that languages vary according not only to language-internal factors but also to language-external factors (i.e. social factors). Community and language are strictly connected with each other, each exerting reciprocal influence on the other. The interaction between these two poles is what sociolinguistics aims at investigating.

Differently from other types of linguistic inquiry, sociolinguistics is not interested in the ideal native speaker. Sociolinguists have rather a speaker-centered approach. They draw their attention to real speakers and above all real linguistic environments. For this

reason, the linguistic community gets directly involved during sociolinguistic fieldwork. To exemplify, some of the issues that they address are: how does language vary according to different social settings (e.g. academic lecture, waiting room, suburbs pub, etc.)? Are there differences between women's and men's speech? More generally, what are the main extra-linguistic factors that influence the way people speak or sign?

Crucially, sociolinguists should be considered as linguists rather than sociologists (Berruto, 1997). Indeed, their research topics are always linguistic phenomena. These topics are usually investigated from a multidisciplinary perspective, including not only linguistics and sociology, but also anthropology and psychology.[9]

Within the sociolinguistic framework, one can find various research areas, namely bilingualism, multilingualism, language contact and language shift, sociolinguistic variation, language planning and policy, interaction and discourse analysis, language attitudes, and politeness norms.

Before discussing sociolinguistic variation, I shall make a short digression on the two approaches that have been developed within sociolinguistics: correlational and functional sociolinguistics (Berruto, 1997). The main goal of correlational scholars, represented by William Labov, is to find correlations between linguistic factors and social factors. They aim at studying how linguistic variation occurs in a community and to what extent independent social variables exert their influence on such variation. The prototypical correlational study was conducted on the variable pronunciation of /r/ in three different shopping malls in New York (Labov, 1966). On the contrary, functional sociolinguists focus on the ethnographic aspects of language and investigate both linguistic and social factors to the same extent. In order to analyze discourse, they focus their attention on social interaction and on the use of contextualization cues. The main exponent of this group of scholars is Gumperz, the founder of interactional sociolinguistics (Gumperz, 1982).

The sociolinguistic work included in this thesis follows Labov's approach. Therefore, in the remainder of the section, I focus on variation from the correlational rather than from the functional perspective. Crucially, the fact that languages are constantly changing

9 Sociolinguistics should not be confounded with sociology of language since they are two distinct disciplines. The former investigates the effects of society on language; the latter studies the effects of language on society.

should not lead to the conclusion that variation is chaotic and accidental. According to Weinreich, Labov & Herzog (1968), languages are characterized by an "orderly heterogeneity", hence variation is not free. This means that the available options are not endless, they are rather limited to precise restrictions and parameters. This reflects the structural essence of grammar (Tagliamonte, 2006).

When studying variation, sociolinguists are interested in understanding how a variable is realized in the linguistic community. A variable, according to Labov (2008), consists of two alternatives meaning the same thing. In other words, as long as a linguistic phenomenon has two possible realizations, a variable can be studied. Sociolinguistic researchers aim at exploring both the linguistic and the extra-linguistic context in which the two variant forms can be observed. But it is not sufficient to say that variables should have at least two variant forms. Rather, they should obey some particular criteria. According to Labov (as reported in Berruto, 1997), good variables should meet the requirements in (40).

(40) a. Frequency: variables should occur frequently in the corpus. Notice that researchers should not exert special prompting during elicitation sessions

b. Unconsciousness: speakers should not have a conscious control over the variable under investigation. The reason is that researchers aim at analyzing spontaneous productions without any possible data manipulation

c. Broad context: variables should not be produced in isolation, but rather within a given context. This condition favors lack of conscious control over the linguistic production

d. Measurability: variables should be easily quantified on a linear scale, so that a quantitative analysis can be conducted

Variables can occur at every linguistic level (phonology, morphology, syntax, pragmatics, etc.). A few examples are listed in (41) (adapted from Tagliamonte, 2006: 10-11).

(41) a. Phonology: the -ing/-in' alternation (e.g. doing/doin')

b. Morphology: adverbs with Ø/-ly alternation (e.g. serious/seriously)

c. Syntax: deontic modality (e.g. I've got to see you again/I have to go away); agreement (e.g. she were a good worker/she was a good worker)

g. Pragmatics: quotative use (e.g. he just said 'Ok'/It was like 'Ok'/I thought 'Ok')

The language-external factors that may influence the realization of a variable are detected within the linguistic community. The most common ones are: age, gender, social class, geographical provenance, urbanization, ethnicity, sexual orientation, religion and level of education.

Given the multiplicity of variables and the possible correlation among them, sociolinguistic variation is both heterogenous and dynamic. There is a general consensus among sociolinguists about the existence of five main dimensions of variation (Flydal, 1951; Coseriu, 1973; Mioni, 1983). This classification is reported in (42) and discussed below with some examples from English.

(42) a. Diachronic variation — social factor(s): age

b. Diatopic variation — social factor(s): provenance, urbanization

c. Diastratic variation — social factor(s): social class, education

d. Diaphasic variation — social factor(s): communicative setting

e. Diamesic variation — social factor(s): communicative medium

Diachronic variation (literally, variation throughout time) is connected with the temporal dimension. It can be intended in terms of language change or language variation. Language change is a crossgenerational phenomenon and involves some linguistic innovations that show up and then propagate over a period of time from one generation to another. Such innovations usually affect first the spoken language and

then can contaminate written language as well. If diachronic variation derives from language-internal factors, we can observe word formation processes such as grammaticalization and lexicalization, as exemplified in (43) and in (44) respectively. The former usually takes a longer time than the latter.

(43) Grammaticalization:

from Old English 'willan' (full verb meaning to want or to wish) to Modern English 'will' (future auxiliary)

(44) Lexicalization:

from Old English 'Crīstes mæsse' to Modern English 'Christmas'

Language-external factors can also trigger diachronic variation. This is the case of borrowings from other languages. Loanwords can be motivated by several factors, such as cultural contact, fashion, trade, war, immigration, colonization. Some examples of loanwords used in English are listed in (45).

(45) Borrowing:

a. Loanwords from Italian: 'viola', 'soprano', 'broccoli'

b. Loanwords from Native American languages: 'skunk', 'moccasin', 'chipmunk'

c. Loanwords from Sanskrit: 'avatar', 'karma', 'yoga'

Differently from language change, language variation through time is a more circumscribed phenomenon. It is determined by linguistic differences characterizing speakers who belong to different age groups. A few examples of lexical variation according to age is given in (46).

(46) Variation according to age:

a. young speakers: 'fridge', 'telly', 'fab'

b. old speakers: 'refrigerator', 'television', 'excellent'

Diatopic variation (literally, variation throughout space) concerns linguistic differences occurring in different places and regions. A 'regiolect' is a variety used in a specific area. Different realizations can be more or less evident and they can be at the phonological, orthographical, or morphosyntactic level. A few examples are reported in (47).

(47) British English (BrE) vs. American English (AmE):

a. d[a]nce in BrE, d[ae]nce in AmE

b. 'different from' in BrE, 'different than' in AmE

Diastratic variation is influenced by a number of social factors. Linguistic choices can be influenced by the social status, gender, profession and education background of the utterer. Everyone is part of a particular social context in which common linguistic features are shared. To illustrate, the lexical differences in (48) are connected with social class.

(48) Variation according to social classification:

a. working class: 'mum', 'bloke', 'pissed'

b. middle class: 'mummy', 'chap', 'drunk'

Diaphasic variation is associated to the setting, the topic, and the participants involved in the communicative event. Utterances may be characterized by either a formal or an informal register, depending on the situation. Besides, if participants are interacting in a professional environment, they are likely to share some technical and subject-specific expressions (i.e. a technical jargon). The difference between a formal and an informal context is visible in various linguistic aspects, as illustrated in (49).

(49) Register differences:

a. formal register: 'I am not'; 'I will investigate the matter'; 'thank you for your e-mail'; 'how do you do?'

b. informal register: 'I ain't'; 'I'll look into it'; 'thanks for your mail'; 'how are you doing?'

Diamesic variation depends on the communication means. The channel through which utterances are transferred from one participant to the other may influence the content and the configuration of utterances themselves. Traditionally, two possibilities are taken into account: written language and spoken language. The examples in (50) show the difference between the two.

(50) Variation according to the communicative channel:

a. Spoken English: 'I'm, like, kinda busy tomorrow'

b. Written English: 'I will be rather busy tomorrow'

By virtue of technologies, a third option needs to be considered, namely transmitted messages. These are messages transmitted over some distance (e.g. television or radio broadcasting, web-pages, etc.).

Notice that each of these dimensions of variation should not be intended as rigidly independent from the other. For instance, diachronic variation may intersect diaphasic variation. Considering the case of young speakers, it is unlikely that they express themselves identically in every social context: it is reasonable to think that they use different linguistic features when they chat with their peers and when they speak to their teachers.

Variationist linguistics, the branch of correlational sociolinguistics that focuses on how languages vary, applies quantitative methods (see Section 1.4.2). This research field has gradually evolved with respect to research goals and methodology. The evolution of this

discipline has been described in Eckert (2012) in terms of "three waves of analytical practice".

The first wave of quantitative studies ("the survey era" in Eckert, 2012), which is the one adopted in this dissertation, was initiated by Labov's (1966) seminal work and is mainly based on the correlation between linguistic variables and major demographic factors, such as gender, ethnicity, age, and social class. This first wave shows the socioeconomic stratification of language variation. Assuming the standard variety to be the one spoken by the educated upper working class, the other varieties (what Eckert call "community vernaculars") reflect different socioeconomic backgrounds. The social hierarchy functions as a map guiding the analysis of sociolinguistic phenomena. Nonetheless, some exceptions have been found. According to Labov (1966), non-standard (i.e. vernacular) variants are also used by individuals belonging to the lower middle class and to the upper working class. Interestingly, people more locally engaged tend to prefer vernacular forms, whereas people living in urban centers more frequently use standard forms. This stage of sociolinguistic research shows that the social meaning of variation is clearly focused on social class and socioeconomic stratifications.

During the second wave of quantitative studies ("the ethnographic approach" in Eckert, 2012), social categories are less abstract and are observed in local social practice. Interestingly, Labov notices that local identity is important in determining the use of one linguistic option or another. In some cases, language change is quickly accepted and integrated in everyday communication, while in other cases it is obstinately rejected. In other words, the involvement of individuals within their community influences their way of expressing themselves and interacting with other members. For this reason, local dynamics, social networks and social identity should be both taken into account during variation analysis.

The third wave of quantitative studies ("practice and the stylistic perspective" in Eckert, 2012) is responsible for shifting attention from social structure to "the role of practice in producing and reproducing structure" (Eckert, 2012). This stage of sociolinguistic research focuses on the everyday practices in which individuals make use of language in order to build social styles and form social meanings. In so doing, attention is shifted from linguistic variants to stylistic practices. This involves ongoing adjustments,

stylistic elaboration, and intentional choices. Unlike first-wave and second-wave researchers, third-wave ones do not observe language variation with the purpose of evaluating how social meaning is reflected on it. They want to study how variation is used to construct style types and social meaning.

1.4.2. Working in the framework

In Section 1.4.1, I sketched the difference between the two main groups of sociolinguists, namely correlational and functional researchers. These two groups differ not only with respect to the theoretical assumptions but also with respect to the research methodology. As discussed in Berruto (1997), the functional approach prefers using qualitative methods: it focuses on interactional procedures, observation in natural settings, evaluation of the participant's perspective and interaction with the communicative context. On the other hand, the correlational approach prefers quantitative methods: it focuses on authentic data from a statistical point of view and tries to identify possible correlations between linguistic factors and extra-linguistic factors. In order to highlight variation, quantitative analysis and statistical computing are systematically employed.

In this methodological section, I focus my attention on the quantitative approach because this is the direction I followed during the empirical part of my study.

With respect to this, the denomination 'quantitative sociolinguistics' is commonly used to indicate the analysis of an amount of data which is sufficiently representative of the relevant linguistic community. On the basis of these data, statistical analysis is used to study distribution and frequencies (Vietti, 2005) and to identify tendencies and regularities (Tagliamonte, 2006). The creation of quantitative models aims at evaluating the strength and statistical significance of pre-established hypotheses. Accordingly, predictions over a larger community can be formulated.

In a quantitative study[10], the first step is identifying the linguistic variable of interest and the possible variant forms. Afterwards, the alternatives should be examined in relation to the linguistic and extra-linguistic contexts in which they occur. Possible subtle

10 For more details on the procedural steps of a quantitative study the reader is referred to Tagliamonte (2006).

differences in meaning or function need to be detected and taken into consideration. Sociolinguistic metadata on the informants should be collected and analyzed, too. The null hypothesis implies that no variation is attested within a population and that no significant difference is observed with respect to a given phenomenon (status quo). On the contrary, the alternative hypothesis suggests that the observed results are instances of language variation. Which of the two hypotheses proves to be correct is what the sociolinguist needs to discover. A pair of examples that clarify the difference between the null and the alternative hypothesis is given in (51).

(51) a. Null hypothesis: there is no relationship between people's age and the frequency of the colloquial expression 'you know'

b. Alternative hypothesis: there is a relationship between people's age and the frequency of the colloquial expression 'you know'

An important distinction that should be kept in mind is the one between dependent variable and independent variable. The linguistic phenomenon under investigation represents the dependent variable. To be more precise, it is a parameter that can be realized in one way or another (Lucas, Bayley & Valli, 2001). It is also referred to with the expression "internal restriction" and it is always a linguistic parameter (e.g. a phoneme, a morphological configuration, a syntactic structure, etc.). Besides the dependent variable, quantitative studies on linguistic issues have to take into consideration a number of independent variables that can be either linguistic or extra-linguistic. These are factors that can have an influence on the realization and distribution of the dependent variable. They are also referred to with the expression "external restriction".

In order to conduct a quantitative study, it is essential to collect authentic linguistic material produced by a subsection of the community (sample). The most common research methods employed to collect data are rapid anonymous surveys, polling techniques, interviews, elicitation sessions, and free conversation (Starks & McRobbie-Utasi, 2001). An important aspect is awareness, as already explained in (40). While participating in the data collection, informants should not have a conscious control over

the production of the variant forms under investigation so that their production is not contaminated by manipulations and hypercorrections. However, as soon as the description of the variable is finished, the community should be able to recognize the same variable as a distinctive feature. Data collection produces a corpus in which the realizations of the variable can be studied.

The next step is data annotation. This should be done without excluding any forms a priori (Tagliamonte, 2006). Annotation could be done manually or following semi-automatic or automatic procedures. The annotated information should be detailed, simple and, most importantly, machine-readable.

As soon as all the variables are annotated according to a pre-established annotation protocol, data can be exported into a dataset. This is not the end point of the entire process. It is rather the raw material that should be imported in a statistical software, like for example R (R Core Team, 2013), in order to start data analysis. These tools allow variationist linguists to evaluate whether the presence or absence of certain independent variables may affect the distribution of the dependent variable. Moreover, the weight of each factor (be they linguistic or extra-linguistic) can be evaluated and statistical inferences can be drawn in order to direct future research.

Depending on the type of data and the hypotheses that need to be tested, sociolinguists choose the most suitable statistical test (a.o. t-test, chi-square test, z-score, f-test). During the statistic calculation, the influence that each factor exerts on the presence of the dependent variable is measured. Such measure is called 'factor weight'. If the influence of a factor is extreme (i.e. not driven by chance), its weight is statistically significant. The p-value helps evaluate how significant a result is. It is used to “assess what proportion of the population has even more extreme values. The smaller this proportion is, the more reason we have for surprise that our test statistic is as extreme as it actually is” (Baayen, 2008: 68). The threshold of significance (α) which is generally assumed in social sciences is 0.05 ($\alpha = 0.05$). If two variables are compared with each other and the corresponding p-value is lower with respect of this threshold ($p < 0.05$), then the null hypothesis is rejected. This is because the difference between the two values is extreme, hence statistically significant. The lower the p-value, the more stringent the evidence against the null hypothesis. Conversely, if the p-value is higher

than the significance threshold ($p > 0.05$), then the correlation between the two variables is not extreme (i.e. the result is not statistically significant). Therefore, the null hypothesis is confirmed.

1.4.3. Major findings in word order

There is a number of sociolinguistic studies dealing with variation at the level of syntax. In particular, word order issues have been investigated both in their synchronic and diachronic dimension (for further details see Kroch, 2000).

As for diachronic variation, syntactic change has not affected all languages to the same extent. Let us think of two different cases. On the one hand, English has witnessed three major syntactic changes in its evolution: from INFL-final to INFL-medial, from verb-second to subject-verb order, and from OV to VO order (Kroch, 2000). On the other hand, Japanese has barely changed over centuries since its final-head pattern has remained stable at all levels of syntactic structure (Kroch, 2000).

The evolution of a language is often reinforced by altered conditions of language transmission and by input reanalyses applied by children during language acquisition. I present now a case study that shows how it is possible to study diachronic variation in word order by looking at the production of different age groups. This study (Barrett, 2008) concerns Sipakapense, a language that is spoken in Guatemala and belongs to the Mayan family. Like other Mayan languages, Sipakapense is regarded as a less prestigious language with respect to Spanish, the language spoken by colonialists. The linguistic convergence from Sipakapense to Spanish is visible in several linguistic aspects, and word order is one of these. As for the basic word order, Barrett claims that VSO is the most common option. The example in (52) shows a declarative sentence having the canonical VSO order (Barrett, 2008: 292).

(52) X-i'-r-to' Pey ri aliit
COM-3pA-3sE-help Pey DET girls
'Pey helped the girls'

In the case of agent focus construction, Sipakapense speakers front the subject at the beginning of the sentence and make use of the SVO order, as in (53) (Barrett, 2008: 292). The community, though, tends to perceive this order as a direct influence of Spanish. According to language activists, the SVO order embodies the Spanish domination.

(53) Pey x-i'-to'-w ri aliit.
Pey COM-3pA-help-AF-PFM DET girls
'It was Pey who helped the girls'

Barrett's (2008) analysis is based on conversational data coming from three age groups (i.e. speakers born between 1914 and 1934, speakers born between 1960 and 1970, and speakers born between 1980 and 1990). The striking result that he highlights concerns the agent focus construction. The prevalent order in the older and middle-aged group is SVO and in the young group is SOV. This shows an ongoing diachronic process: older and middle-aged speakers are still influenced by the dominant language whose canonical order is SVO, whereas younger speakers avoid using Spanish word order and provide an impulse to revitalization. This structural change put forward by the younger generation implies an overgeneralization against SVO order. This pattern was used to express agent focus construction in previous stages of the diachronic evolution of Sipakapense, but it is currently rejected because it is perceived as too similar to the language of the colonialists.

With respect to diatopic variation, I present a case study on three-verb clusters in some Dutch non-standard varieties (Barbiers, 2005). In standard Dutch the canonical order with these constructions is 1>2>3, as in (54) (Barbiers, 2005: 233).

(54) dat hij moet$_1$ kunnen$_2$ zwemmen$_3$
that he must can.INF swim.INF

'that he must be able to swim'

On the basis of data coming from a questionnaire filled out by 368 informants, the author shows that a high degree of geographical variation can be observed. In other words, three-verb clusters display different order options throughout Dutch dialects. In particular, transitional areas between the Dutch and the German language area are more subject to this variation with respect to other zones. For example, in the area of Dutch Limburg and the south-eastern part of North Brabant three word orders are attested, namely 1>2>3, 1>3>2 and 3>1>2. Barbiers hypothesizes that the grammatical system allows for a number of word orders. The types of word ordering in three-verb clusters to which speakers are exposed in the environment determine their preferences. Speakers living in transitional areas get exposed to more different orders and, as a consequence, tend to use more alternatives than speakers living in non-transitional areas.

1.4.4. The relevance of sign languages for sociolinguistics

With respect to diamesic variation (cf. Section 1.4.1), sign languages display some interesting characteristics. These are listed and discussed in (55).

(55) a. sign languages are transmitted in the face-to-face modality in the sense that they require the interlocutors to be present in the same physical environment. The face-to-face interaction is the prevalent communication modality among signers.

b. sign languages do not have a systematized written form. Even if some writing systems have been tentatively proposed over the years (Mantovan & Celo, 2008), they have not been widely accepted and used, yet.

c. sign languages can be transmitted across long distances through recorded videos and online video calls. This transmitted modality may have an influence on sign language (e.g. reduction in the signing space, adjustments due to the transmission of a two-dimensional space, new metalinguistic signs, etc.).

The use of sign languages is influenced by a number of peculiar sociolinguistic aspects: atypic language acquisition, debatable definition of the 'good signer', pressure from the dominant spoken language, absence of standard variety. Given these circumstances (which will be shortly discussed one by one), a high degree of variation is expected. For this reason, from a sociolinguistic perspective, the study of sign languages is particularly appealing.

First of all, the composition of the family nucleus is atypical with respect to any sociolinguistic setting of spoken languages. Only 5-10% of Deaf children are born to Deaf parents and hence acquire sign language as their first language in early childhood and in a natural environment. The vast majority of Deaf individuals are born to hearing parents. As a result, they often do not get exposed to sign language in early childhood. They start acquiring it when they enter the school system. If the school does not offer a bilingual program and if there are no signers in the educational environment, the exposition to sign language can be even delayed to early adulthood after the first contacts with signers. This anomaly is interesting because it introduces a social factor based on family composition that can correspond to different levels of competence and different linguistic outputs.[11] Sign languages are unique in this respect and offer a privileged view to the theory of sociolinguistics.

Another consequence depending on the family composition is that native signers are only a small minority in the Deaf community. An interesting research question is: who drives language change? The minority of native signers or the majority of non-native signers? And how do the two groups influence each other? Who should be involved in experimental sessions (native signers, near-native signers, non-native signers)? This is another uncommon circumstance that sociolinguistics has to deal with when considering sign languages.

Another interesting phenomenon that may be of interest for sociolinguists is language contact. Sign languages are in constant contact with spoken languages since they are both used by the Deaf community on a daily basis. It could resemble the cases of many dialects spoken by local communities. The exceptionality of sign languages consists in

11 Crucially, information on family composition should be included in metadata questionnaires.

the fact that they are expressed though a different modality (spoken vs. signed) and there is no difference in geographical dimension between the minority and the dominant language. In some cases, inter-modal language contact is influenced by the fact that the dominant spoken language is regarded as more prestigious than the sign language used by the Deaf minority. The fact that some countries have not officially recognized the local sign language yet may have a negative impact on their social status. As a consequence of the influence exerted by the language spoken by the hearing majority, signers may be induced to overuse fingerspelling and to follow the word order of the dominant spoken language. This is more likely to happen in very formal settings. However, the more aware the Deaf are of their culture and language, the fewer interferences from spoken to sign language emerge.

Lastly, sign languages provide a peculiar empirical basis for sociolinguistic studies because most of them do not have a codified standard variety and hence display a high degree of variation. This is due to a number of reasons, for example the lack of official recognition in some countries, the paucity of projects promoting bilingual education, the small number of available descriptive grammars, the absence of a systematized written form.

1.5. Conclusion

This introductory chapter presented the three theoretical frameworks on which the thesis is based. The goals and the methodology characterizing typological linguistics, generative linguistics, and sociolinguistics have been illustrated in some detail. Even if each approach inevitably has its own peculiarities, there are some common aspects. For example, both typology and sociolinguistics are interested in studying in what respect languages vary and differ from each other. Both generative grammar and typology look for specific constraints that guide the production of linguistic utterances. Both generative linguists and sociolinguists are interested in interfaces: formal linguists investigate the interface between language and human cognition, while sociolinguists explore the interface between language and society.

All three approaches deal with variation, although to different extents. An aspect that

challenges both generativists and typologists is language-internal variation, namely the possibility to find different word orders in a single language or even in one and the same speaker. This phenomenon is problematic at the syntactic level where consistency is expected. It is problematic because in the generative and typological frameworks language is considered a relatively uniform abstraction and a stable system. Under this perspective, the possibility to arrange linguistic items in a relatively free way is hard to explain. The main theoretical innovation of this thesis consists in examining a typological universal (Universal 20, cf. Section 2.2.), showing that it is subject to language-internal variation in LIS, and explaining how such variation obeys precise linguistic constraints even in absence of a standard form.

In the next chapters, I will follow a multidisciplinary approach. In particular, I will show that it is possible to integrate typological generalizations and sociolinguistic facts into a unique formal account.

2. Literature review on nominal modification

2.1. Introduction

This chapter provides an overview on the significant literature on nominal modification. This is not intended as a comprehensive account of what has been published with respect to this research topic. Rather, it offers background information explaining the starting point from which this research work developed.

The term 'nominal expression' is intended as the extended projection of the Noun Phrase (hereafter NP) (in the sense of Grimshaw, 1991) containing the noun and its modifiers. In current theories, this means that nominal expressions include the NP, the DP, and all other projections hosting nominal modifiers. Nominal modification is to be regarded as the possibility to enrich the structure of the NP providing extra information about the noun. 'Nominal modifiers' is here considered as an umbrella-like expression including all the elements that are part of the extended projection of the NP (e.g. possessives, demonstratives, adjectives, relative clauses, etc.).

The complexity of the NP can be increased by adding nominal modifiers. The use of more or less complex nominal expressions in discourse may be influenced by several factors. I list some of them in (1) (for a more extensive discussion see Rijkhoff, 2002).

(1) a. Familiarity: the more familiar referents are, the more likely it is to use non-complex nominal expressions, or even pronominal forms.

b. Style: formal discourse triggers the production of circumlocutory polite phrases and complex nominal expressions.

c. Processing constraints: in spontaneous non-elicited discourse, complex nominal expressions are not frequent and they are limited to two, maximally three, modifiers.

d. Culture: unfamiliar notions are likely to be referred to by employing periphrastic expressions, hence complex nominal expressions.

In early generative works (cf. Section 1.3.), the complexity of NPs is captured by a recursive four-level structure composed of the lexical head N° and three superordinate levels projected by the head itself, namely N', N'', and N''' (Jackendoff, 1977). Such template allows more than one specifier and more than one complement within the same projection. According to Jackendoff (1977), the highest specifier, the one dominated by N-triple-bar, hosts modifiers that are marked for the feature [+Det], namely determiners or genitive NPs. The lowest specifier, the one dominated by N-double-bar, hosts modifiers that are marked for the feature [-Det], namely quantifiers. The problematic aspect of this analysis is that different types of modifiers compete for the same syntactic position (e.g. articles and demonstratives in [Spec, N''']). Crucially, no clear prediction is made on the relative order of modifiers.

Following studies on the nominal domain reveal the undesired contrast between a rather simple syntactic structure and the presence of more and more elements to accommodate. Arguing that nominal expressions need a more articulated structure, Abney (1987) proposes that Jackendoff's forth level (i.e. N''') belongs to a different syntactic category. According to Abney, the NP is embedded in a functional projection. He refers to such projection as Determiner Phrase (hereafter DP) and claims that it can accommodate determiners and genitive NPs. Therefore, the determiner is not a specifier within the NP, it is rather the head of a functional projection taking the NP as its complement. Unlike Jackendoff, Abney does not resort to multiple specifier positions in one single projection. Abney's proposal is well-known under the name of 'DP hypothesis'.

One of the main tenets of the DP hypothesis is that nominal domain and clausal domain display a similar syntactic configuration (Abney, 1987; Szabolcsi, 1987, 1994). They are both based on a lexical projection: NP for the nominal domain and VP (Verb Phrase) for the clausal domain. The VP is anchored to discourse by the functional projection IP (Inflection Phrase) which entails temporal and agreement specifications. The NP is also linked to the discourse context. The linker consists in the functional projection DP which conveys information on the reference of the noun.

The DP hypothesis sets the ground for the discussion included in this chapter. However, as the following sections will show, the structure of nominal expressions is more complex than originally thought. As we will see, Cinque (2012) applies the cartographic approach also to the nominal domain proposing the explosion in the number of functional projections populating the extended projection of the NP. In this chapter, Section 2.2 introduces Greenberg's Universal 20 and discusses the distribution of demonstratives, numerals, and adjectives both from a typological and a generative perspective. Section 2.3 presents the extended projection of the NP from the generative perspective with the purpose of discussing the fine-grained hierarchy hosting nominal modifiers. Section 2.4 provides an overview of what has been documented with respect to the nominal modification in sign languages. First, word order facts relative to several sign languages are illustrated, then the discussion focuses on LIS.

2.2. Universal 20

After having introduced the notion of 'universal' from a typological perspective in Section 1.2.1., I focus on one of Greenberg's universals, namely Universal 20. This generalization represents the empirical domain of this dissertation. The main goals of the discussion is to find how three different theoretical approaches can converge, to integrate different methodologies in a multi-method research, and finally to enrich the description of the grammar of a language, LIS. The generalization contained in Universal 20 concerns word order at the nominal level and it has triggered a lively debate over the years. In the next sections, both formal analyses (Cinque, 2005; Abels & Neeleman, 2009) and typological analyses (Cysouw, 2010; Dryer, 2011) are presented.

In his influential paper published in 1963, Greenberg presented 45 universals sub-divided into three main parts: typology, syntax, and morphology. Most of Greenberg's universals are formulated as implicational universals (cf. Section 1.2.1.). Taking a deeper look at universals on syntax, I further restrict my attention to Universal 20 which concerns the position of three nominal modifiers, namely demonstrative, numeral, and adjective. Greenberg's original formulation of Universal 20 is reported in (2) (Greenberg, 1963: 87).

(2) “When any or all of the items (demonstrative, numeral, and descriptive adjective) precede the noun, they are always found in that order. If they follow, the order is either the same or its exact opposite.”

According to the generalization in (2), the possible word order options are three, one with prenominal modifiers, (3)a, and two with postnominal modifiers, (3)b and (3)c.

(3) a. Dem>Num>Adj>N (e.g. English, Turkish)

b. N>Dem>Num>Adj (e.g. Noni, Abu)

c. N>Adj>Num>Dem (e.g. Yoruba, Thai)

Greenberg’s formulation is however too strong and fails to capture the attested crosslinguistic variation. In particular, the mixed orders with both prenominal and postnominal modifiers are not accounted for. For example, Num>N>Adj>Dem is the canonical order found in Celtic, Hebrew, Indonesian, and other languages but it is not captured by Greenberg’s generalization.

On the basis of a larger database and some counterexamples to Greenberg’s statement, Hawkins (1983) proposes the less restrictive reformulation in (4) (Hawkins, 1983: 119-120).

(4) “When any or all of the modifiers (demonstrative, numeral, and descriptive adjective) precede the noun, they (i.e., those that do precede) are always found in that order. For those that follow, no predictions are made, though the most frequent order is the mirror-image of the order for preceding modifiers. In no case does the adjective precede the head when the demonstrative or numeral follow.”

The crosslinguistic data included in Section 2.2.1 will show that, when all modifiers occur before the noun, only one possible order is attested (Dem>Num>Adj>N). On the contrary, when all modifiers follow the noun, several possible orders are attested. This asymmetry between prenominal and postnominal modifiers is captured by Hawkins' version in (4).

2.2.1. A formal derivation

After forty years from its original formulation, Greenberg's Universal 20 gave rise to a lively debate in which generativists and subsequently linguists with different theoretical backgrounds provided an account for crosslinguistic variation in the nominal domain. This section offers an outline of the formal accounts proposed by Cinque (2005) and Abels and Neeleman (2009).

According to Cinque (2005), the problem with Hawkins' (1983) formulation is that the distribution of postnominal modifiers is apparently unrestricted ("for those that follow no predictions are made" in Hawkins, 1983). Still, some of the word order options containing postnominal modifiers are not attested in the world's languages (e.g. * N>Num>Dem>Adj).

In recent works on the distribution of demonstrative, numeral, and adjective, Cinque (2003, 2005) shows that of the 24 logically possible permutations (factorial 4 = 1x2x3x4 = 24) only 14 orders are effectively attested in the languages of the world. For the sake of completeness, the attested and unattested orders are listed in (5) (Cinque, 2005).

(5) a. Attested orders (14): Dem>Num>Adj>N, Dem>Num>N>Adj, Dem>N>Num>Adj, N>Dem>Num>Adj, Adj>N>Dem>Num, N>Adj>Dem>Num, Dem>Adj>N>Num, Dem>N>Adj>Num, N>Dem>Adj>Num, Num>Adj>N>Dem, Num>N>Adj>Dem, N>Num>Adj>Dem, Adj>N>Num>Dem, N>Adj>Num>Dem

b. Unattested orders (10): *Num>Dem>Adj>N, *Num>Dem>N>Adj, *Num>N>Dem>Adj, *N>Num>Dem>Adj, *Adj>Dem>Num>N,

*Adj>Dem>N>Num, *Dem>Adj>Num>N, *Num>Adj>Dem>N, *Adj>Num>Dem>N, *Adj>Num>N>Dem

Given these data, Cinque provides an account in the generative tradition with the purpose of showing how this macroscopic variation emerges out of a very restricted set of possibly independent rules shaping the grammar of human language. Under this perspective, the fact that not all possible permutations are found is not accidental, but derives from the universal rules of phrasal combination which exclude certain orders and not others.

Cinque's analysis starts from three assumptions characterized by different levels of tenability. First, the order with prenominal modifiers (i.e. Dem>Num>Adj>N) reflects the universal hierarchical order. If the noun remains in situ, this order does not allow any permutation (cf. (5)b) and thus is assumed as the order of Merge. Second, all syntactic movements involve a subtree containing the noun and target a c-commanding position. This is a stipulation that essentially disallows remnant movement within the DP. Third, the X-bar structure has the rigid organization proposed by Kayne (1994)[1], namely Spec>Head>Complement (cf. Section 1.3.1.). This represents one of the general assumptions adopted in the generativist framework. Besides, each nominal modifier comes with two projections: a lower one which hosts the modifier and a higher one which is a functional agreement projection and is the target of DP-internal movements. The schema of the DP structure assumed by Cinque is illustrated in Figure 1.

1 Kayne's (1994) analysis is based on the Linear Correspondence Axiom (LCA). This principle states that the hierarchy of non-terminals is linearly reflected in the order of terminals. This means that there is a strict correspondence between the hierarchical structure and the linear ordering of words.

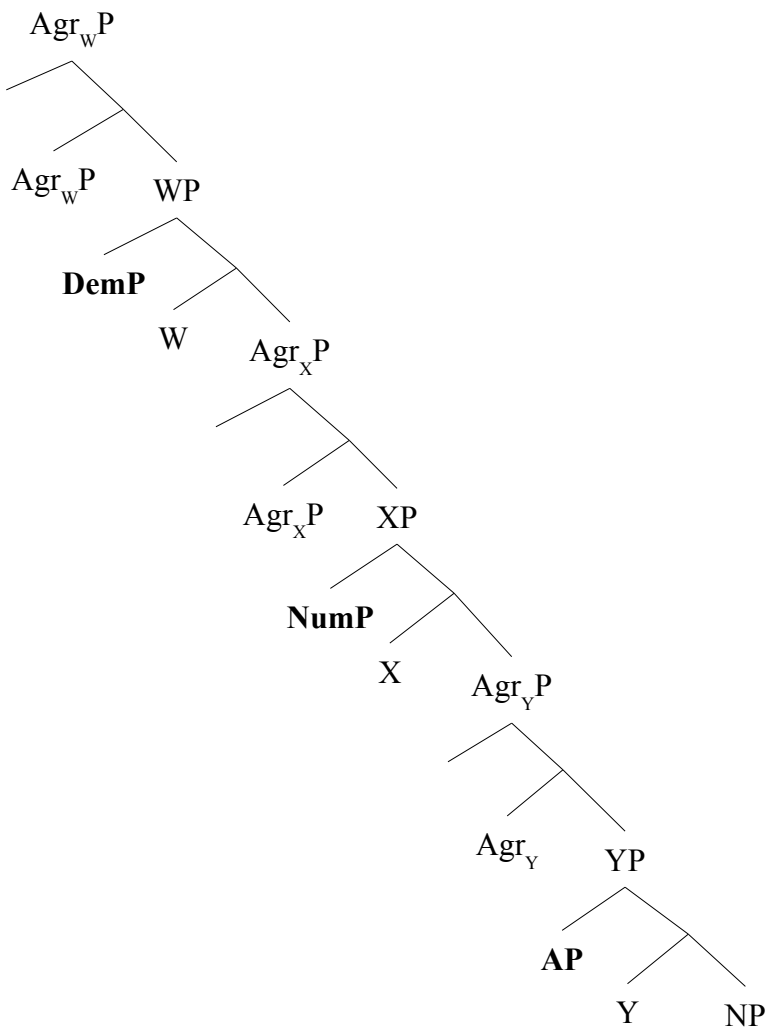


Figure 1: DP structure according to Cinque (2005)

Given that Dem>Num>Adj>N is assumed as the first Merge order (i.e. base generation), Cinque derives the other 13 attested orders listed in (5)a by different movement operations, targeting different portions of the structure. The available movement operations are listed in (6). The frequency at which the attested orders occur is predictable on the basis of the nature of the movements involved (marked vs. unmarked).

(6) a. No syntactic movement

[XP [NP]]

This is an unmarked option

b. NP-movement with pied-piping of the 'whose picture' type

$[[NP\ [XP]]_i \ldots [\ldots t_i]]$

This is an unmarked option

c. Simple NP-movement (without pied-piping)

$[[NP]_i \ldots [XP\ t_i]]$

This is a marked option

d. NP-movement with pied-piping of the 'picture of whom' type

$[[XP\ [NP]]_i \ldots [\ldots t_i]]$

This is a more marked option

The NP or the portion of structure containing the NP may raise all the way up (i.e. total movement) or partially (i.e. partial movement). The former option is unmarked, whereas the latter is marked.

This movement model allows also to predict the unattested orders in (5)b. The two possible violations are: movement of portions of the structure not containing the noun (e.g. * $[Adj]_i$ Dem Num t_i N), and movement of a non-constituent (e.g. * [Num N] Dem Adj).

Abels and Neeleman (2009) argue that Kayne's (1994) Linear Correspondence Axiom (LCA) is not necessary to explain the word-order asymmetries of Universal 20. In other words, they claim that none of the impossible permutations in (5)b is excluded by a rigid Spec>Head>Complement order of Merge. The assumptions that Abels and Neeleman (2009) start from are identical to the ones mentioned in Cinque (2005) with the only exception of antisymmetry. So, the DP structure that they use in their argumentation can be either leftward or rightward branching. The hierarchy they assume is architecturally simpler than Cinque's because it does not include functional projections intervening between the modifiers. The possibility to have more than one specifier in each projection provides the landing site for moved constituents. The attested orders that cannot be captured by base-generation are obtained by applying one single movement. The result is a theory which is more liberal on the generation of

syntactic structures, but it is more restricted in terms of movement. The resulting schema can be observed in Figure 2.

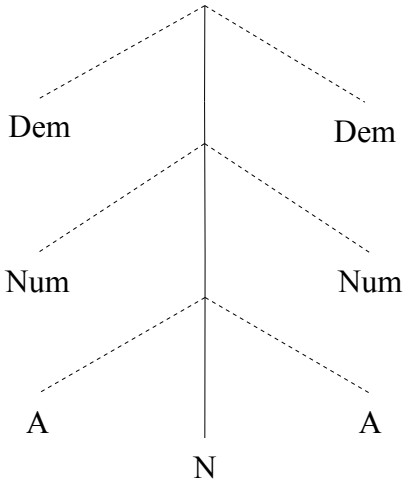


Figure 2: DP structure according to Abels and Neeleman (2009)

Abels and Neeleman (2009) obtain the same results as Cinque (2005), namely the same 14 attested orders and the same 10 unattested orders in (5). However, their analysis is different in that 8 attested orders are base-generated and do not need the application of movement operations. The other 6 attested orders are derived by leftward movement, which is motivated by parsing requirements (Abels & Neeleman, 2006). This symmetrical analysis leads the two authors to the conclusion that LCA is not a strict requirement to derive Universal 20 and therefore to capture Cinque's results.

Despite appearances, Cinque's (2005) proposal and Abels and Neeleman's (2009) proposal are empirically almost equivalent (so much so that a mechanical procedure can transform Cinque's tree into Abels and Neeleman's tree and vice versa, see Abels & Neeleman, 2006: sec. 4.1). However, this dissertation follows the antisymmetric analysis proposed by Cinque because in more recent works (Cinque, 2010, 2012) he proposed a finer-grained structure accounting for a more complex empirical domain than the one emerging from Universal 20. Currently, it is not obvious whether Abels and Neeleman's (2006) machinery can account for that pattern.

2.2.2. Recent typological developments

Word order facts on the distribution of demonstrative, numeral, and adjective in the nominal domain inspired some recent typological studies, such as Cysouw's (2010) work and Dryer's (2011) alternative to Cinque's (2005) analysis.

Cysouw's (2010) proposal is based on probability of occurrence and statistical preferences. In general, his perspective on typological data departs from Cinque's (2005) and Abels and Neeleman's (2009) formalizations, specifically with respect to the attested/unattested opposition. According to Cysouw, this dichotomy is not reliable since "it is highly dependent on the selection of languages investigated" (2010: 256). A particular linguistic configuration might be attested in a language, but if this language is not included in the sample under investigation, then it will be judged as unattested leading to wrong conclusions. The wrong conclusion would be considering the unattested pattern impossible, and hence ungrammatical. Actually, the fact that one particular order is not attested in any world's language does not necessarily mean that it is an impossible order. Besides, impossible does not necessarily mean ungrammatical: a particular sequence of items might not be produced by speakers because of high processing costs, rather than grammatical proscriptions.[2] What Cysouw suggests is to shift attention from possibility to probability. Hence, counter-examples and exceptions are not problematic in Cysouw's analysis, they are simply considered as less probable options. Indeed, working on Dryer's (2006) sample, Cysouw (2010: 272) reports that there are six languages attesting orders claimed to be impossible under Cinque's (2005) framework.[3]

With respect to the case study on the distribution of demonstrative, numeral, and adjective, Cysouw (2010) compares different models, including Cinque's, and looks for the one offering the best generalizations. According to his view, the optimal model is measured on the basis of the following criteria: the possibility to cover a great number of languages and the balance between accuracy of predictions and model weight.[4]

2 A typical case that causes extreme processing difficulty is represented by central-embedded relative clauses (for further details see Miller & Chomsky, 1963).

3 Notice that some of the data in Dryer''s (2006) sample, which are claimed to be problematic for Cinque's theory, are analyzed as cases of reduced relative clauses. See also footnote 6.

4 Simplifying a bit, 'weight'' is intended as the measure of theoretical complexity required to derive a generalization. If I understood correctly, when a generalization can be captured by two different models, the simpler model (the one requiring less assumptions or linguistic operations) is superior. Notice that this criterion favors descriptive adequacy over explanatory adequacy.

However, the choice of the final model may depend on the researcher's priority: highly accurate predictions or lower model weight. According to Cysouw, Cinque's model, although fairly accurate, does not result in the best option because of its heaviness (i.e. number of assumptions on the hierarchical structure and on linguistic operations).

After comparing different models, Cysouw singles out the one he considers the best. This model combines the following strong points: low residual deviance (i.e. only few observed options cannot be accounted for), low weight (i.e. simpler models are considered more elegant), and good accuracy (i.e. the more accurate the predictions, the better the model). The statistical preferences characterizing this model are listed in (7).[5]

(7) a. N and Adj are adjacent

b. Dem occurs at the edge of the DP

c. N occurs at the edge of the DP

d. N is followed by Adj

The basic idea behind Cysouw's (2010) approach is that a complex linguistic phenomenon can be analyzed by choosing the best payoff between empirical coverage and complexity of the theoretical model. As noticed in footnote 4, this line of reasoning would produce a reasonable descriptive adequacy without considering the burden of explanatory adequacy. Indeed, Cysouw explicitly claims that exceptional data do not need to be explained (Cysouw, 2010: 261). As long as frequency matters, they are only less likely options. However, the task of a linguistic enterprise should go beyond descriptive adequacy looking for deeper explanations which should also be compatible with models of language acquisition (Chomsky, 2004). Under this view, models in which generalizations are independently derived by the mechanisms of language generation are in principle to be preferred to those based on frequency accounts. In the case of Universal 20, we can clearly see how a generative model like Cinque's derives the generalizations in (7). Specifically, (7)a , (7)b, and (7)c are captured by Cinque's assumption on the universal order of Merge. Syntactic movement (an independently

5 If a language adheres to the four generalizations in (7), the order N>Adj>Num>Dem is instantiated. Indeed, this is the most frequent order according to Dryer (2011).

motivated operation within the theory) derives the generalization in (7)d. It seems that the heaviness (in terms of Cysouw, 2010) introduced by the adoption of a structural assumption is compensated by the reduction on the number of assumptions needed. The fact that this assumption is somehow motivated in Cinque's work further adds explanatory power to the generalization emerging from the data. All in all, although the weight assigned by Cysouw (2010) to Cinque's structural assumption might be higher than the weight characterizing the generalizations in (7), the explanatory adequacy of Cinque's model provides a more comprehensive linguistic account.

The analysis discussed by Dryer (2011) is based on a sample of 341 languages. First of all, he argues that some of the ten orders that are claimed to be unattested by Cinque (2005) (cf. (5)b) are indeed attested. These orders, reported in (8), are found in very few languages.[6]

(8) a. Num>N>Dem>Adj (Katu, Kilivila, Teop, Drehu, Yapese)

b. Dem>Adj>Num>N (Dhivehi, Ingush, Chechen)

c. N>Num>Dem>Adj (Haya)

In the light of Cinque's (2005) analysis, the data in (8) are unexpected. The order in (8)a is underivable because movement involves a non-constituent: according to the universal order of Merge, numeral and noun do not form a constituent (unless the adjective is also included), hence cannot undergo syntactic movement. The order in (8)b cannot be derived because movement involves a portion of the structure that does not contain NP: in this case, only the adjective undergoes movement and the head noun remains in its Merge position. Moreover, if no movement occurs, the three modifiers are in the wrong order of Merge. Similarly to (8)a, the order in (8)c is underivable because movement involves numeral and noun which do not form a syntactic constituent.

Generally speaking, Dryer's (2011) account departs from the ideas proposed by Cinque

6 With respect to Dryer's (2011) data and analysis, Cinque (2013a) argues that the three orders in (8) are probably not genuinely attested. For example, it is important to assess whether the adjective is of the direct modification type or it is rather derived from a reduced relative clause (Cinque, 2010). Full discussion on these special cases will be considered in future works (see Cinque, in preparation).

(2005) because he proposes that nominal modifiers should not be considered as syntactic categories, but rather as semantic categories. The author points out that in the world's languages the notion of demonstratives, numerals, and adjectives may be realized syntactically in different ways. For example, numerals may be realized as a distinct word class, as a subclass of adjectives, as verbs included in relative clauses, etc. Dryer claims that the different syntactic realizations are not relevant, therefore word order generalizations should rely on semantic categories instead. Consequently, Dryer (2011) suggests that the attested orders, including the three patterns claimed to be unattested in Cinque (2005), cannot be explained through a syntax-based analysis. His alternative approach consists of five independent principles, reported in (9) (Dryer, 2011: 6).

(9) a. Symmetry Principle I: "the Adj and the Num tend to occur closer to the N than the Dem when they (the Adj and the Dem and/or the Num and the Dem) occur on the same side of the noun."

b. Symmetry Principle II: "the Adj tends to occur closer to the N than the Num when they occur on the same side of the N."

c. Asymmetry principle: "the Symmetry Principles apply more strongly to prenominal modifiers than they do to postnominal modifiers; exceptions to the Symmetry Principles will occur only with postnominal modifiers."

d. Greenberg's Universal 18: "when the Adj precedes the N, the Dem and the Num, with overwhelmingly more than chance frequency, do likewise."

e. Intracategorial Harmony: "the Dem, Num, and Adj tend to all occur on the same side of the N."

According to Dryer, the number of violations to these principles can predict how frequent word orders are. Zero violations are expected to correspond to 44 genera (i.e. language families) or more, one violation to 4-22 genera, two violations to 3 genera or less. If a given word order violates 3 or 4 principles, then Dryer's (2011) analysis predicts no corresponding languages. As Dryer admits, his model does not yield correct

predictions for a few word orders. For example, the order Dem>Adj>N>Num violates two principles, namely (9)d and (9)e, and hence it is predicted to be attested in three or less language genera, but it is actually attested in six genera in Tibeto-Burman and New Guinea (Bodic, Kuki-Chin-Naga, Mirish, Eastern Highlands, Pawaian, Middle Sepik). The order Adj>N>Dem>Num is predicted to be unattested since it violates three of Dryer's principles, namely (9)a, (9)d and (9)e, but it is actually found in two language genera (Ubangi, Bai).

Taking into account Dryer's (2011) view on Universal 20, three considerations are here discussed. First, semantics may be influencing the position of modifiers in the structure. For instance, demonstratives may be high in the structure due to compositionality reasons. However, this would mean that demonstratives are in a peripheral position at Logical Form, but not necessarily in Surface Structure. Therefore, semantics may somehow exert an influence but it cannot explain the whole picture. Second, the fact that in some languages the modifiers involved in Universal 20 are categorized differently is not a point against Cinque's (2005) model. It simply shows that, in some cases, Universal 20 does not apply tout court because some relevant category is missing. Third, the criteria proposed by Dryer in (9) are not semantic at all. Rather, they are generalizations pertaining to the syntax-phonology interface. Indeed, they can be quite easily derived within the generative framework. As similarly discussed with respect to Cysouw's analysis, the explicatory adequacy of the model proposed by Dryer is inferior to Cinque's model.

2.3. Fine-grained nominal modification

The internal structure of nominal expressions has been investigated from a generative perspective in many studies. In what follows, I do not intend to offer an historical excursus on the major studies that offered insights into this research area.[7] Rather, I offer the necessary background information to set the theoretical context of the case studies that will be presented in Chapters 4 and 5.

7 For a more detailed discussion on this topic, the reader is referred to Coene and D'Hulst (2003) and Alexiadou, Haegeman, and Stavrou (2007).

First, I present the extended projection of the NP (in the sense of Grimshaw, 1991) showing the fine-grained syntactic structure characterizing it (a.o. Cinque, 2005, 2012). This section prepares the theoretical ground for the quantitative analysis on nominal modification (cf. Chapter 4). Then, I focus on the syntax of quantified expressions, in particular of cardinal numerals. Cardinaletti and Giusti's (2006) proposal is presented. This analysis will be adopted in the qualitative analysis of cardinals in LIS (cf. Chapter 5).

2.3.1. The extended projection of the NP

The parallelism between DP structure and clause structure originally discussed by Abney (1987), is further developed by Grimshaw (1991) who suggests that both NP and VP are dominated by a number of functional projections. The syntactic structure containing the projection of a lexical head, be it a N or a V, and the functional projections associated to it represents an extended projection. Therefore, in the spirit of Grimshaw (1991), the extended projection of the NP is to be intended as the array of functional projections whose heads are projected by the lexical head (N) at the bottom of the structure.

This structural layout can be observed in Figure 1, where I presented Cinque's (2005) syntactic hierarchy containing demonstrative, numeral, and adjective. For ease of reference, it is linearly represented in (10). For the sake of convenience, the labels referring to the functional projections hosting the modifiers are not specified.

(10) [Dem … [Num … [Adj … [N]]]]

As mentioned before, Cinque's idea is that all modifiers enter a fixed syntactic structure resulting from repeated applications of Merge. Applying the same principles illustrated in Section 1.3.3., Cinque assumes the hierarchy in (10) as the underlying hierarchical structure shared by all human languages. The variety of orders that superficially appear in world's languages is derived via internal displacement of the head noun or larger

portions of the structure containing the noun.[8] This assumption is motivated in Cinque's (2005) analysis by the fact that when the noun remains in its Merge position at the bottom of the extended projection, no word order permutations are allowed (e.g. *Num>Adj>Dem>N). In other words, if all three nominal modifiers occur in prenominal position, the order of Merge must be obeyed (Dem>Num>Adj>N). Given this observation, Cinque assumes that syntactic movement within the nominal domain involves the noun or a portion of structure containing it.

At the end of his paper, Cinque (2005) tentatively expands the hierarchy by adding four other modifiers, namely universal quantifier, ordinal numeral, numeral classifier, and relative clause (RC). The expanded hierarchy is provided in (11).

(11) $[Q_{univ} \ldots [Dem \ldots [Num_{ord} \ldots [RC \ldots [Num_{card} \ldots [Cl \ldots [Adj \ldots NP]]]]]]]$

Under the cartographic approach (cf. Section 1.3.3.) and on the basis of crosslinguistic considerations, Cinque (2012) proposes an even more refined hierarchy. This is shown in Figure 3. In this detailed hierarchy, each nominal modifier is hosted in dedicated functional projections. Notice that the structure contains both phrases (QP, DemP, AP, etc.) and heads (Det°, Number°, Diminutive°, Pejorative°, Endearment°).

8 The various linear orders attested in the world do not depend on semantic constraints. However, the intralinguistic variation may be motivated by the presence of additional semantic features, as I will show in Chapter 5.

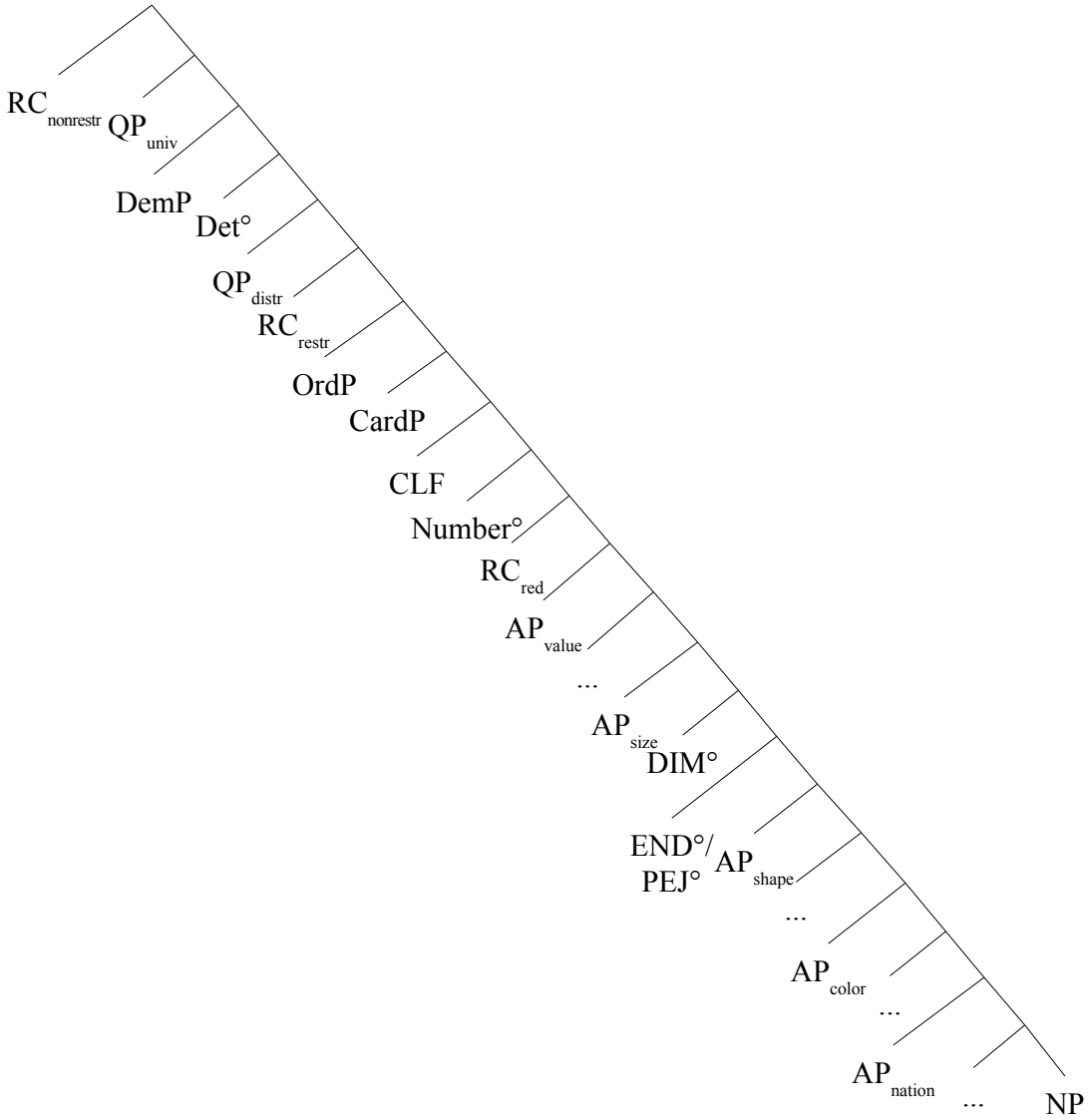


Figure 3: Cartographic hierarchy of the NP extended projection (Cinque, 2012)

In the remainder of this section, Cinque's (2012) hierarchy is discussed in a top-down fashion by presenting empirical evidence in support of it.[9]

As proposed by many scholars (see Sportiche, 1988; Cardinaletti & Giusti, 1992, 2006), quantifiers occur in DP-external positions. According to Cardinaletti and Giusti (2006), quantifiers are in a dedicated projection (Quantifier Phrase, QP) embedding the DP. This

9 I do not focus on all the modifiers included in Cinque's (2012) hierarchy, but only on the modifiers relevant to the quantitative analysis on the distribution of modifiers in LIS (cf. Chapter 4). For instance, the discussion in this section does not focus on the fine-grained classification of RCs (i.e. restrictive, non-restrictive, and reduced RCs) since it has not been further investigated in this thesis. Currently, there is no general consensus on the status and structure of RCs in LIS (for further discussion the reader is referred to Cecchetto, Geraci & Zucchi, 2006 and Branchini & Donati, 2009).

is shown in the structure in (12).

(12) [$_{QP}$ [$_{Q}$ all] [$_{DP}$ [$_{D}$ the] [... [$_{NP}$ students]]]]

This analysis is supported by the Italian data in (13). The equivalent structure of (12) is given in (13)a. In the sentence in (13)b (Cardinaletti & Giusti, 2006: 40), the Italian universal quantifier 'tutti' selects the clitic pronoun 'li' which undergoes clitic extraction leaving the quantifier in situ.

(13) a. [$_{QP}$ [$_{Q}$ tutti] [$_{DP}$ [$_{D}$ gli] [... [$_{NP}$ studenti]]]]

b. Li ho visti [$_{QP}$ tutti [$_{DP}$ ~~li~~]]

(I) them have seen all

If we assumed a DP-internal position for quantifiers, their floating behavior could not be accounted for. For more details on Cardinaletti and Giusti's analysis (2006), I refer the reader to Section 2.3.2.

Early proposals on the position of determiners were mainly based on English data. It was commonly assumed that article and demonstrative, being in complementary distribution in English (* this the house; * the this house), occur in the same syntactic position (D). Giusti (1993) however shows that in some languages, article and demonstrative may actually co-occur, as in (14).

(14)	a. afto	to	vivlio	Modern Greek
	this	the	book	
	b. ika	n	anak	Javanese
	this	the	child	
	c. ez	a	fiú	Hungarian

this the boy

These data show that the two elements are not in competition for one and the same syntactic position. Besides, as reported in Alexiadou, Haegeman, and Stavrou (2007), demonstratives constitute a universal category (i.e. they are present in all the languages of the world), whereas articles show great crosslinguistic variation. Some languages use articles, some others do not (see Bošković, 2009 for the properties of languages without definite article). Giusti (2002) argues that articles are functional heads: if present, they can be realized either as free dummies or as inflectional morphemes. Also, Giusti points out that articles cannot be merged in the structure if the lexical head is absent, as in (15).

(15) a. * I saw the English

b. * Ho visto il Italian

On the other hand, demonstratives are modifier-like elements and convey semantic referential features. They are claimed to occupy the Spec position of a functional projection. Demonstratives can be used either transitively with an NP complement or intransitively without any NP complement, as exemplified in (16).

(16) a. I saw this English

b. Ho visto questo Italian

As shown in Figure 3, Cinque assumes that numerals occupy a DP-internal position. The examples in (17) show that, in English, numerals are in an intermediate position between demonstrative (or definite article) and adjective.

(17) a. These three red cars are broken

b. All the five senses are important

In the hierarchy, ordinal numerals precede cardinal numerals. Fassi-Fehri (1999, as cited in Shlonsky, 2004: 1481) shows that when in Standard Arabic both types of numerals precede the noun, the relative order is Ord>Card, as in (18).

(18) a. ʔawwal-u xams-i muħaadaraat-in Arabic
first-NOM five-GEN lectures-GEN
'the first five lectures'

b. * xams-u ʔawwal-i muħaadaraat-in Arabic
five-NOM first-GEN lectures-GEN

Notice that the examples in (17) and in (18) show cases of cardinals that appear in definite environments. According to Cardinaletti and Giusti (2006), these are instances of a class of DP-internal modifiers, called quantitative adjectives. They should be kept distinct from cardinals used to quantify referents (e.g. 'three cars'), which sit in a DP-external position. The distinction between these two types of cardinals will be addressed in Section 2.3.2.

Looking at adjectives, there has been a lively debate on the distribution of these elements with respect to the head noun. I will shortly present Cinque's (2010) account in which it has been proposed that adjectives come in two flavors: direct modification and indirect modification. These two types of adjectives are associated with different syntactic and semantic properties. For example, adjectives of the direct modification type are non-restrictive, non-intersective, and have an individual-level reading. Adjectives of the indirect modification type are restrictive, intersective, and have a stage-level reading. On the basis of these differences, Cinque claims that adjectives have two different structural sources, as represented in Figure 4.

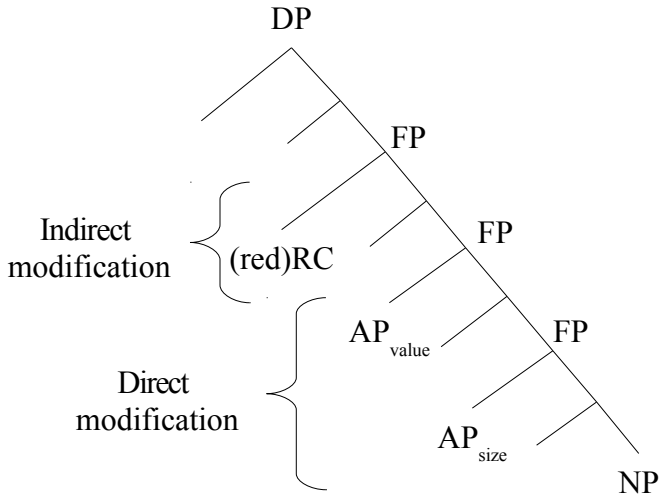


Figure 4: Two structural sources for adjectives (Cinque, 2010)

Direct modification adjectives (e.g. $AP_{value,}$, AP_{size}, etc.) are merged in the specifier position of various dedicated functional projections. Indirect modification adjectives are merged in a higher position and enter the DP as reduced relative clauses. As noted by Sproat and Shih (1991: 565), direct modification shows adjectival ordering restrictions as in (19) in which shape has to follow size, whereas indirect modification is characterized by freedom of order as in (20).

(19) a. small square table

b. * square small table

(20) a.the table [that is small] [that is square]

b. the table [that is square] [that is small]

The complete hierarchy that adjectives in direct modification must conform to is provided in (21) (Sproat & Shih, 1991: 565).[10]

(21) a. quality/value>size>shape>color>nationality

10 For an even more refined hierarchy accounting for adjectival ordering restrictions, see Scott (2002).

b. beautiful big round white Indian vases

It is worth noting that the extended projection illustrated in Figure 3 does not include four classes of modifiers that have been considered for the quantitative analysis of LIS data discussed in Chapter 4. These four categories are: i) possessives; ii) genitives (intended as lexical possessors, à la Alexiadou et al., 2007); iii) higher adjectives; iv) relational (or classificatory) modifiers.

In Cardinaletti (1998), possessive modifiers are classified into three groups. On the basis of their morphosyntactic properties, they can be classified as clitic, weak, or strong possessive modifiers. Cardinaletti's proposal involves a unique position in which possessives are base-generated. The three categories of possessives differ from each other in terms of syntactic derivation. Strong possessives can remain in their Merge position. Weak possessives are moved to the specifier position of a projection below D. Clitic possessives undergo head movement and move to D. These elements are in complementary distribution with articles. Alexiadou et al. (2007) give an outline of the main proposals on possession (including Cardinaletti, 1998). The items included in this analysis are: i) strong possessives as (22)a; ii) weak possessives as (22)b; iii) clitic possessives as (22)c; iv) genitives as the English example in (22)d.

(22) a. il cane suo — Italian
the dog his

b. il suo cane — Italian
the his dog

c. son chien — French
his dog

d. John's dog — English

Alexiadou et al. (2007) show that the distribution of these items can be accounted for involving different syntactic positions. Under this framework, possession is established

in a low position corresponding to nP. Depending on their typology, possessive modifiers can stay in situ or undergo syntactic movement. Strong possessives remain in their Merge position in the specifier of nP. Weak possessives undergo XP movement targeting the specifier position of AgrP, a functional projection above NumP. Clitic possessives undergo head movement and are cliticized to the head of the DP. Genitives undergo XP movement targeting the specifier position of the DP. Figure 5 shows how each possessive modifier is accommodated in the structure (adapted from Alexiadou et al., 2007: 575).

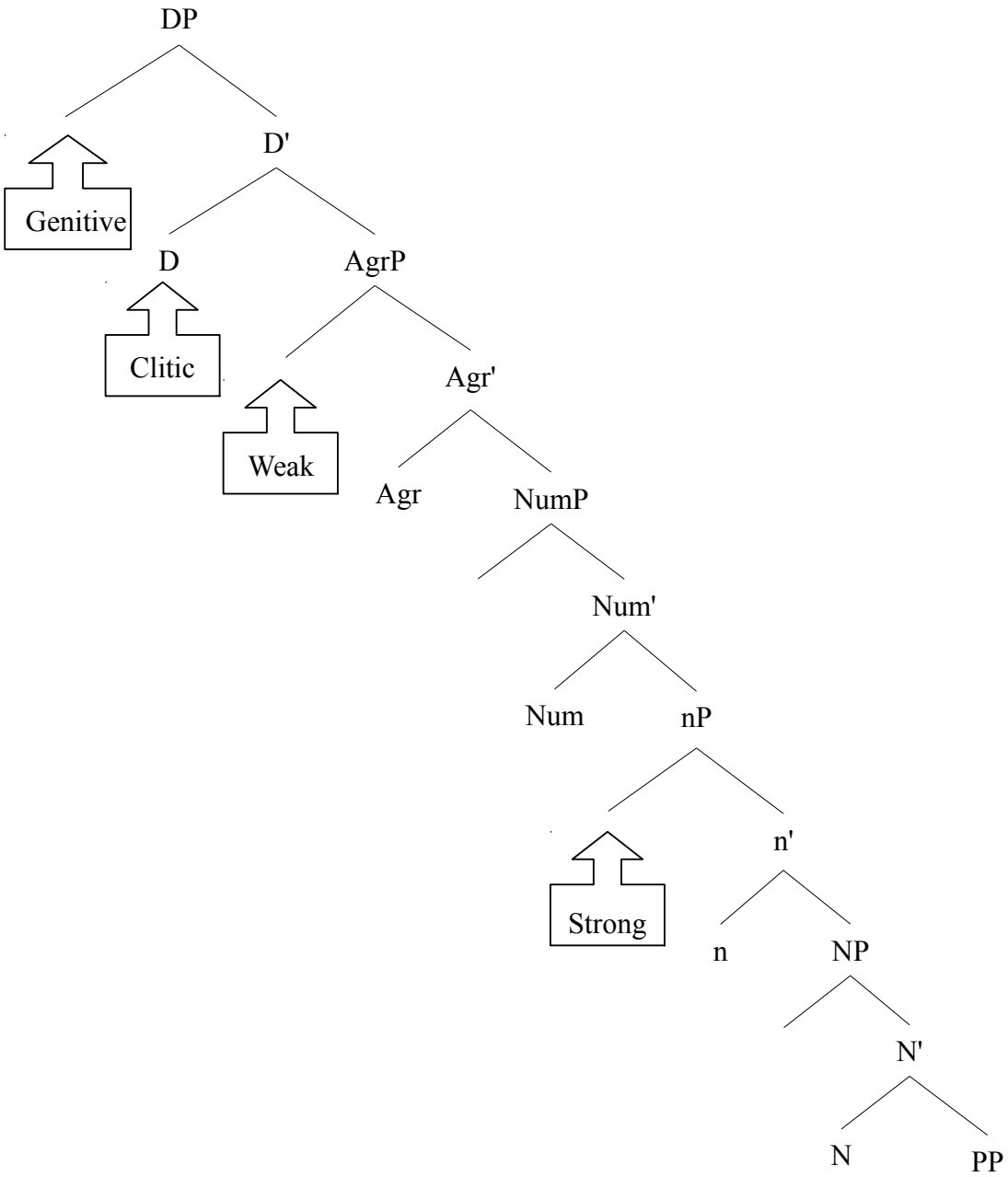


Figure 5: Positions available for possessive modifiers (Alexiadou et al., 2007)

This proposal can be made compatible with Cinque's (2005, 2010, 2012) framework assuming that the syntactic positions illustrated in Figure 5 are specialized functional projections in which possessive elements are base-generated. The possibility to move the head noun, rather than the possessive, across these positions allows accounting for language variation.

In some languages, possessive and (prepositionless) genitive may co-occur in the same

nominal expression. (23)a (Janssen, 1975, as cited in Alexiadou et al., 2007: 594) and (23)b (Ramat, 1986, as cited in Alexiadou et al., 2007: 594) provide two examples of this particular configuration which is known as possessor doubling construction.

(23) a. Peter z'n kat — Dutch

Peter his cat

b. Epkema syn plan — Frisian

Epkema his plan

It is worth noting that the proposal according to which different types of possessive elements are base-generated in different functional projections can successfully accommodate all the lexical entries involved in the possessor doubling construction.

Anticipating what will be discussed in Section 4.2.1., corpus data will reveal that in LIS there are two types of possessives showing a different syntactic behavior. Their distribution will be addressed by assuming two different syntactic positions.

Turning to higher adjectives, the modifiers to be included in this subclass of adjectives are 'other', 'same', 'only', 'following', and 'previous'. The label 'higher' is due to the fact that they usually precede other adjectives in the nominal domain, as shown in (24).

(24) a. The other beautiful pictures are in the drawer

b. * The beautiful other pictures are in the drawer

Therefore, they are assumed to be structurally higher than other adjectives.

Relational (or classificatory) modifiers (e.g. 'technical', 'international', 'political', 'pulmonary', etc.) are very close to the noun. Usually other modifiers cannot intervene between relational modifiers and noun, as in (25)a'. Also, relational modifiers are not gradable, as in (25)b, and cannot be used as predicates in copular sentences, as in (25)c. For more details on the morphosyntactic properties of relational modifiers, see McNally

and Boleda (2004).

(25) a. a pulmonary disease

a' * a pulmonary rare disease

b. * a more pulmonary disease

c. * this disease is pulmonary

Given the close relation between relational modifier and noun, the former is assumed to occupy a low position in the DP structure, possibly adjacent to the noun.

2.3.2. The syntax of quantified expressions

With respect to the syntactic position of cardinals, I will adopt the analysis in Cardinaletti and Giusti (2006). In order to explain syntactic asymmetries in the distribution of cardinals in spoken languages, they propose that cardinals may sit in two distinct projections: a higher quantifier projection and a lower quantitative adjective projection. Cardinals in QP sit in a DP-external position, as illustrated in (26). The corresponding tree structure is provided in Figure 6.

(26) $[_{QP}\ [_{Q}$ two$]\ [_{DP}\ [_{D}\ \emptyset]\ [\ldots\ [_{NP}$ children$]]]]$

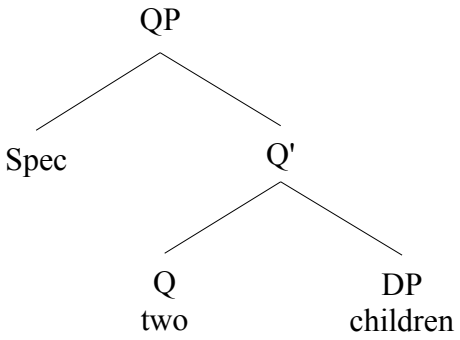


Figure 6: Position of cardinals functioning as quantifiers in the syntactic tree

In article languages, cardinals in QP display a number of syntactic properties. For the sake of clarity, they are listed and exemplified below.

(27) a. they can co-occur with pronouns

e.g. you two

b. they can occur in discontinuous position

e.g. ne ho visti tre, '[I] NE have seen three'

c. they can select a partitive PP

e.g. ho visto tre dei partecipanti, '[I] have seen three of the participants'

On the contrary, cardinals in the quantitative adjective projection (hereafter, QuantAP) sit in a DP-internal position, as illustrated in (28). The corresponding tree structure is provided in Figure 7.

(28) $[_{DP}$ the $[_{AgrP}$ $[_{QuantAP}$ two] … $[_{NP}$ children]]]

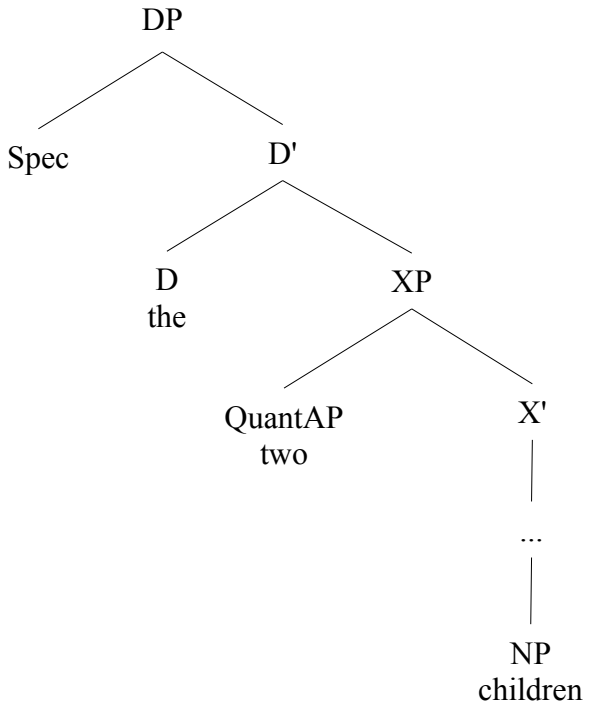


Figure 7: Position of cardinals functioning as quantitative adjectives in the syntactic tree

In article languages, cardinals in QuantAP display the syntactic properties listed and exemplified in (29).

(29) a. they are preceded by a determiner

e.g. I remember the two cities that are mentioned in the book

b. they cannot co-occur with pronouns

e.g. * you the two

c. they cannot appear in discontinuous position

e.g. * ne ho visti i tre, '[I] NE have seen the three'

d. they cannot select a partitive PP

e.g. * ho visto i tre dei partecipanti, '[I] have seen the three of the participants'

2.4. Nominal modification: the view from sign languages

Little work has been done with respect to Universal 20 in sign languages. This section presents some works that have been published on the internal structure of nominal expressions. The description of ordering patterns is proposed with a particular focus on the position of demonstrative, numeral, and adjective with respect to the noun. In sign languages, demonstratives are substantially realized by pointing signs or, alternatively, by other hand configurations directed toward a point in the signing space. In several cases, it is not straightforward to tell the difference between pointing signs functioning as demonstratives and those functioning as determiners. Despite their apparently iconic nature, numerals display a considerable degree of crosslinguistic variation. For a classification of numeral systems in crosslinguistic perspective, the reader is referred to Zeshan, Escobedo Delgado, Dikyuva, Panda, and de Vos's (2013) work which will be outlined in Section 5.2.

First, I report data coming from Hong Kong Sign Language (HKSL) (Tang & Sze, 2002), Taiwan Sign Language (TSL) (Zhang, 2007), American Sign Language (ASL) (MacLaughlin, 1997; Neidle & Nash, 2012), and Polish Sign Language (PJM) (Rutkowski, Czajkowska-Kisil, Łacheta & Kuder, 2014).

Then, I focus my attention on the case of LIS presenting the most representative works on nominal modification (Branchini, 2007; Bertone, 2007, 2009; Brunelli, 2011).

2.4.1. DP-internal order in sign languages

Tang and Sze (2002) discuss word order issues relative to the nominal domain in HKSL. These data, mainly coming from narratives, are crosslinguistically compared with respect to ASL, English, and Cantonese. Looking at complex combinations, Tang and Sze claim that the prevalent order in HKSL nominal expressions is Det>Num>N. However, the two authors point out that nominal modifiers in HKSL show a certain degree of variability since both pointing signs and numerals can appear either before or after the noun. The picture on HKSL with respect to Universal 20 appears incomplete since data on adjectives are not provided.

Definite nominal expressions are conveyed through the presence of a pointing sign and/or the presence of special non-manual markers (hereafter NMMs), usually eye gaze toward the location of the referent(s) in the signing space. Pointing signs with a definite reading can either be prenominal or postnominal, as shown in (30) (adapted from Tang & Sze, 2002: 299).

(30) a. [$_{DP}$ IX MALE] EAT-RICE

'That man eats rice'

b. [$_{DP}$ MALE IX] SLEEP

'That man is sleeping'

According to Tang and Sze, whether pointing signs with a definite reading are to be intended as determiners or demonstratives is an open issue.

Like pointing signs, numerals can either precede or follow the noun, as the examples provided in (31) show (adapted from Tang & Sze, 2002: 303).

(31) a. [$_{DP}$ HAVE THREE MALE] STEAL DOG

b. [$_{DP}$ HAVE MALE THREE] STEAL DOG

'Three men stole a/the dog'

On the basis of these data, modifiers included in nominal expressions in HKSL show considerable distributional freedom.

Turning to TSL, another sign language used in the southeast of Asia, I present Zhang's (2007) study. The author discusses TSL data relative to the nominal domain in the light of Cinque's (2005) antisymmetric analysis and Abels and Neeleman's (2009) symmetric analysis (cf. Section 2.2.1). The data for Zhang's study mainly consist in elicited sentences coming from the corpus project "A Study of Taiwan Sign Language: Phonology, Morphology, Syntax and Digital Graphic Dictionary" (see Lai, 2005). On

the basis of the data at her disposal, Zhang shows that TSL displays a considerable degree of order flexibility. If the noun is associated with one modifier, be it a demonstrative, a numeral, or an adjective, both modifier>N and N>modifier are possible orders, as exemplified in (32) (Zhang, 2007: 63).

(32) a. (IX_DEM) AIRPLANE (IX_DEM) Dem>N or N>Dem
'this plane'

b. (THREE) BOOK (THREE) Num>N or N>Num
'three books'

c. (CUTE) CAT (CUTE) A>N or N>A
'cute cats'

If the nominal expression contains two modifiers, three possibilities are attested: i) both modifiers precede the noun, as in (33)a; ii) one modifier precedes and the other follows the noun, as in (33)b and in (33)c; iii) both modifiers follow the noun, as in (33)d and in (33)e (Zhang, 2007: 63-68). The examples below include numeral and adjective but the same pattern is found with numeral and demonstrative and with adjective and demonstrative. The relative order of two modifiers occurring on the same side of the noun shows an interesting asymmetry. In the case of prenominal modifiers, the order is rigidly constrained in that only the universal order of Merge (cf. Section 2.2.1) is possible, as in (33)a. The reversed order is not accepted, as (33)a' shows. Such restriction is not found in the distribution of postnominal modifiers which can be arranged as in (33)d or in (33)e.

(33) a. FIVE CUTE CAT Num>A>N

a'. *CUTE FIVE CAT *A>Num>N

b. FIVE CAT CUTE Num>N>A

c. CUTE CAT FIVE A>N>Num

d. CAT CUTE FIVE N>A>Num

e. CAT FIVE CUTE N>Num>A

'five cute cats'

If the nominal expression contains all three modifiers, four order options are found. These are exemplified in (34) (Zhang, 2007: 65). While numerals and adjectives show a relatively flexible distribution, demonstratives can only be produced at the very beginning of the nominal expression.

(34) a. IX_DEM FIVE NAUGHTY BOY Dem>Num>Adj>N

b. IX_DEM NAUGHTY BOY FIVE Dem>Adj>N>Num

c. IX_DEM FIVE BOY NAUGHTY Dem>Num>N>Adj

d. IX_DEM BOY NAUGHTY FIVE Dem>N>Adj>Num

'these five naughty boys'

Zhang points out that different combinations do not yield any interpretation differences. Unfortunately, Zhang (2007) does not indicate whether some ordering options are attested only with specific NMMs. The possibility of arranging the modifiers according to different ordering patterns shows that a unique basic order does not emerge. Crucially, Zhang points out that even if TSL shows a considerable degree of intralinguistic variation, all the attested combinations in which more than one modifier is considered are derivable within Cinque's model.[11]

ASL is one of the most widely used sign languages and is the sign language that has received most attention in the linguistic literature. In order to describe word order issues relative to the nominal expressions in ASL, I present here data coming from MacLaughlin (1997), Boster (1996), and Neidle and Nash (2012).

According to MacLaughlin (1997), pointing signs occurring in the nominal domain

11 Order flexibility is typical of sign languages. Different distributional patterns are found also in LIS and, most importantly, they can all be derived by Cinque's (2005) analysis.

should not be uniformly accounted for. Only prenominal indexes convey the notion of definiteness, whereas postnominal indexes are adverbial modifiers and are not restricted in terms of definiteness (so they are not further discussed here). An example of prenominal pointing sign is given in (35) (adapted from MacLaughlin, 1997: 117).

(35) [$_{DP}$ IX MAN] ARRIVE

'The/that man is arriving'

On the basis of the English translation provided in (35) and MacLaughlin's discussion, it is not clear whether the prenominal pointing signs serve as articles or demonstratives. Neidle and Nash (2012) confirm that these determiners correlate with definiteness and propose that they are "somewhat intermediate between the definite article and demonstrative of English" (Neidle & Nash, 2012: 271). The two authors argue that when prenominal pointing signs are phonologically stressed, then the demonstrative reading is forced.

With respect to the distribution of cardinal numerals, both Boster (1996: 160) and MacLaughlin (1997: 133) affirm that they can either precede or follow the noun. The acceptability judgments reported in (36) (adapted from Boster, 1996: 160) show the two ordering possibilities.

(36) a. IX-1 WANT THREE BOOK

b. IX-1 WANT BOOK THREE

'I want three books'

To my knowledge, no differences in meaning have been reported, so apparently the two orders (Num>N, N>Num) are semantically equivalent.

Adjectives also show a certain flexibility (Boster, 1996; MacLaughlin, 1997). This is confirmed by the examples in (37) (adapted from Boster, 1996: 171).

(37) a. IX-1 WANT RED BOOK

b. IX-1 WANT BOOK RED

'I want the/a red book'

However, MacLughlin (1997) points out that these two order options are not equivalent. The author analyzes prenominal adjectives as attributive modifiers. When two or more than two adjectives precede the noun, ordering restrictions appear (for a more detailed discussion, see MacLaughlin, 1997: 193-194). This is the same pattern discussed for direct modification adjectives in Section 2.3.1. Moreover, some adjectives in ASL can only appear in prenominal position. This is the case of REAL, FORMER, BASIC. On the contrary, adjectives in postnominal positions do not show ordering restrictions and they are claimed to function as predicative modifiers (see indirect modification adjectives in Section 2.3.1).

Turning to the co-occurrence of demonstrative, numeral, and adjective, Neidle and Nash (2012) claim that ASL conforms to Universal 20. The relative canonical order of the three nominal modifiers is the one shown in (38).

(38) Dem>Num>Adj>N

The order in (38) reflects the universal order of Merge (cf. Section 2.2.1). Having all modifiers in prenominal position, ASL seems to show a head final pattern. However, as indicated by Neidle and Nash (2012), deviations from the canonical word order are also found. The examples shown above confirm that.

In addition to these nominal modifiers, the distribution of possessive modifiers and genitives has been investigated. According to Abner (2012), ASL possessives basically function as verbal predicates. Similarly to transitive verbs, they express verbal agreement spatially. Indeed, possessives move from the location associated with the possessee to the location associated with the possessor. When they are used as

attributes, as in (39)a, they enter the nominal structure via reduced RC formation. Abner's proposal is represented in (39)b (Abner, 2012: 76-79).

(39) a. BRUNO$_i$ POSS$_i$ BOOK

'A book of Bruno's'

b. [$_{DP}$ … [[$_{RedRC}$ ~~BOOK~~ BRUNO$_i$ POSS$_i$] [$_{NP}$ BOOK]]]

If the possessor is leaved unexpressed, as in (40)a, Abner proposes the same derivation via reduced RC with the only difference that the possessor position is occupied by a null pronoun ('pro'), as illustrated in (40)b (Abner, 2012: 80).

(40) a. POSS$_i$ BOOK

'A book of his'

b. [$_{DP}$ … [[$_{RedRC}$ ~~BOOK~~ pro$_i$ POSS$_i$] [$_{NP}$ BOOK]]]

What is relevant at this point of the discussion is that, according to Abner (2012), the possessive occupies a low position in the hierarchical structure.

Finally, I present some partial data relative to a European sign language, namely PJM. A corpus-based study (Rutkowski et al., 2014) reveals that the canonical surface order of nominal modifiers in PJM is the one reported in (41).

(41) Dem>Num>N>Adj

As the authors point out, this pattern is different from the one attested in Polish nominal expressions showing that the two languages do not share the same ordering rules (at least in the nominal domain).

The order in (41) shows that PJM conforms to Universal 20. The authors observe that,

following Cinque's (2005) analysis, this order can be derived by assuming NP-movement across the adjective. However, the co-occurrence of demonstrative, numeral, and adjective in a single nominal expression is not frequent in the corpus. Therefore, the canonical order reported in (41) is a tendency that has been reconstructed on the basis of the distribution of the annotated modifiers and of partial combinations. Rutkowski et al. (2014) also underline that various factors may intervene influencing the overall order (a.o. numeral incorporation, focalization, special uses of adjectives).

Before moving on giving an outline of previous works on the DP-internal order in LIS, I shall conclude this section by remarking the considerable distributional freedom characterizing all the four sign languages that have been considered.

2.4.2. DP-internal order in LIS

In this section, the most relevant findings on the DP in LIS are presented. In particular, data from Branchini (2007), Bertone (2007, 2009), and Brunelli (2011) are shown. The empirical base of the three studies consists in, to a large extent, elicited data. [12] Since the three authors are in substantial agreement, the core part of this section is based on the most extensive analysis of the DP in LIS, namely that of Bertone (2007, 2009). Additional data from Branchini (2007) and Brunelli (2011) are provided to complement the picture.

Bertone (2007, 2009) claims that nominal modifiers in LIS usually occur in postnominal position. The unmarked order is represented and exemplified in (42) (adapted from Bertone, 2009: 22).[13]

(42) N>Adj>Num>Dem

BOOK NEW TWO IX_DEM$_i$ MINE

12 Branchini (2007) also takes into consideration some spontaneous data provided by informants collaborating with the National Research Council (Consiglio Nazionale delle Ricerche, CNR) located in Rome.

13 As a methodological note, Bertone (2007, 2009), Branchini (2007), and Brunelli (2011) notice how difficult it is to observe the co-occurrence of all nominal modifiers in one and the same sentence. To bypass this issue and collect evidence about the relative order among the modifiers, partial combinations of these elements have been elicited until the total order was derived.

'These two new books are mine'

The order in (42), which contains the three nominal modifiers considered under Greenberg's Universal 20, is the one reported also in Branchini (2007) and Brunelli (2011), and it is enriched by including order information for other nominal modifiers. Specifically, they claim that: i) possessives usually appear between the noun and the adjective; ii) quantifiers tend to appear after all other modifiers. The revisited ordering schema is illustrated in (43) (after Branchini, 2007 and Brunelli, 2011).

(43) N>Poss>Adj>Num>Dem>Quant

Besides the unmarked order represented in (42), Bertone (2007) reports that other orders are also attested, namely the ones given in (44) with the corresponding examples (adapted from Bertone, 2009: 22-23). These orders are claimed to be infrequent and require additional prosodic marking, mainly tensed cheeks and raising eyebrows.

(44) a. N>Num>Adj>Dem

BOOK TWO NEW IX_DEM$_i$, MINE

b. Dem>N>Adj>Num>index

IX_DEM$_i$ BOOK NEW TWO IX$_i$, MINE

c. Dem>N>Num>Adj>index

IX_DEM$_i$ BOOK TWO NEW IX$_i$, MINE

'These two new books are mine'

Branchini (2007) points out that, although nominal modifiers are mainly postnominal as in (45) (adapted from Branchini, 2007: 48), some cases of prenominal modifiers are also attested as in (46) (adapted from Branchini, 2007: 49).

(45) CHILDREN$_i$ IX_DEM$_i$ SOCCER PLAY

'These children play soccer'

(46) IX_DEM$_i$ CHILDREN$_i$ SOCCER PLAY

'These/those children play soccer'

Another case of prenominal modifiers discussed in Branchini (2007: 51) is that of heavy NP complements[14] shown in (47). In these cases, it would be interesting to check the non-manual contour characterizing the nominal expression so to have a better grasp of its phrasal boundaries.

(47) a. [[[FRIEND$_k$ POSS] SISTER POSS$_k$] THREE] MARRY DONE

b. ?? [THREE [[FRIEND$_k$ POSS] SISTER POSS$_k$]] MARRY DONE

'Three of my friend's sisters are married'

The order of LIS modifiers shown in the canonical ordering in (42) is the mirror image of the universal (and underlying) order of Merge proposed in Cinque (2005). Both Bertone (2007, 2009) and Brunelli (2011) adopt Cinque's framework to account for these data. I briefly illustrate their analysis by deriving the surface order for the LIS example in (42). The derivation of the full structure proposed by Bertone (2007) and Brunelli (2011) requires three steps in which the syntactic operations of Merge and Move apply in turn repeatedly. These are summarized in (48). A simple NP movement derives the order N>Adj, as in (48)b, and two complex movements (i.e. movement of larger phrases by pied-piping of the 'whose picture'' type) derive the order N>Adj>Num and N>Adj>Num>Dem, as in (48)d and in (48)f, respectively.

(48) a. [$_{AP}$ NEW [$_{NP}$ BOOK]]

14 I think that this construction is better analyzed as the equivalent of the English genitive construction. Unfortunately Branchini (2007) does not provide further data to support this claim.

b. [$_{XP}$ [$_{NP}$ BOOK] [$_{AP}$ NEW ~~BOOK~~]]

c. [$_{NumP}$ TWO [$_{XP}$ [$_{NP}$ BOOK] [$_{AP}$ NEW ~~BOOK~~]]]

d. [$_{YP}$ [$_{XP}$ [$_{NP}$ BOOK] [$_{AP}$ NEW ~~BOOK~~]] [$_{NumP}$ TWO ~~BOOK~~]]

e. [$_{DemP}$ IX_DEM [$_{YP}$ [$_{XP}$ [$_{NP}$ BOOK] [$_{AP}$ NEW ~~BOOK~~]] [$_{NumP}$ TWO ~~BOOK~~]]]

f. [$_{ZP}$[$_{YP}$[$_{XP}$ [$_{NP}$ BOOK] [$_{AP}$ NEW ~~BOOK~~]] [$_{NumP}$ TWO ~~BOOK~~]]] [$_{DemP}$ IX_DEM ~~BOOK~~]

Neither of the two authors takes into consideration the cases of prenominal modifiers that have been briefly alluded to by Branchini (2007). Consequently, their derivation is expressly designed for postnominal modifiers and does not account for data like (46) and (47)b.

2.5. Conclusion

The internal structure of nominal expressions has received significant attention in the linguistic literature. This chapter offered the "state of art" on nominal modification by selecting the studies relevant to this dissertation.

First, I presented the tenets of typological Universal 20 showing the word order facts relative to the nominal head and its modifiers (demonstrative, numeral, and adjective). The generalizations emerging by Universal 20 have been discussed in the light of both typological and generative formalizations.

Second, I presented the fine-grained structure of the extended projection of the NP that can accommodate and account for the nominal modifiers considered in the next chapters. The discussion on each relevant projection has been supported by examples mainly from spoken languages.

In the third part of this chapter, I provided an overview of the main facts and findings concerning the DP-internal structure in sign languages. Apparently, HKSL, TSL, ASL, and PJM are languages displaying a significant distributional variation. Although it is

possible to observe combinational tendencies and formulate the respective canonical word order, intralinguistic variation has been found in all the four languages under investigation (to various degrees). Conversely, the existing literature on LIS nominal modification shows a consistent and regular pattern. Except for a few marked orders characterized by special prosodic contours, LIS nominal modifiers are claimed to be systematically arranged according to the mirror image of the order of Merge. Overall, the possibility to produce modifiers in prenominal position is an understudied issue.

In Chapter 4, I will show how corpus data call into question the idea that modifiers are systematically postnominal in LIS. Indeed, I will present data revealing the presence of both prenominal and postnominal modifiers. With respect to word order issues, this dissertation will show that LIS is rather similar to the other sign languages in that order flexibility is attested in the nominal domain. Crucially, order flexibility does not depend on absence of structure. Indeed, I will show that the distribution of nominal modifiers, far from being random, appears highly constrained by the syntactic options that natural languages allow.

3. Methodological issues

3.1. Introduction

In this chapter, methodological issues are presented. More specifically, I explain what kind of tools I used to conduct my analysis and the reasons that led me to choose such tools.

The methodology that I adopted involves two types of data: corpus data (free narration), and elicited data (acceptability judgments and narration tasks). Corpus data were taken from the LIS corpus project (Geraci, Bayley, Branchini, Cardinaletti, Cecchetto, Donati, Giudice, Mereghetti, Poletti, Santoro & Zucchi, 2010; Cardinaletti, Cecchetto & Donati, eds., 2011), while the other two sources of data were specifically collected for this dissertation.

Methodology is here illustrated through an overview of the main procedures involved: data collection, data annotation, and statistical analysis. With respect to data collection, I treat corpus data and elicited data separately in order to capture the specificities of each type of data. As for data coding and annotation, the procedure applied to corpus and elicited data is similar, therefore it is discussed without any specific differentiation according to the type of data. Finally, data extraction and statistical computing are described with special reference to corpus data since these procedures were applied to the quantitative analysis of data coming from the corpus. This quantitative approach was adopted for the two case studies that will be discussed in Chapter 4.

Section 3.2 describes how data collection has been conducted and what kind of linguistic tasks have been proposed to informants. In Section 3.3, I explain what kind of linguistic information I coded for and how I did it by using the annotation software ELAN. Section 3.4 deals with the quantitative analysis of corpus data: first, I explain how I obtained a dataset by extracting corpus data from the annotation files; second, I illustrate how I codified linguistic variation into statistical variables; third, I discuss how I imported the dataset in the statistical software R and how I conducted the statistical

analysis.

3.2. Data collection

The data used for this thesis are of two types: corpus data and elicited data.

On the one hand, corpus data represented the main basis for the quantitative analysis. This source of data has been considered primarily because it allows the investigation of linguistic variation in different sociolinguistic environments. Besides, the production of a large group of signers, rather than of only a few signers, may be more accurately representative of the larger Deaf Italian community. It also shows those facets that may depend on the various situations in which the language has been developed (see also Section 1.4.1.). The role of sociolinguistic factors can be evaluated on a large scale. So, corpus data coming from fluent signers, be they not necessarily native, provide a more accurate picture of the language in use. Another advantage of using corpus data, especially free conversations and spontaneous narratives, is that signers express themselves in a natural setting reducing at a minimum the risk of conscious control. However, what corpus data do not offer is negative evidence. This means that they cannot be used to understand what is disallowed in the language (also see Van Herreweghe & Vermeerbergen, 2012).

On the other hand, elicited data and narration tasks were included in this research work in order to develop qualitative considerations. They have been used to look for very specific constructions or long sequences of signs that would be difficult to observe in spontaneous settings. An advantage of using elicited data is that linguists have the possibility to collect negative evidence in the form of acceptability judgments. This would be the main source of evidence to elaborate grammatical theories (also see Van Herreweghe & Vermeerbergen, 2012).

An important issue that should be taken into consideration at this first methodological step is what Labov (1972) calls 'The Observer's paradox'. This phenomenon happens when the informant is aware that s/he is being observed by researchers and/or recording devices and consequently tends to produce more self-conscious responses. In order to

minimize this undesirable effect, a few expedients have been adopted in both the collection of corpus data and the elicited data that I specifically collected for this dissertation. Corpus data were collected in informal and familiar settings and at the constant presence of a local Deaf interlocutor. As for the other types of data, the signers involved are informants that have been collaborating with linguists for a long time. Given their experience in this field, the influence exerted by the presence of cameras and researchers is likely to be minimal.

3.2.1. Corpus data

The corpus data that have been considered for the quantitative analysis presented in this thesis come from the 'LIS corpus project'[1] (Geraci, Battaglia, Cardinaletti, Cecchetto, Donati, Giudice & Mereghetti, 2011; Cardinaletti, Cecchetto & Donati, eds., 2011). This is the largest corpus for LIS currently available. The project was funded by the National Research Fund PRIN[2] and involved deaf and hearing researchers coming from three Italian institutions: University of Rome La Sapienza, University of Milan Bicocca, and Ca' Foscari University of Venice. Local branches of the National Deaf Association (Ente Nazionale Sordi, ENS) have collaborated giving assistance especially during data collection. Some American and European consultants have shared their experience providing suggestions on technical issues. The procedures relative to data collection and data annotation have been inspired by previous studies on the sociolinguistic variation of ASL (Lucas, Bayley & Valli, 2001) and AUSLAN (Johnston & Schembri, 2006). However, some adjustments were required in order to adapt the project to the Italian context (Geraci et al., 2010).

Data collection has been conducted at the ENS local associations. Active members of the local communities have collaborated in selecting participants. The 165 native or near-native LIS signers who participated in the LIS corpus took part in four linguistic tasks. The tasks are described in (1).

(1) a. individual narration: each participant was supposed to talk about any personal

1 http://w3.uniroma1.it/progettolis/index.php (20 August, 2014)
2 PRIN 2007 project ('Dimensions of Variation in Italian Sign Language')

experience connected with his or her education, family, job, etc. As part of the procedure of data collection, participants were encouraged to produce short narratives (of about 5 minutes) in front of the contact person, namely a signer of the same local variety of LIS. This was done to minimize potential interference from non-local varieties of LIS and to avoid the unpleasant situation of a signer placed in front of a camera with nobody around.

b. question-answer elicitation task: this task has been specifically designed to test wh-questions. Two participants belonging to the same age group were encouraged to talk about a car accident. One of the two had access to the visual representation of the accident, the other had to fill in a word-less questionnaire in order to reconstruct the dynamics of the fact.

c. free conversation among three people: three participants belonging to the same age group were asked to have a conversation together for 45 minutes. No specific topic was provided by the local contact person. The production of each participant was filmed by a dedicated camera.

d. picture-naming task: each participant was shown 42 images and was asked to produce the corresponding LIS signs. The elicited signs were of different types: colors, months, family members, classifiers, initialized signs, fingerspelled signs, compounds, and signs subject to diachronic variation.

For the purposes of this research, I focused on the individual narration task analyzing data from 162 signers out of 165 (three signers who participated in the free conversation session did not participate in this specific task). I decided to do so because narration is probably the ideal place to look for both complex and simple DPs. Indeed, narration often includes detailed descriptions of situations, events, participants, etc., and nominal modifiers are more likely to be found in these linguistic situations.

Signers' productions have been recorded by digital cameras so that video recordings in mpg2 format were collected. At the end of each session, metadata have been collected through a questionnaire. Each signer was asked to answer questions about his or her age, provenance, educational background, etc. The metadata questionnaire is essential in order to code for sociolinguistic variables.

The 162 signers involved in the narration task were balanced with respect to the various social factors. The nine factors that have been considered are provided in (2).

(2) Social factors:

Geographical provenance, Residence status, Gender, Education, Job, Age, Age group, Family, Register

Having a perfectly uniform distribution of all social factors within the corpus is virtually impossible (Cecchetto, Giudice & Mereghetti, 2011). In order to obtain an internal balance as much as possible, the selection of participants followed three guidelines: i) for each city, participants should be equally divided into three age group (the optimal distribution according to age group consists of 6 young, 6 middle-aged, and 6 old signers for each city); ii) for each age group, both genders should be represented leading to an overall gender balance; iii) for each city, the group of native signers (i.e. with Deaf parents) should be represented. In this section, I do not show how the social factors are distributed with respect to the dependent variable. This will be discussed in Chapter 4 when I will present the results related to the analysis of corpus data. In this methodological chapter, I only describe what each factor refers to. Linguistic factors will be discussed in detail in Section 3.3.2.

Geographically, participants were recruited in 10 Italian cities. As shown in (3), the whole country is represented, form North to South.

(3) a. Northern Italy: Turin, Milan, Brescia, Bologna

b. Central Italy: Florence, Rome

c. Southern Italy: Trani, Salerno, Lamezia, Ragusa

The selection of these ten cities is motivated not only by their geographical location, but also by the presence of institutes for deaf pupils (e.g. Pavoni Institute in Brescia, Smaldone Institute in Salerno, Prinotti Institute in Turin, Gualandi Institute in Florence).

Another factor relative to the place where the participants come from is the residence status. This implies whether the participant lives in a rural area or an urbanized area. Living in a city tends to favor the adoption of new linguistic features because of higher circulation of people, goods, and ideas. On the contrary, rural areas usually display a more conservative attitude toward linguistic innovations. So, according to the factor residence status, participants are divided into two subgroups: signers living in urbanized areas and signers living in rural areas.

Gender is another social factor that has been taken into consideration. In principle, women and men could show some important peculiarities (Labov, 1990). On the one hand, men tend to use more nonstandard forms. On the other hand, women are more sensitive to prestige issues and use fewer stigmatized forms. With respect to LIS, variation according to gender may depend on single-gender education (now less frequent than earlier) and on social contacts which may encourage signers to adopt new variant forms or, conversely, to reject them.

The educational qualification can also affect the way signers express themselves. LIS corpus participants are divided into three subgroups: the ones who did not continue their education past the primary school, the ones who completed middle school, and the ones who have a high-level education since they completed either high school or a university program. The educational background may lead to language variation. For example, the fact of having a graduate degree usually entails the acquisition of competences as well as sector-specific knowledge that can have linguistic implications (e.g. access to Italian, exposure to academic registers, acquisition of technical jargon, etc.).

The socio-economic class of participants may be inferred on the basis of their occupation. The possibilities included in the questionnaire are the following: white collar, blue collar, professional, student, and unemployed. Notice that the level 'professional' includes also Deaf educators and LIS teachers. These job profiles require signers to use LIS in a professional environment. Therefore, these signers are more likely to express themselves with conscious control including also prestigious variant forms.

Possible diachronic variation can be detected by looking at age. This is a continuous numeric variable and may provide evidence for a linguistic change in progress.

In addition to the standard sociolinguistic information mentioned so far, information specific to the Deaf world was also collected.

One of the Deaf-specific predictors[3] that have been considered is age group. Participants have been divided into three groups: young signers (from 18 to 30 years old), middle-aged signers (from 31 to 54 years old), and old signers (over the age of 55). This classification reflects the special education policies that have been adopted in Italy in the last decades. Before 1977, deaf children used to attend special institutes for the deaf (i.e. residential schools). An important change was determined by the law 517/77 that established the possibility for disabled students, and hence deaf students too, to enroll in mainstream schools. In virtue of this freedom to choice, many families opted for mainstream schools. Consequently, the great majority of institutes for the deaf witnessed a gradual enrollment decrease and were forced to shut down. So, the three age groups reflect the evolution of the education system: old signers attended residential schools, middle-aged signers were in the middle of the transition from one system to the other, and the majority of young signers attended mainstream classes.[4]

Another important Deaf-specific predictor is the presence of Deaf members in the family. The possible values included in this factor are: hearing parents and relatives, hearing parents and at least one Deaf relative, Deaf parents. This social variable can say something about the type of language acquisition (cf. Section 1.4.4.). Differently from spoken languages, sign languages are characterized by typical language acquisition only in the case of a small minority. Indeed, only 5-10% of Deaf children are born to Deaf parents, and hence acquire sign language as their first language in early childhood and in a natural environment.

The third and last factor linked to the Deaf world is register. This factor summarizes two distinct categories: involvement in the Deaf community and access to the LIS version of TV news. Involvement in the community is measured as institutional roles on behalf of ENS. The levels of this variable are the following: president, vice-president, council member, no institutional role. If a signer has carried out an institutional role, this means

3 The use of the term 'factor' and the use of the term 'predictor' are intended with the same meaning.

4 Age group is inevitably collinear with the continuous variable Age. This means that the two factors provide the same information. Age group is also partially collinear with the nominal variable Job. This happens because the younger signers fall into the category 'student'. This aspect has been taken into consideration during the statistical analysis (cf. Section 4.2.).

that s/he has probably given formal presentations and taken part in formal debates or meetings. Such contexts are likely to provide exposure to high registers of LIS. Similarly, access to the LIS version of TV news (either with a Deaf journalist or with an interpreter) is connected to controlled registers of the language. In the metadata questionnaire participants were asked whether they had watched TV news in LIS the day before. The fact that involvement in the Deaf community and access to the LIS version of TV news are both connected to register motivates their fusion into one single factor.

The social predictors that have been considered for the present quantitative analysis are summarized in Table 1. Under the second column, the possible values for each predictor (i.e. levels) are reported.

Social predictors	**Levels**
Geographical provenance	Turin, Milan, Brescia, Bologna, Florence, Rome, Salerno, Trani, Lamezia, Ragusa
Residence status	urban, rural
Gender	female, male
Education	primary, middle, high
Job	white collar, blue collar, professional, student, unemployed
Age	Range: 18 – 81
Age group	Young, middle-aged, old
Family	hearing, deaf parents, deaf relatives
Register	low, medium, high

Table 1: Social predictors and their levels

3.2.2. Elicited judgments and narration tasks

In this section, I focus on the data that I collected personally (i.e. elicited judgments and

narration tasks). The consultants participating in these tasks are three Deaf LIS signers. They are all native signers since they have Deaf parents and have been exposed to LIS since birth. They have experience as LIS teachers and have frequent contacts with the Deaf community. Mirko Santoro is a 33-year-old man who was born in Salerno and grew up in Modena and Padua. He attended institutes for deaf students, earned a Master's degree in linguistics and he is now a PhD student in Paris. Gabriele Caia is a 37-year-old man who was born in Siracusa and grew up in Verona and Padua. During his school years, he attended institutes for deaf students. He has a Master's degree in Art, Music and Performing Arts. He is now a teacher of LIS at the University of Venice. Rosella Ottolini is a 46-year-old woman, who was born, grew up in Brescia, and then later moved to Milan. As for the primary school, she attended an institute for deaf pupils. Afterward, she attended mainstream middle school and high school. She is now a trainer for teachers of LIS.

The recording sessions took place at the Ca' Foscari University of Venice and at the Sign Language Lab at Institut Jean-Nicod (CNRS) in Paris[5]. During the elicitation sessions, I used a digital camera supported by a tripod. Special attention has been paid to the setting of lighting so that facial expressions were clearly visible. Another important requirement was the frontal position of informants with respect to the camera, as shown in Figure 1. In order to make informants feel at ease, elicitation sessions took place preferably in informal settings, although not as informal as the Deaf associations. However, as previously mentioned, all of them are used to collaborate with linguists, therefore they are familiar with both the technical settings and the different linguistic tasks.

5 Elicited data were partially collected in Paris because one of the informants was attending his PhD program there. My sincere thanks are due to the Sign Language Lab group at Institut Jean-Nicod (CNRS) for allowing me to use their space and their facilities.

Figure 1: Informant in frontal position

Once the elicitation session was concluded, I extracted the memory card from the camera and inserted it into the SD card slot of my MacBook Pro.

Using the software iMovie (version 8.0.6), it was possible to import the videoclips directly from the card. The main features of the iMovie interface are represented in Figure 2.

Figure 2: iMovie dialog box and its components

For the sake of clarity, I present here the import/export procedure that I followed after each elicitation session. After successful import from the SD card, the software displays raw clips in the event browser which is at the bottom of the screenshot. Then, the relevant clips need to be selected and dragged into the project library. Once the clips are correctly ordered and the project is completed, it can be extracted by creating a new file. The resulting video is a file in m4v format.

During the elicitation sessions, I used several techniques in an integrated approach. Each session was prepared in advance with a list of constructions I wanted to elicit. However, the protocol was not rigidly followed: deviations and on-the-spot adjustments depending on the informants' answers happened quite often. In the remaining of this section, the elicitation techniques are briefly explained, and the peculiarities and advantages of each of them are described.

For the narration tasks, I used visual material. Basically, I asked informants to describe either a single picture or a picture story. The aspect I paid particular attention to is absence of written language. In this way, materials did not give rise to inter-linguistic influence (i.e. influence from Italian to LIS). To illustrate, one of the pictures I used is shown in Figure 3.[6]

6 http://johnabisaab.com/2013/12/16/this-is-lebanon-get-used-to-it/ (20 August, 2014).

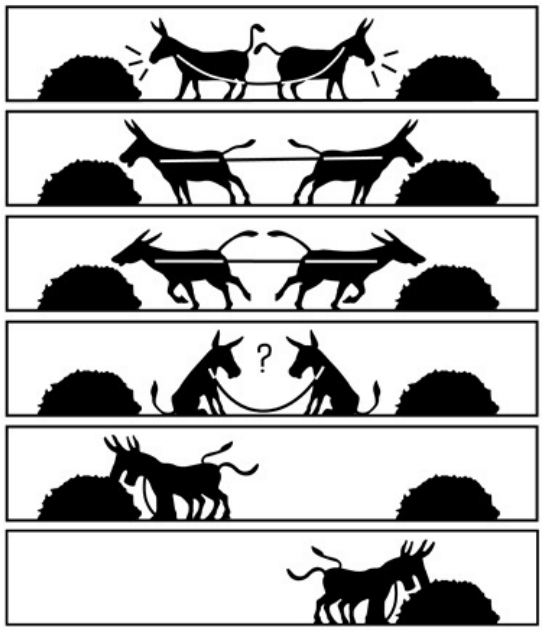


Figure 3: Picture story for the narration task

I also used short guided compositions in which the informants were asked to say something about a given topic. For example, in order to elicit a Measure Phrase (i.e. a complex expression in which a cardinal is combined with a measure noun referring to length, weight, capacity, etc.), I asked the informants to pretend to be educational scientists describing the peculiarity of giraffes and their long necks. The short composition I obtained is reported in (4).

(4) GIRAFFE IX NECK LONG CAN TWO METRE

'A giraffe’s neck can measure 2 meters in length'

In this case, I also used the strategy of role play by asking informants to assume the role of somebody else. An advantage of this technique is that informants are so concentrated on simulation and improvisation that they exert less control on what they are going to produce.

In order to test particular sign orders, I provided informants some lexical items and asked them to combine these items into grammatical sentences. For example, in order to

elicit a DP containing two different adjectives, I asked informants to think about possible combinations using the four signs listed in (5). As a result, I obtained the three sentences in (6).

(5) Lexical items:

{TABLE, SQUARE, BEAUTIFUL, BREAK}

(6) Acceptable combinations:

a. $\overline{\text{TABLE}_{\overline{3}}}$ $\overline{\text{IX BEAUTIFUL SQUARE}}$ IX-3 BREAK

b. $\overline{\text{TABLE}_{\overline{3}}}$ $\overline{\text{BEAUTIFUL IX SQUARE}}$ IX-3 BREAK

c. $\overline{\text{TABLE}_{\overline{3}}}$ $\overline{\text{SQUARE IX BEAUTIFUL}}$ IX-3 BREAK

'The beautiful square table broke'

In the case of multiple acceptable combinations, I always asked whether the same basic meaning was preserved or not. For example, according to an informant who produced the three options in (6), the three sentences are semantically equivalent.

As mentioned at the beginning of the section on data collection, I used elicited data also to test long sequences of signs. This was done by first eliciting a simple construction with the method explained above. The outcome is exemplified in (7)a. Then I asked my informants to add to this simple construction an item referring to the sign TABLE (e.g. GREEN), as in (7)b. This procedure has been reiterated several times and at each step a new nominal modifier has been added.

(7) a. CHILD TABLE BUY

'A child buys a table'

b. CHILD TABLE GREEN BUY

'A child buys a green table'

c. CHILD TABLE GREEN SMALL IX BUY

'A child buys a small green table'

This elicitation technique was used to observe how complex structures are progressively built without influencing the signer on the relative order among the elements.

In several cases, I tapped informants' intuitions about the acceptability of some sentences. I presented them either sentences signed by myself or sentences produced by the informants themselves during previous elicitation tasks. For example, one of the informants produced the sentence in (8) while he was spontaneously narrating a short story. In order to investigate the distribution of the numeral with respect to the noun, I reversed the order between these two elements obtaining the sentence in (9). Then I asked the same informant whether or not he would accept the sentence in (9) in the same context.

(8) CHILDREN TWO FIGHT FIST

'Two children got into a fist fight'

(9) TWO CHILDREN FIGHT FIST

In case of a negative reply (i.e. unacceptable judgment), I asked the informant to produce the option s/he would have used instead.

3.3. Data annotation

The next step was data annotation which is an essential operation that allows researchers to add the relevant linguistic information and to keep track of them for further refined analyses. This section offers a general overview of the annotation software ELAN and its main functions.[7] With respect to the procedure I applied during data annotation, I discuss the following steps in detail: creation of the protocol of

7 For more details, the reader is referred to the official full manual which is available at the following link: http://www.mpi.nl/corpus/manuals/manual-elan.pdf (20 August, 2014).

annotation, generation of a template, and data annotation.

3.3.1. ELAN

Annotation for both corpus data and elicited data has been conducted by using ELAN[8] (Eudico Linguistic Annotator). This is a software application developed by the Max Planck Institute for Psycholinguistics Nijmegen, the Netherlands (Crasborn & Sloetjes, 2008). Since its first release in 2002, ELAN has been receiving growing consensus among the scientific community, especially in the field of sign language linguistics.

In order to start a new annotation file, ELAN requires at least two source files, namely one (or more than one) multimedia file(s) and one template in etf (ELAN Template File) format. An unlimited number of annotations can be time-aligned to the original video and/or audio resource. A single annotation can consist of a sentence, a complex constituent, a lexical item, or a particular linguistic feature. At the end of the annotation session, a file in eaf (ELAN Annotation Format) format can be saved and stored for further analysis and comparison.

Following the most common procedures in sign language corpus linguistics, each sign needs to be associated to an ID-gloss (Johnston, 2001). This consists in an English word and is used as an identifying label to systematically refer back to the annotated sign (see Figure 8). Particular attention should be paid to internal consistency, so that each sign is associated only with one type of annotation and vice versa. To clarify this point, I briefly discuss the case of pointing signs[9] that display an ambiguous status since they can be used as determiner-like elements,[10] possessives, and locatives. This ambiguity could be problematic for the annotation, unless different annotations are used. The escape hatch consisted in using different glosses, as shown in (10).

(10) a. IX_DET determiner-like pointing sign

8 http://www.lat-mpi.eu/tools/elan/ (20 August, 2014).

9 A pointing sign is usually produced by extending the index finger and literally pointing at some particular location in the signing space.

10 In some cases, determiner-like pointing signs have been coded differently. For instance, the pointing signs articulated with repeated movement and clearly used as demonstratives have been glossed with the label 'IX_DEM'. The group of determiner-like signs share the same Part of speech level in the same tier (see fn 13), so regrouping 'IX_DET' and 'IX_DEM' can be done easily by using filters.

b. IX_POSS possessive pointing sign

c. IX_LOC locative pointing sign

It is worth noting that this gloss-based annotation is not intended as a translation and is not context-dependent. Neither it is a transcription because it does not code for the phonetic features associated to signs. In other words, ID-glosses do not describe how signs are articulated, they are rather used to univocally identify signs. The fact that nowadays annotations can be time-aligned with video recordings has reduced the practice of (phonetic) transcription. Despite some pioneering works (e.g. David J. Peterson's SLIPA),[11] at present a standardized International Phonetic Alphabet (IPA) for sign languages is not available. Some researchers are currently working with HamNoSys,[12] which is a transcription system internationally applicable. Thanks to some recent developments, the HamNoSys unicode-based font can be imported and used in ELAN tiers, but unfortunately there are still some issues concerning the possibility to look for HamNoSys transcriptions in the search dialog box.

The ID-gloss is usually associated with a number of details (i.e. dependent and independent variables) organized on multiple layers (tiers). Tiers are displayed in a hierarchical structure and can be either independent or linked to each other.

In ELAN, it is possible to choose between different interfaces. The default working mode is the annotation mode. It gives access to all possible functions. More specific interfaces are the synchronization mode, the transcription mode, and the segmentation mode.

The ELAN window displays different panels, as illustrated in Figure 4.

11 http://dedalvs.conlang.org/slipa.html (20 August, 2014).
12 http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/hamnosys-97.html (20 August, 2014).

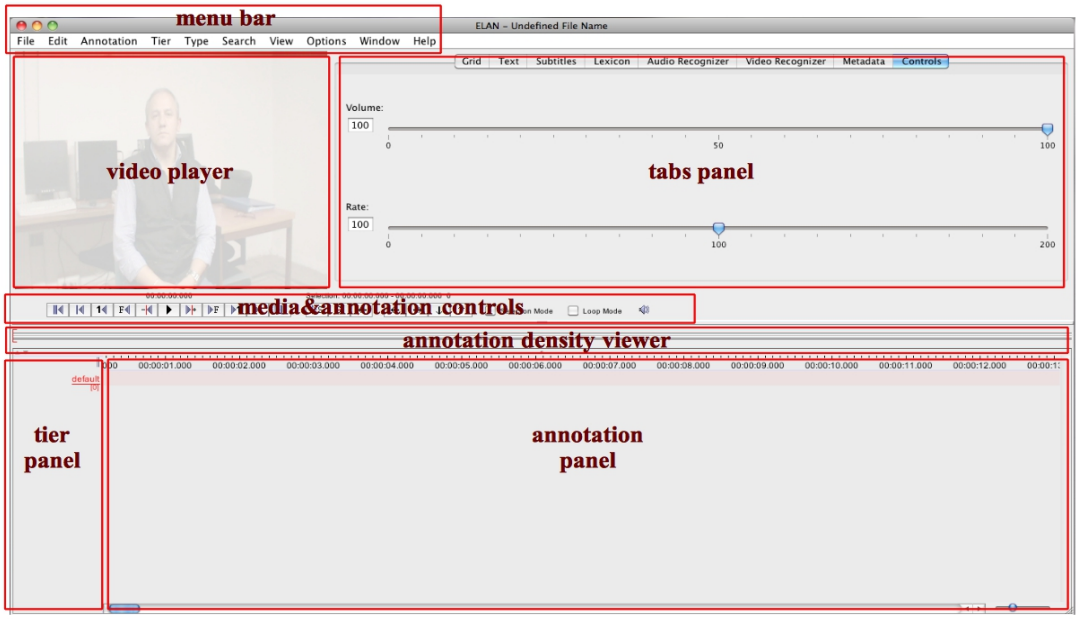


Figure 4: ELAN dialog box and its components

In the upper left corner, users can see the menu bar which gives access to all the main functions of the software, namely file, annotation, tier, type, search, view, options, window, and help.

Below the menu bar, the video viewer can show up to four videos at the same time. By using the video player controls below the video area, it is possible to start or pause the playback, go to the beginning or to the end of the video stream, go to the previous or next frame, go to the previous or next pixel, etc. Next to the video player options, users can find selection controls which can be used to examine a specific portion of the media file or navigate from one section to another in a quick way. In the annotation density viewer, users can observe how annotations are distributed throughout the whole media file. The timeline clarifies which portion of the video stream is shown in the video player.

The tabs panel in the upper right corner allows users to visualize the annotations in different formats such as grid, text, and subtitles. It is also possible to adjust the volume and the rate at which the video plays.

The tier panel in the bottom left corner presents the different tiers, which can be visualized in different ways: in hierarchical order, in alphabetical order, by linguistic

type, by participant or by annotator. Linguistic information is inserted in the annotation panel. If some variables hint at a limited number of possible values, ELAN users can enter these values in a controlled vocabulary so that drop-down menus can be used in the annotation panel. This is a time-saving tool since it spares manual insertion of the annotations. Another advantage of entering a list of entries in the controlled vocabulary is that annotators do not run the risk to mistype information. As a result, consistency is guaranteed throughout the annotation files.

To illustrate how ELAN has been used for annotation, a representative screenshot is shown in Figure 8. The active tier is highlighted in red. The blue vertical strip indicates that the sign glossed as 'SORDO' (DEAF) is selected for annotation.

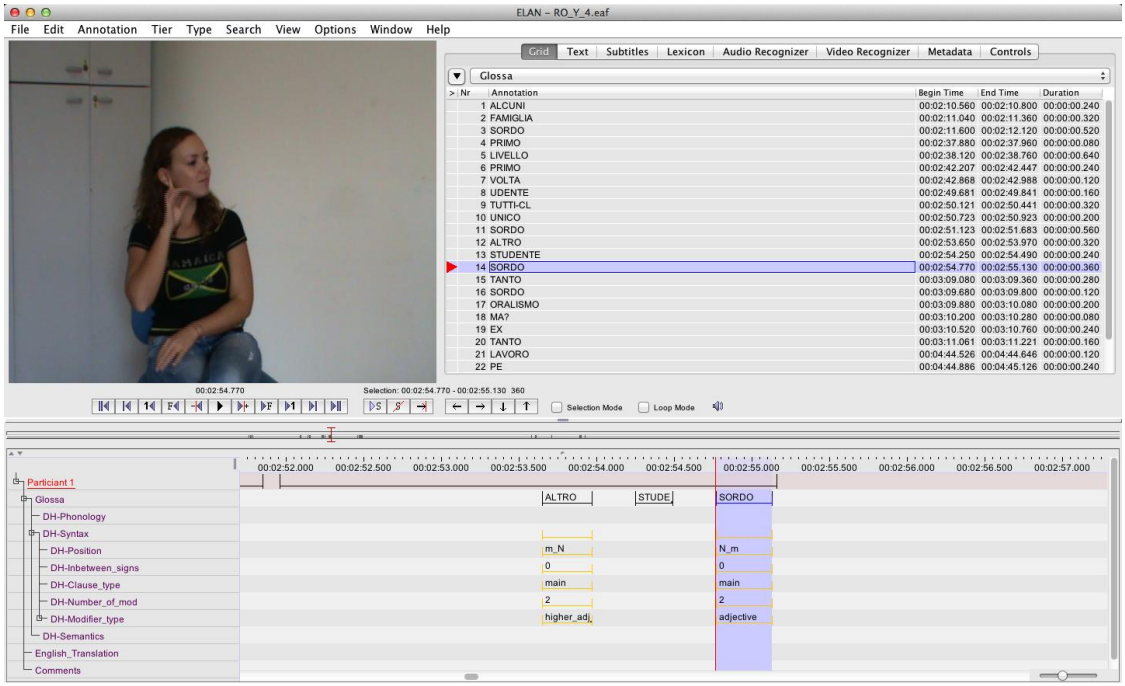


Figure 5: ELAN annotations

The possibility of permanently tracing eaf files back to their original sources has been preserved by using identification codes as filenames (Tagliamonte, 2006). The code assigned to each file indicates geographical provenance (e.g. 'MI' for Milan), age group (e.g. 'M' for middle-aged) and an informant number which is linked to the recording order. For example, the identifying string reported at the top of the screenshot in Figure 5 is “RO_Y_4”. It refers to the fourth participant belonging to the group of young

signers from Rome. One of the main advantages of using this system of pseudonyms is that the eaf files can be identified in a quick and straightforward way.

3.3.2. Coding scheme

In this section, I illustrate the protocol that has been adopted for the annotation of linguistic information to be used in the two quantitative studies on nominal modification (see Chapter 4).[13] This has been implemented in an ELAN template file. Each ELAN tier is associated with a specific linguistic information.[14]

The main requirement for inserting a new annotation was the presence of a noun and of at least one related modifier. In particular, nouns have been coded for ID-glosses only, modifiers have been coded for ID-glosses and all the linguistic variables listed in (12) below. The annotated modifiers directly modify the head noun and share with it the same specific prosodic contour and NMMs (for more details, see Bertone, 2007). No prosodic breaks have been found in between.

Before introducing the coding scheme, I point out what has not been considered for the purposes of data annotation. The few special cases that have not been annotated are listed in (11).

(11) Protocol of exclusion

a. nouns appearing without any modifier (e.g. HOUSE)

b. pronominal forms (e.g. IX-3)

c. modifier(s) co-occurring with a silent/unpronounced noun (e.g. IX_DEM)

d. incorporated modifiers (e.g. INTENSE-LIGHT)

e. nominal expressions that are only partially uttered (false starts, abrupt

13 In addition to the five tiers illustrated in this section, other three tiers have been annotated, namely Part of speech, Adjective degree, and Semantic class of the adjective. Part of speech has not been considered for the analyses because Modifier type offered a more refined modifier classification. Semantic class of the adjective is a variable that does not offer information with respect to all types of modifiers. Therefore, it could not be used as statistic variable. Also Adjective degree provided only partial information. Moreover, the cases of comparative and superlative adjectives were too few to be considered.

14 The results of the analysis on nominal modification in LIS will be presented in Section 4.2.

interruptions)

f. nominal expressions that have been repeated twice or more than twice (in these cases only one DP is considered)

Now, I illustrate the tiers included in the coding scheme. The first tier is named 'Participant' and codes for the utterance level. Utterance is defined as the concrete realization of the abstract linguistic competence. It is a prosodic unit and coincides with the speaker's communicative intention.

The second tier hosts glosses. Following standard conventions, glosses have been written in capital letters. ID-glosses for LIS signs were consistently provided in Italian so that independent Deaf evaluators could access and verify the soundness of the annotations. As specified above, both nouns and nominal modifiers have been segmented and glossed. This was done by paying special attention to the beginning and ending point of each sign.

In addition to the utterance and gloss tiers, the template contains five tiers coding for the syntactic (in the broad sense) properties of nominal modifiers. These five tiers are grouped together under the parent tier 'Syntax'. I list the syntactic tiers in (12) and for each of them I provide the corresponding levels.

(12) a. Position: the position of nominal modifiers with respect to the head noun. This information is collected to shed light on the distribution of modifiers. Levels: prenominal modifier, postnominal modifier, repetition of the modifier folding the noun, repetition of the noun folding the modifier.

b. In-between signs: the number of signs occurring between the nominal modifier and the noun. This variable concerns specifically linear order since it shows whether or not modifier and noun occur in adjacent positions. Levels: no intervening signs (i.e. modifier and noun are adjacent), one intervening sign, two or more than two intervening signs.

c. Number of modifiers: the number of modifiers included in the DP. This factor helps distinguish between simple and complex DPs. Levels: one modifier, two

modifiers, three or more than three modifiers.

d. Clause type: the type of clause that includes the DP under investigation. This information may give some insights on the linguistic macro-context containing the DP under investigation. Levels: main clause, argument clause, adjunct clause.

e. Modifier type: the category in which the modifier falls into. The classification of nominal modifiers mainly follows Cinque's (2012) fine-grained hierarchy. Cinque's classification has been implemented with some new categories (i.e. possessives, genitives, higher adjectives, and relational modifiers). This issue will be further discussed in Section 4.2.1. Levels: quantifier, pointing sign, ordinal numeral, cardinal numeral, possessive, higher adjective, adjective, genitive, relative clause, relational modifier.

For each linguistic factor, I entered a set of predefined entries in order to generate a controlled vocabulary. The values included in the controlled vocabularies will be shown in Table 2 below.

The last two tiers of the coding scheme are deputed to host the English translation and some comments. The latter is particularly useful in that it allows the annotation team to share doubts, opinions, and intuitions.

By using the function 'Add new linguistic type', I defined a type for each tier. 'Type' is defined as a property constraining a tier. For each type it is possible to specify four attributes: i) type name; ii) stereotype; iii) controlled vocabulary (if available); iv) ISO data category. Specifically, the stereotype determines the nature of the annotation that can be entered. The available options are listed in (13) and for each of them a corresponding example is provided.

(13) a. None: the annotation is linked to the time axis in an independent way.

Example: utterances.

b. Time subdivision: the annotation is a subdivision (i.e. a smaller unit) of a parent annotation. There cannot be any gap between two units.

Example: words.

c. Symbolic subdivision: the annotation is a subdivision of a parent annotation and there is no link to a time interval.

Example: morphemes within a word.

d. Included in: the annotation is a subdivision of a parent annotation. There can be a gap between two units.

Example: separate words.

e. Symbolic association: the annotation exactly corresponds to the parent annotation.

Example: free translation with respect to a sentence, gloss with respect to a word.

So, before starting with data annotation, I developed the protocol of annotation which is schematized here in Table 2.

Linguistic predictors	Type	Annotation
Participant	default-It	-
Gloss	Included in	ID-gloss
DH-Syntax	Symbolic association	-
DH-Position	Symbolic association	Prenominal modifier, postnominal modifier, reduplicated modifier, reduplicated noun
DH-Inbetween_signs	Symbolic association	Adjacent signs, one intervening sign, two intervening signs
DH-Clause_type	Symbolic association	Main clause, argument clause, adjunct clause
DH-Number_of_mod	Symbolic association	One, two, three or more modifiers
DH-Modifier_type	Symbolic association	Quantifier, pointing, ordinal numeral, cardinal numeral, possessive, higher adjective, adjective, genitive, relative clause, relational modifier
English_translation	Symbolic association	-
Comments	Symbolic association	-

Table 2: Protocol of annotation

The label 'DH' refers to the dominant hand. In principle, the distinction between dominant ('DH') and non-dominant ('NH') hand can be coded and accounted for. NH factors could be useful for further investigations, but for the purposes of this dissertation they have not been taken into consideration. So, for the sake of simplicity, all annotations have been entered into the DH tiers.

For the purposes of the case study on cardinal numerals in LIS, the coding scheme in Table 2 has been further implemented by adding some tiers intended for NMMs Phonetics. More specifically, these tiers are: NM-Head, NM-Eyebrows, NM-Body, and NM-Eyes. NMMs have not been systematically annotated throughout the corpus. Other NMMs (e.g. chin, shoulders, etc.) might implement the coding scheme and be relevant for further studies. Unlike elicited data, corpus data have not been annotated for NMMs because facial expressions could not be clearly evaluated. Such technical problem would have resulted in an unbalanced dataset and, as a consequence, it would not have been possible to perform an adequate statistical analysis.

On the basis of this protocol of annotation, I created a brand new template which has been tested with a few videos to detect potential inconsistencies. The template is hierarchically organized as shown in Figure 6.

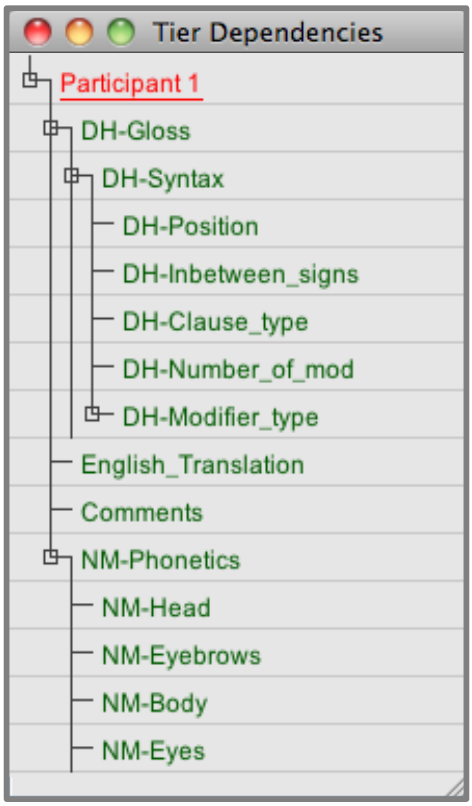


Figure 6: ELAN: template

A feature that has been investigated for a separate study included in the quantitative

analysis (Section 4.2.2.) is represented by the duration of the modifier signs. Duration is not associated with any specific tier, but its value in milliseconds can be extracted from the segmentation in the gloss tier. Sign duration is measured in terms of the time interval including holds and transition from the initial to the final location of a sign.

Adopting the coding scheme illustrated in Figure 6, I annotated both corpus and elicited data. As for corpus data, I annotated for each signer the first 15 complex DPs (when available). The first two minutes of production were neither considered nor analyzed because in this initial lapse, participants usually get gradually acquainted to the linguistic task and the surrounding environment. The overall number of annotated modifiers in the corpus is 2118. As for elicited data, I identified all the nouns co-occurring with at least one modifier, and then I annotated the linguistic features associated with each modifier.

The annotation has been performed by myself and checked by my supervisor, a LIS CODA[15] and expert of sign language annotation. The dubious cases have been double-checked by two Deaf native signers, teachers of LIS at the Ca' Foscari University of Venice who already collaborated in other projects based on the LIS Corpus.

3.4. Quantitative analysis

Quantitative research (Baayen, 2008; Gries, 2013) by definition deals with a large amount of data. In order to extract valuable information from data produced by 162 signers and make predictions about the linguistic behavior of the larger Deaf community, I needed to extract data from the annotation files and import them in a statistical software. Specifically, I developed a mixed-effects model that allowed me to evaluate to what extent language varies and what is the role played by each variable with respect to variation.

15 'CODA' (literally Child of Deaf Adult/s) is an acronym used to indicate a hearing person who was raised by one or two Deaf parents.

3.4.1. Data extraction

Once the annotation was completed, linguistic data have been extracted from ELAN and a dataset has been created. The first step of the extraction procedure was to preset a domain for each city. This could be done straightforwardly in ELAN through the search option.

The search function allows ELAN users to conduct both a single layer search or a multiple layer search. The former can be used if we have a single item in mind and we want to look for it. The latter can be used if we want to look for two or more annotations on the basis of temporal or structural constraints. Multiple layer search is laid out on a grid composed by some white cells, as shown in Figure 7.

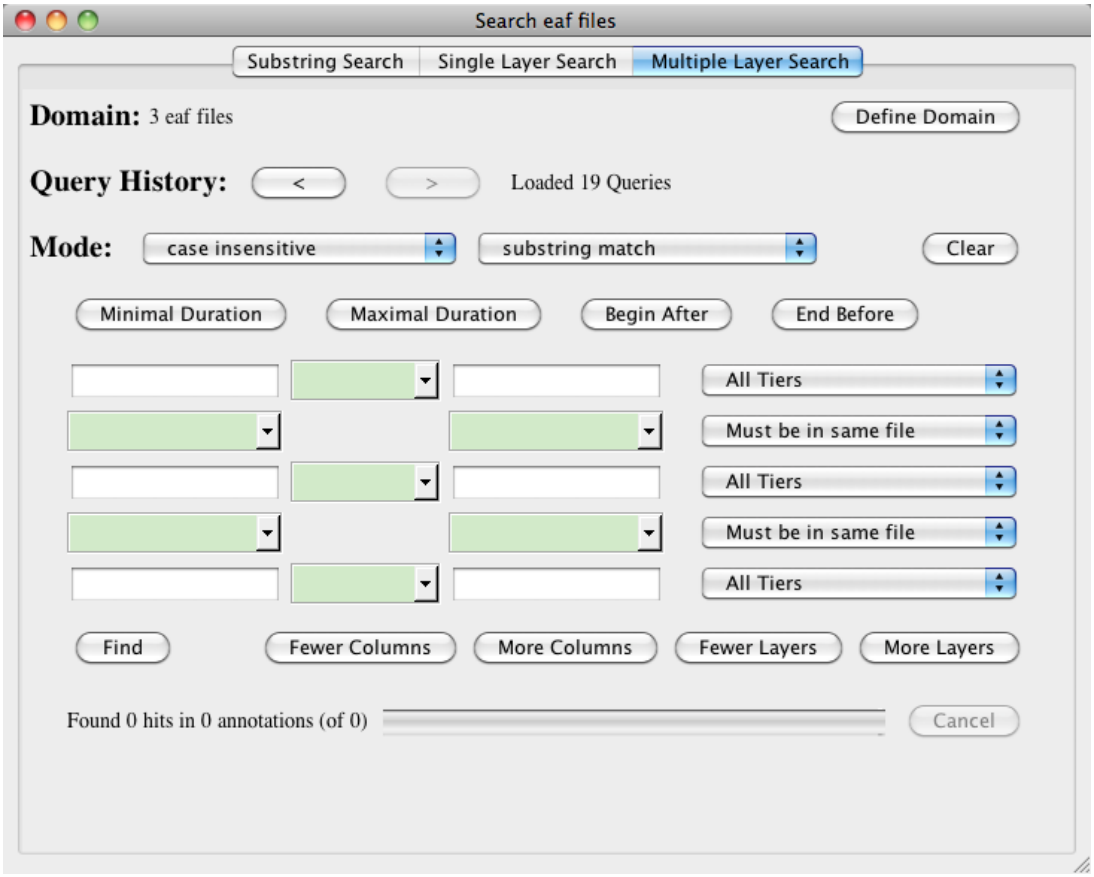


Figure 7: ELAN multiple layer search

The search query can be customized by adding or removing layers and/or columns. After the annotations are inserted in the cells, the connections between them need to be

specified in the green boxes (e.g. fully aligned, overlap, surrounding, etc.). The names of the tiers we want to take into consideration should also be indicated.

The main advantage of the multiple layer search is that values belonging to different tiers can be cross-checked at the same time. For instance, considering the corpus data used here, one might be interested in studying the position of pointing signs when they occur within particularly complex nominal expressions. In the search dialog box, it is necessary to specify the ID-gloss 'IX' in the tier Gloss, the value '3+' (i.e. three or more than three modifiers) in the tier DH-Number_of_mod, and the symbols '.+' (i.e. any value) in the tier DH-Position. In this case, looking for 'IX' in the gloss tier, it is possible to isolate the modifiers associated with that gloss and hence to single out only pointing signs. The second filter retains the indexes that occur in DPs containing three or more than three modifiers. Last, the third filter specifies that the signs selected by the first and second filter should be retrieved regardless of their position with respect to the head noun. The results obeying these three constraints are illustrated at the bottom of Figure 8.

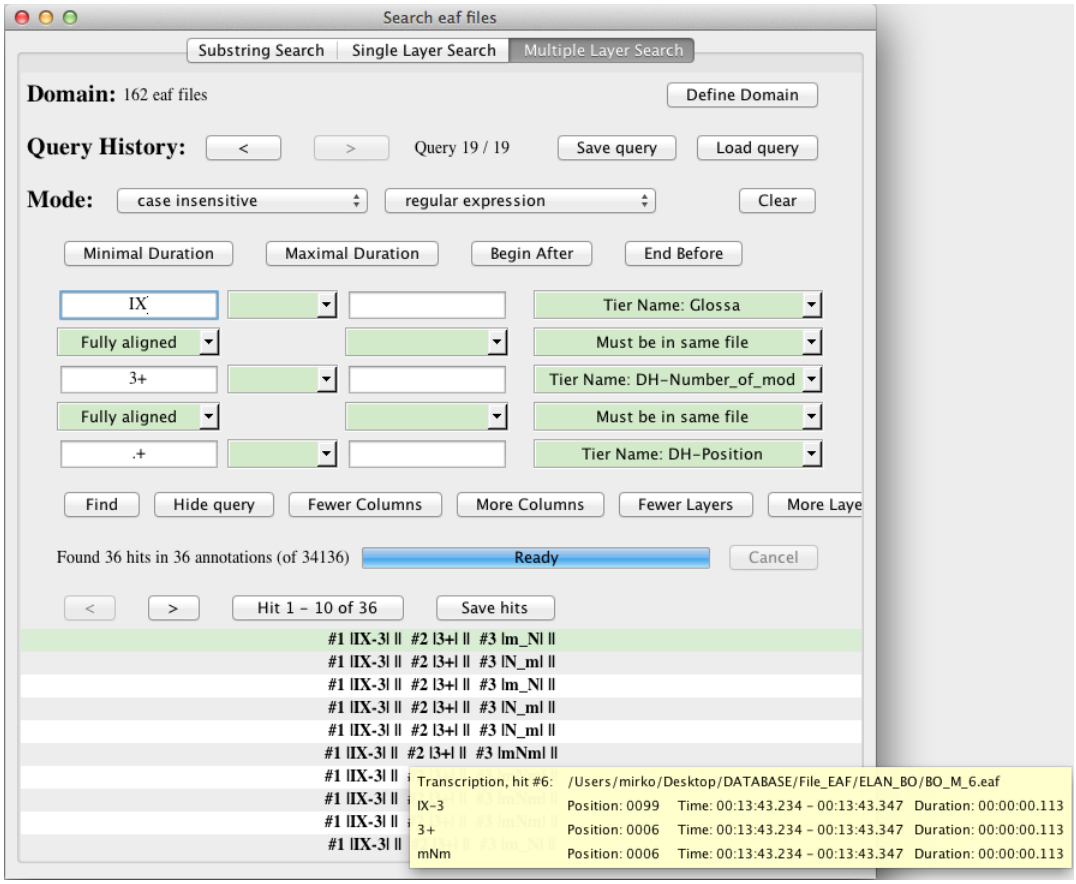


Figure 8: Results of the multiple layer search in ELAN

The search results appear quickly at the bottom of the dialog box: each annotation is associated to an info balloon showing details on source file, tier, begin time, end time, and duration. In dubious cases, annotations can be visualized and double-checked in their original eaf files by simply clicking on them. Most importantly, the search results can be exported to a tab-separated text file and they are accessible to other softwares.

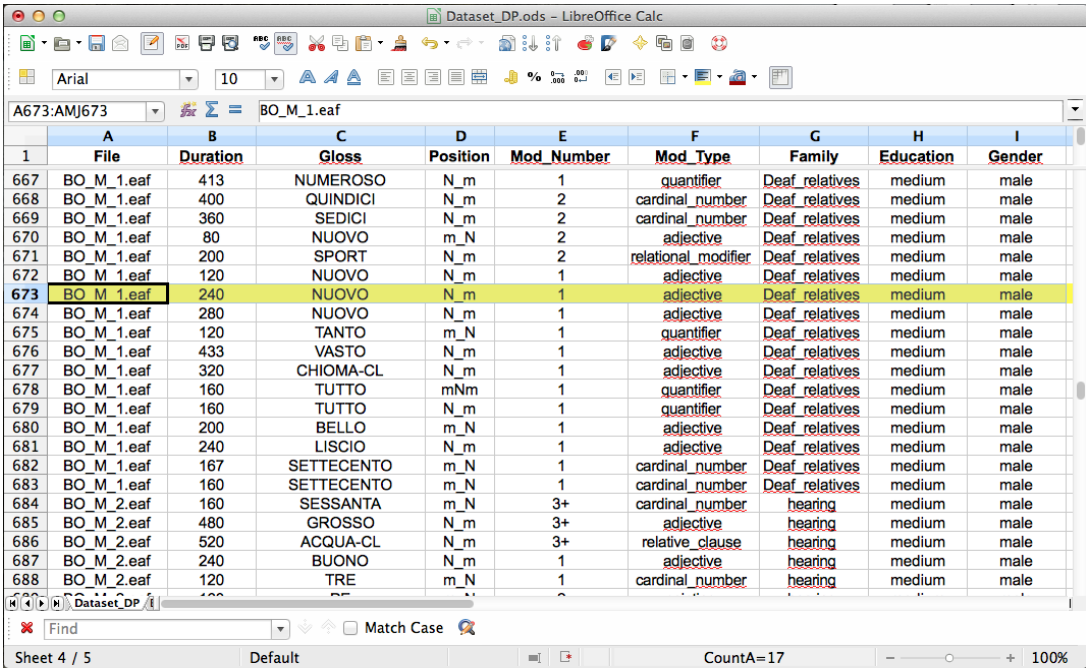
Turning back to the procedure for data extraction, I opened the dialog box 'Structured search multiple eaf' and then I defined ten different domains, one for each of the cities included in the corpus. This function ('define domain') is shown at the top-right corner of Figure 7. So, the eaf files matching for city have been grouped together under the same label (e.g. all annotation files of signers from Rome under the label 'Rome_narr', all annotation files of signers from Milan under the label 'Milan_narr', etc.).

The next step was to select 'Export multiple files as Tab-delimited Text' from the menu bar. This operation allows the generation of columns in the dataset. The source that has

been selected for data exportation consists of the ten domains corresponding to the ten cities. Then, the relevant tiers have been selected and details on begin time, end time, and duration have been specified. After this initial configuration, a file in txt format has been saved for each domain (e.g. 'Rome_narr.txt', 'Milan_narr.txt', etc.).

As a result, I obtained ten files in txt format, one for each city. In order to create the dataset, I had to transfer data from txt files to ten distinct spreadsheet files. So, I copied the whole content of the txt files and then I pasted it into the spreadsheet files. The correct tabulation was maintained by separating values with 'Tab' and 'Space' options. Data related to the ten cities were first imported separately on different worksheets, then they were merged together in a unique spreadsheet file.

Notice that this large spreadsheet file containing a total number of 2118 tokens includes linguistic information only. Each row (i.e. each annotation) has been associated with the name of the corresponding signer. Then after the rightmost cell of each row, sociolinguistic data (gender, city, age, etc.) have been added accordingly. As a result, the file included both linguistic and sociological information. To illustrate how the general spreadsheet for my dataset was generated, a representative screenshot is shown in Figure 9.



	A	B	C	D	E	F	G	H	I
1	File	Duration	Gloss	Position	Mod_Number	Mod_Type	Family	Education	Gender
667	BO_M_1.eaf	413	NUMEROSO	N_m	1	quantifier	Deaf_relatives	medium	male
668	BO_M_1.eaf	400	QUINDICI	N_m	2	cardinal_number	Deaf_relatives	medium	male
669	BO_M_1.eaf	360	SEDICI	N_m	2	cardinal_number	Deaf_relatives	medium	male
670	BO_M_1.eaf	80	NUOVO	m_N	2	adjective	Deaf_relatives	medium	male
671	BO_M_1.eaf	200	SPORT	N_m	2	relational_modifier	Deaf_relatives	medium	male
672	BO_M_1.eaf	120	NUOVO	N_m	1	adjective	Deaf_relatives	medium	male
673	BO_M_1.eaf	240	NUOVO	N_m	1	adjective	Deaf_relatives	medium	male
674	BO_M_1.eaf	280	NUOVO	N_m	1	adjective	Deaf_relatives	medium	male
675	BO_M_1.eaf	120	TANTO	m_N	1	quantifier	Deaf_relatives	medium	male
676	BO_M_1.eaf	433	VASTO	N_m	1	adjective	Deaf_relatives	medium	male
677	BO_M_1.eaf	320	CHIOMA-CL	N_m	1	adjective	Deaf_relatives	medium	male
678	BO_M_1.eaf	160	TUTTO	mNm	1	quantifier	Deaf_relatives	medium	male
679	BO_M_1.eaf	160	TUTTO	N_m	1	quantifier	Deaf_relatives	medium	male
680	BO_M_1.eaf	200	BELLO	m_N	1	adjective	Deaf_relatives	medium	male
681	BO_M_1.eaf	240	LISCIO	N_m	1	adjective	Deaf_relatives	medium	male
682	BO_M_1.eaf	167	SETTECENTO	m_N	1	cardinal_number	Deaf_relatives	medium	male
683	BO_M_1.eaf	160	SETTECENTO	m_N	1	cardinal_number	Deaf_relatives	medium	male
684	BO_M_2.eaf	160	SESSANTA	m_N	3+	cardinal_number	hearing	medium	male
685	BO_M_2.eaf	480	GROSSO	N_m	3+	adjective	hearing	medium	male
686	BO_M_2.eaf	520	ACQUA-CL	N_m	3+	relative_clause	hearing	medium	male
687	BO_M_2.eaf	240	BUONO	N_m	1	adjective	hearing	medium	male
688	BO_M_2.eaf	120	TRE	m_N	1	cardinal_number	hearing	medium	male

Figure 9: Spreadsheet file containing the dataset

Rows from 667 to 683 in Figure 9 show the annotations relative to a middle-aged signer from Bologna (file name in the first column: 'BO_M_1'). To illustrate, I concentrate on the token highlighted in yellow and I clarify what each cell exactly refers to. The gloss of this annotated sign is 'NUOVO' (NEW) and it has a duration of 240 milliseconds. From Figure 9, we can see that the sign is in postnominal position ('N_m'), occurs in a simple DP since it is the only present modifier, and it is classified as adjective. Figure 9 also shows part of the coding for social variables: specifically, this sign is produced by a man with a medium level of education and with at least one Deaf relative in the family.

The creation of the final dataset required one further adjustment. It concerns the variable Position. As already mentioned, the levels contained in this variable are the following: 'm_N' (prenominal modifier), 'mNm' (repeated modifier), 'N_m' (postnominal modifier), 'NmN' (repeated noun). The case involving modifier repetition is problematic because each pair of modifiers consists of two distinct annotations, instead of one. This redundancy depends on the coding type that I have adopted and is problematic because it leads to the overestimation of the number of repeated modifiers. I solved this complication by modifying the spreadsheet file. I added an extra column, Position2, in which I split the mNm cases into 1Nm and mN2. I did this in order to eliminate redundant data points (mN2) provided by the doubling configuration. This operation was not necessary for NmN since there is only one modifier in this type of construction. The dataset resulting from the exclusion of the redundant data points contains a total number of 2023 tokens.

Finally, the spreadsheet file has been saved in csv format by using a comma (',') as value for separation. The resulting csv file is to be considered as the final dataset as it contains the names of the signers, the linguistic values, the social values, and details on duration. This is what the statistical software R needs in order to initiate a quantitative analysis.

3.4.2. Statistical analysis

The software R was developed at Bell Laboratories by John Chambers and colleagues and it is currently downloadable from the Comprehensive R Archive Network (CRAN).[16] It is available as open-source software and functions thanks to the large

16 http://cran.r-project.org/ (20 August, 2014).

community of people working with it. Its functionalities are constantly being updated and improved with new packages. R is not just a statistical software: it can also handle spreadsheet data, databases, graphics, and other functions. In order to work successfully, R requires the user to master its programming language which is made up of a number of commands and compositional rules (for detailed instructions on how to use R, the reader is referred to Baayen, 2008 and Gries, 2013).

For the purposes of this dissertation, R has been used to analyze the LIS corpus data. After the startup of R, what I did was attach a number of useful packages as 'languageR', 'MASS', 'lattice', etc. Another fundamental operation was the import of my data into R. Figure 10 shows how it is possible to load some packages and the dataset in csv format into the R Console.

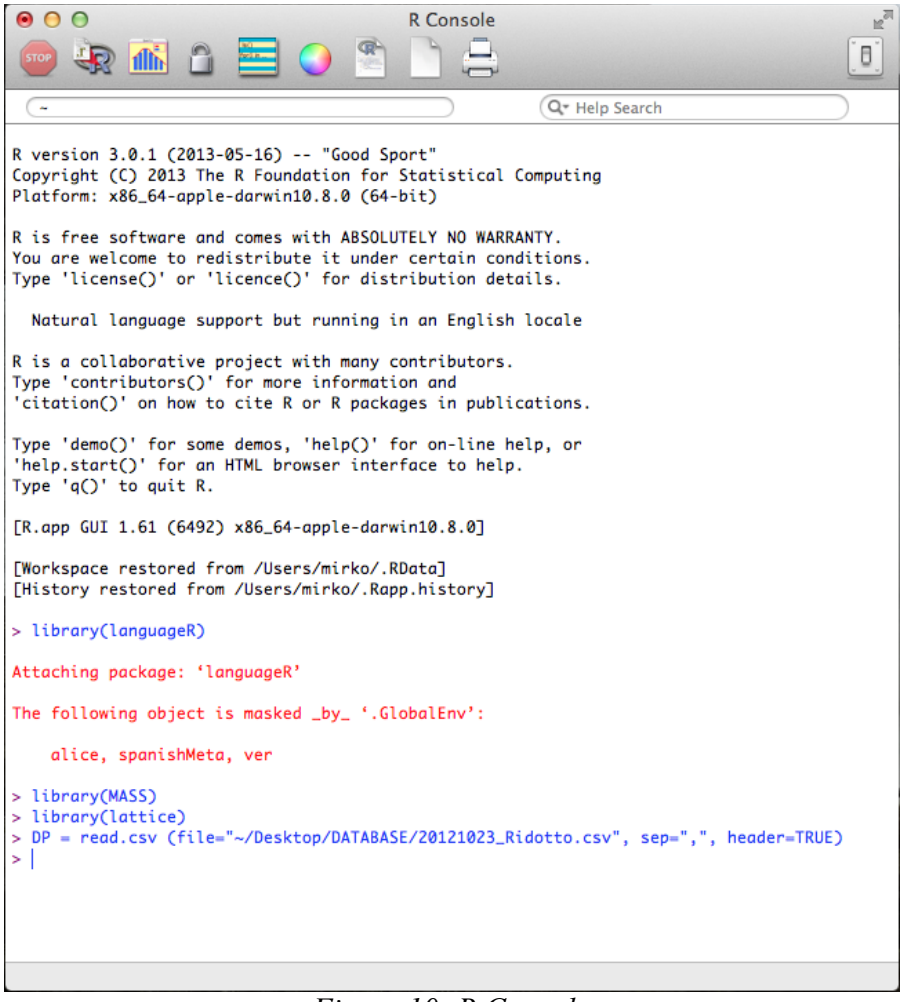


Figure 10: R Console

After the prompt sign ('>'), any complete R command can be typed. For the sake of convenience, the uploaded dataset has been saved under the name “DP”. Thanks to this label, it is thus straightforward to refer back to the dataset for any further operation.

3.4.3. Statistical variables

Once the dataset is available in the R Console, it is important to determine what kind of information the software should deal with. As previously discussed (see Section 1.4.2.), the quantitative linguist needs to classify variables as dependent or independent. In a nutshell, the dependent variable is what the researcher wants to shed light on it, whereas the independent variable is either a linguistic or a social factor that may determine the variation of the dependent variable.

In statistics, a 'variable' is any property that shows variance and that can be either measured or counted. The variables used in this work are of two types: numerical data and nominal data. The difference between these two categories is explained in (14).

(14) a. Numerical data: this variable can be measured by means of numerical measurements (i.e. scores). Values are usually arranged in a scale and follow a precise order. A numerical variable is, for example, the age of the signers participating in the data collection. In this case, possible values are 31, 67, 45.

b. Nominal data: this variable contains two or more levels described in words (i.e. categories). Each item is categorized in one and only one of the levels and, importantly, levels do not overlap. If there are only two different levels, then the nominal variable is defined as binary. Differently from numerical data, nominal data are not ranked in any scale. Pursuing the example of the age of the signers, we could think of some age groups. The levels of this nominal variable may be young, middle-aged, and old.

In order to capture the variation characterizing the dataset, it is necessary to shift from linguistic variation to statistical variables which, as shown in (14), can be either numerical or nominal. Variation must be codified into statistical variables. Variables

need to be operationalized, intending that it must be clear what is to be observed or measured for each of them (Gries, 2013). They have to categorize linguistic variation exhaustively and without overlapping. To illustrate, the study on the distribution of nominal modifiers in LIS aims at examining the linguistic variation found in the nominal domain. Variation is captured by the statistical variable Position which includes four possible options, namely prenominal modifiers, postnominal modifiers, modifier repetition, and noun repetition. The variable Position categorizes variation so that it includes exhaustive (no options are left out) and non-overlapping (levels are not redundant) information. This procedure has been systematically applied to all other variables.

In the light of what was explained in (14), the linguistic and social variables included in the dataset can be classified as in (15).

(15) a. Numerical variables: (sign) Duration, Age

b. Nominal variables: Position, Modifier type, In-between signs, Number of modifiers, Clause type, Age group, Gender, City, Residence, Family, Education, Job, Register

The R function 'xtabs ()' allows the generation of contingency tables, so that two or more than two factors can be cross-tabulated. The formula referring to this function is exemplified in (16). This contingency table cross-tabulates the position of nominal modifiers with the signers' educational background.

(16) xtabs (~DP$Position + DP$Education)

The output of (16) is given in (17). The four ordering options ('m_N', 'mNm', 'N_m', 'NmN') are vertically displayed under 'DP$Position'. The three education groups are visible under 'DP$Education' ('p' stands for primary school, 'm' for middle school and 'h' for high school and university). For each education group, the contingency table shows how the modifiers are distributed with respect to the head noun.

(17)

	DP$Education		
DP$Position	p	m	h
m_N	35%	39%	31%
mNm	6%	5%	4%
N_m	57%	55%	64%
NmN	2%	1%	1%

In some cases, this function may highlight the fact that two values show a similar distribution. If this is the case, the coding scheme can be honed by collapsing the two similar values into one (Tagliamonte, 2006). For instance, assuming that signers with a primary-school certification and signers with a middle-school certification behave similarly, it is possible to collapse these two categories into one. The new category can thus be compared to the group of the signers who have completed high school or have graduated from university. As a result, a new factor is created, Education2. It contains two levels: primary and middle school vs. high level of education.

Whether the statistical model should include Education or Education2 (i.e. the recoded variable with collapsed values) is something that should be decided by observing how the model reacts to the presence of one or the other. Whether to include Education or Education2 depends also on the explicatory strength of each level contained in the two factors. This procedure has been systematically used during the quantitative analysis (see Section 4.2.).

I anticipate here that the repetition cases (1Nm and NmN) are a small minority in the corpus (cf. Section 4.2.1.). They are too few to conduct a systematic quantitative analysis. Therefore, a new dataset has been created without considering these special cases. This dataset, containing 1908 data points, has been considered for the binomial analysis (i.e. analysis based on a two-value dependent variable).

3.4.4. Mixed model

A mixed-effects model (or multi-level model) is a statistical model which is able to handle at least one fixed effect and at least one random effect. Fixed effects are constant across individuals and they are characterized by repeatability. An example of fixed effect is gender because its levels (i.e. male, female) do not vary according to different groups of people. On the contrary, random effects show variation due to chance. Their levels represent just a set of all possible levels, hence do not cover all possibilities. Usually subjects and items are non-repeatable, hence vary randomly (Baayen, 2008: 241-242).

The main advantage of using a mixed model is that it can deal with different types of factors and it can measure by-signer and by-gloss correlations. It actually accounts not only for variation between different groups (e.g. people living in urbanized areas vs. people living in rural areas) but also for variation within different groups (e.g. individual signers).

In this section, I illustrate the procedure to compute the mixed-effects model for the study of the position of nominal modifiers. Since the same procedure applies to the other study, the one on the duration of nominal modifiers, I do not discuss it here.

The starting point for the mixed model that I used for the quantitative study of modifier position is summarized in (18).

(18) a. Dependent variable: Position2

b. Random factors: Gloss (corresponding to modifier), Name (corresponding to subject)

c. Fixed factors: 4 linguistic predictors (Clause type, In-between signs, Number of modifiers, Modifier type), 9 sociolinguistic predictors (Education, Job, Register, Age, Group, City, Gender, Residence, Family)

One single item or one single subject may behave anomalously. Random factors, listed in (18)b, are used to correct these possible distortions. Fixed factors, listed in (18)c, are used as predictors of the variability of the position of nominal modifiers with respect to

the head noun.

In order to run a mixed model, the LME4 package is required. Depending on the type of dependent variable, one can use the function 'lmer ()' or 'glmer ()'. Lmer () is required for continuous dependent variables, and hence it has been used for the quantitative study on the duration of modifiers. Glmer () is needed for nominal dependent variables, and hence it has been used for the quantitative study on the position of modifiers. The prototypical formula that has been used to run the model with glmer () is given in (19).

(19) model_name = glmer (dependent variable ~ independent_variable$_1$ + independent_variable$_2$ … + (1|Random effect$_1$) + (1|Random effect$_2$), family = binomial, data = xxx)

The glmer () formula in (19) contains six elements: i) the dependent variable; ii) a '~' operator separating the dependent variable from the rest of the formula; iii) the fixed effect(s) separated by a '+'; iv) the random effect(s); v) the type of model ('family'); vi) the name of the dataset. Random effects are introduced by parenthesis. The number '1' appearing inside the brackets indicates the intercept. This consists in a sort of baseline mean. The vertical line following '1' stands for 'given' or 'conditional on' and it is used to separate "the grouping factor (to its right) from the fixed-effects terms for which random effects have to be included" (Baayen, 2008: 244). The whole model is associated to a model name which functions as a label to refer back to the model.

In order to identify the most explanatory model, two main procedures are involved: the step-up (or forward entry) and the step-down (or backward elimination). Both of them consist of a series of multiple comparisons and are organized in a number of subsequent steps proceeding in opposite directions.

The step-up procedure starts with a model containing only the dependent variable and the random effects and then it proceeds with several rounds. At round 1, the contribution of each single fixed effect to the fit of the model is evaluated against the model without fixed effects. At the end of round 1, the model containing the most significant fixed effect is retained and the procedure goes to round 2. At round 2, the model containing

the significant fixed effect is augmented with another factor and its significance is evaluated. At the end of each round the best model contains one fixed effect more than the model of the previous round. The procedure continues until no further additions result in a statistically significant improvement of the current model (Tagliamonte 2006: 140).[17]

To illustrate this procedure, I show how I developed the mixed model relative to the study on the position of nominal modifiers by considering a subset of fixed effects, namely Family, Gender, Clause type, and In-between signs. The starting model includes only the random factors (i.e. Gloss and Name), as reported in (20).

(20) su.00.01 = glmer (Position2 ~ (1|Gloss) + (1|Name), family=binomial, data=DP)

As seen so far, the basic model contains only dependent variable, intercept, and random effects. This is the starting point for the step-up procedure. Round 1 starts by generating four models. Each of them includes the basic model in (20) plus one fixed effect. The models generated in round 1 are reported in (21).

(21) a. su.01.01 = glmer (Position2 ~ Family + (1|Gloss) + (1|Name), family=binomial, data=DP)

b. su.01.02 = glmer (Position2 ~ Gender + (1|Gloss) + (1|Name), family=binomial, data=DP)

c. su.01.03 = glmer (Position2 ~ Clause_type + (1|Gloss) + (1|Name), family=binomial, data=DP)

d. su.01.04 = glmer (Position2 ~ Inbetween_signs + (1|Gloss) + (1|Name), family=binomial, data=DP)

These models differ from each other only for one variable: Family in (21)a, Gender in

17 When needed, factors with several levels have been collapsed and their effect on the general model re-tested.

(21)b, Clause type in (21)c, In-between signs in (21)d. Once these operations are done, each model is compared with the basic one (su.00.01). This is done with the 'anova' function, as in (22).[18] The output of the anova is reported in Table 3.

(22) anova (su.00.01, su.01.01)

Models	**Df**	**AIC**	**BIC**	**logLik**	**deviance**	**Chisq**	**Chi Df**	**Pr(>Chisq)**
su.00.01	3	2063.4	2080.1	-1028.7	2057.4			
su.01.01	5	2030.9	2058.6	-1010.5	2020.9	36.489	2	1.193e-08 ***[19]

Table 3: Anova at round 1 of the step-up

The output in Table 3 shows the difference between the two models in terms of degrees of freedom ('Df'), Akaike and Bayesian information criteria ('AIC' and 'BIC'), log likelihood[20] and the probability value (Pr>Chisq). According to the Pr>Chisq value in Table 3, the second model (su.01.01) is statistically different with respect to the first model (su.00.01). The best model is the one that: i) is significantly different from the model of the previous round (in this case, the basic model) and ii) has the log likelihood value closer to 0.

The same comparing procedure is adopted for the other three models of round 1 (su.01.02, su.01.03, su.01.04). The models resulting significantly different from the basic model are the one including Family (su.01.01) and the one including In-between signs (su.01.04).

Which of the two should be added to the basic model depends on the log likelihood value. The schema containing the models' log likelihood (Table 4) helps detect the winning model.

18 The 'anova' function applies the chi-square test to establish whether the models are significantly different (Baayen, 2008: 166-167).

19 Significant codes in R: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

20 The log likelihood "measures the goodness of fit of the analysis. Figures closer to 0 represent better models than those further removed from 0" (Tagliamonte, 2006: 156).

Models	logLik
su.00.01	-1028.7
su.01.01	-1010.5
su.01.02	-1028.7
su.01.03	-1028.4
su.01.04	-1018.6

Table 4: Models' log likelihood at round 1 of the step-up

Looking at Table 4, the resulting best model is the one containing the predictor Family (su.01.01) because it is significantly different than model su.00.00 and its log likelihood value is the closest to 0. This model is considered for the second round of the step-up analysis.

Following this procedure, the model is enriched by adding one significant fixed factor at a time. At the end of round 2, the model that is significantly different from the winning model of round 1 and has the log likelihood closer to 0 is the one containing Family and In-between signs. The resulting model (su.02.03) is given in (23).

(23) su.02.03 = glmer (Position2 ~ Family + Inbetween_signs + (1|Gloss) + (1|Name), family=binomial, data=DP)

For each new predictor added to the model, I ensure that it is not collinear with respect to the other predictor(s) already contained the model. This is done in order to assess whether they are correlated and provide overlapping information. In the case of (23), I check collinearity between the two selected factors (Family and In-between signs) by examining the contingency table based on them (Tagliamonte, 2006: 181-187). The table is illustrated in (24).

(24)

DP$Family	DP$Inbetween_signs 0	1	2+
h	46%	45%	50%
p	33%	33%	36%
r	21%	22%	14%

The contingency table in (24) shows that Family and In-between signs are not collinear, since there are no values below 5% (Tagliamonte, 2006). Also, the two variables are not distributed in terms of "irregular" clusters of data. Data appear rather uniformly distributed.

The third and last round of the step-up procedure tries to augment the model in (23) containing two fixed effects (Family and In-between signs) by adding Gender, as in (25)a, and Clause type, as in (25)b.

(25) a. su.03.01 = glmer (Position2 ~ Family + Inbetween_signs + Gender + (1| Gloss) + (1|Name), family=binomial, data=DP)

b. su.03.02 = glmer (Position2 ~ Family + Inbetween_signs + Clause_type + (1| Gloss) + (1|Name), family=binomial, data=DP)

The two models in (25) only differ by the third fixed effect. The anova function shows that including either Gender or Clause type does not significantly improve the fit of the model. The log likelihood values reported in Table 5 confirm that the winning model of round 2 and the two models of round 3 are very similar with respect to each other.

Models	logLik
su.02.03	-999.73
su.03.01	-999.71
su.03.02	-999.13

Table 5: Models' log likelihood at round 3 of the step-up

Both models of round 3 are discharged and the best model of round 2 is kept (su.02.03). At this point, I end the step-up procedure because no other significant predictors can be entered into the model.

Differently from the step-up, the step-down procedure starts with a complex model and from this model non-significant factors are removed one by one. The starting model contains the dependent variable, all the independent variables, and the random effects. At round 1, each model is generated by excluding one of the independent variables. The worst explicatory model (i.e. the least significant one) is singled out and considered as the winner of round 1 and the basic model for round 2. At round 2, another independent variable is excluded on the basis of the worst significance value. This step-wise procedure continues until there are no other non-significant independent variables that can be removed from the model. So, the step-down ends with the most explicatory simple model.[21]

This procedure begins with a complex model including all the independent variables. The formula for the starting model is reported in (26). As with the examples introduced to illustrate the step-up procedure, only four fixed effects are taken into consideration, namely Gender, Family, Clause type, and In-between signs.

(26) sd.00.01 = glmer (Position2 ~ Family + Inbetween_signs + Gender + Clause_type + (1|Gloss) + (1|Name), family=binomial, data=DP)

Round 1 starts by generating four models. Each of them excludes one of the fixed

21 Since the inspection on collinearity was conducted during the step-up, it was not necessary to repeat it during the step-down procedure.

effects included in (26). The models generated in round 1 are reported in (27).

(27) a. sd.01.01 = glmer (Position2 ~ Family + Inbetween_signs + Gender + (1|Gloss) + (1|Name), family=binomial, data=DP)

b. sd.01.02 = glmer (Position2 ~ Family + Inbetween_signs + Clause_type + (1|Gloss) + (1|Name), family=binomial, data=DP)

c. sd.01.03 = glmer (Position2 ~ Family + Gender + Clause_type + (1|Gloss) + (1|Name), family=binomial, data=DP)

d. sd.01.04 = glmer (Position2 ~ Inbetween_signs + Gender + Clause_type + (1|Gloss) + (1|Name), family=binomial, data=DP)

In order to assess whether one of the fixed effects in (26) can be removed from the model without compromising its statistical significance, the four models of round 1 are compared with respect to the basic model of the step-down. I do this by applying the anova function and I obtain the outcome illustrated in Table 6.

Models	**logLik**
sd.01.01	-999.7
sd.01.02	-999.13
sd.01.03	-1009.9
sd.01.04	-1018.1
sd.00.01	-999.11

Table 6: Models' log likelihood at round 1 of the step-down

On the basis of the output in Table 6, I pinpoint the factor whose "loss least significantly reduces the likelihood" (Tagliamonte, 2006: 143). Given the fact that the basic model (sd.00.01) has a log likelihood of -999.11, the model of round 1 with the closest value is sd.01.02 (the one without Gender) which has a log likelihood of -999.13. Afterwards, I

run the anova function once more to check whether the removal of the factor Gender, as in (27)b, is legitimized or not. The corresponding output is in Table 7.

Models	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
sd.01.02	9	2016.3	2066.1	-999.13	1998.3			
sd.00.01	10	2018.2	2073.6	-999.11	1998.2	0.0313	1	0.8597

Table 7: Anova at round 1 of the step-down

Table 7 confirms the possibility to remove Gender. Being statistically non-significant, the presence of this factor does not offer a significance improvement to the model.

The same procedure applies for round 2 which starts with sd.01.02 (i.e. the model without Gender) as basic model. Three new models are generated: sd.02.01 excluding Clause type, sd.02.02 excluding In-between signs, sd.02.03 excluding Family. Given the fact that the basic model (sd.01.02) has a log likelihood of -999.13, the model of round 2 with the closest value is sd.02.01 which has a log likelihood of -999.73. The anova function is applied to check the legitimacy of the removal of Clause type and the corresponding output is given in Table 8.

Models	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
sd.02.01	7	2013.5	2052.2	-999.73	1999.5			
sd.01.02	9	2016.3	2066.1	-999.13	1998.3	1.1983	2	0.5493

Table 8: Anova at round 2 of the step-down

According to the Pr(>Chisq) value in Table 8, the factor Clause type does not offer a significance improvement to the model. Therefore, at the end of round 2, this fixed effect can be removed.

Round 3 starts with sd.02.01 as basic model. Then, two new models are generated: sd.03.01 (without Family) and sd.03.02 (without In-between signs). Given the fact that

the basic model (sd.02.01) has a log likelihood of -999.73, the model of round 3 with the closest value is sd.03.02 which has a log likelihood of -1010.51. In order to assess whether the removal of In-between signs is legitimized or not, the anova function is run. The output is provided in Table 9.

Models	**Df**	**AIC**	**BIC**	**logLik**	**deviance**	**Chisq**	**Chi Df**	**Pr(>Chisq)**
sd.03.02	5	2031.0	2058.7	-1010.51	2021.0			
sd.02.01	7	2013.5	2052.2	-999.73	1999.5	21.569	2	2.071e-05 ***

Table 9: Anova at round 3 of the step-down

The output in Table 9 shows that In-between signs significantly contributes to the fit of the model, therefore it is a significant factor. At the end of round 3 no independent variable is allowed to be removed. The best model resulting from the step-down procedure is sd.02.01, reported here in (28).

(28) sd.02.01 = glmer (Position2 ~ Family + Inbetween_signs + (1|Gloss) + (1| Name), family=binomial, data=DP)

At the end of the step-down, I check whether the model resulting from the step-up coincides with the model resulting from the step-down. Indeed, the final model resulting from the step-up, su.02.03 in (23), is identical to the final model resulting from the step-up, sd.02.01 in (28). The two procedures converge on the same model and thus the resulting model can be used to explain the data (Tagliamonte, 2006).

At this point, I refine the model by evaluating the possible interactions, namely fixed effects produced by the combination of two (or more) predictors. In R it is possible to check for interactions by inserting an asterisk between the factors that presumably interact, as in (29).

(29) sd.02.01.i = glmer (Position2 ~ Family * Inbetween_signs + (1|Gloss) + (1|

Name), family=binomial, data=DP)

In order to evaluate whether the interaction is significant or not, the model containing it (sd.02.01.i) is compared to the model resulting from the step-up and step-down procedure. The output of the anova is given in Table 10.

Models	Df	AIC	BIC	logLik	deviance	Chisq	Chi Df	Pr(>Chisq)
sd.02.01	7	2013.5	2052.2	-999.7	1999.5			
sd.02.01.i	11	2014.0	2074.9	-996.0	1992.0	7.4074	4	0.1159

Table 10: Anova comparing the same model with and without interaction

The output in Table 10 shows that the interaction between Modifier type and Age is significant, therefore it should not be included in the model.

Another essential operation to refine the model is excluding the outliers. These are the data points that are surprisingly distant from all other observations (Baayen, 2008). Outliers may reveal some atypical responses or even experimental errors. Since outliers could obscure the trend shown by the other data points and hence compromise the whole analysis, it is convenient to eliminate them. This could be done by excluding those data points that are more distant than 2 or 2.5 or 3 points with respect to the standard deviation. For example, considering a standard deviation cutoff of 2.5, the procedure for removing outliers requires the string 'subset = abs(scale(resid(model_name)))<2.5' at the end of the glmer formula. The complete formula is shown in (30).

(30) subset_name = glmer (dependent variable ~ independent_variable$_1$ + independent_variable$_2$ … + (1|Random effect$_1$) + (1|Random effect$_2$), family = binomial, data = xxx, subset = abs(scale(resid(model_name)))<2.5)

The function in (30) creates a subset in which the data points that are more distant than

2.5 points from standard deviation are ignored.

Once the best model is obtained, it is important to assess in which direction the effect of each level goes. This can be done by applying the function summary to the model resulting from the step-up and step-down procedure. The output of this operation is illustrated in Table 11.

Fixed effects	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
(Intercept)	2.0618	0.3319	6.211	5.25e-10 ***
Family_p	-0.2669	0.1559	-1.712	0.086935 .
Family_r	-0.4114	0.1745	-2.358	0.018394 *
Inbetween_signs1	0.7518	0.2229	3.372	0.000746 ***
Inbetween_signs2+	1.8727	0.6585	2.844	0.004459 **

Table 11: Summary of the fixed effects contained in the final model

In Table 11, the level of the dependent variable taken as default value is the prenominal position. The direction of the effects can be observed by looking at the estimate in the second column of the table. Considering participants with hearing parents as default value, we can observe that: i) participants with Deaf parents are significantly more likely to produce nominal modifiers in prenominal position (the estimate has a negative value) and ii) participants with Deaf relatives are even more likely to produce nominal modifiers in prenominal position (the estimate has a negative value which is farther from zero with respect to the previous estimate value). The effect exerted by In-between signs goes in the opposite direction. Indeed, the estimate values corresponding to the levels of this predictor are all positive, rather than negative. Considering nominal modifiers that are adjacent to the noun as default value, we can observe that: i) modifiers separated from the noun by one sign are significantly more likely to appear in postnominal position and ii) modifiers separated from the noun by two signs are even more likely to appear after the noun.

3.5. Conclusions

This chapter focused on the methodological aspects involved in this dissertation. I discussed the tools that I used and the linguistic issues I dealt with.

First, I presented the potentialities and limitations of all the types of data that I considered. The quantitative analysis on nominal modification is based on corpus data coming from the LIS Corpus project (Geraci et al., 2011). After introducing how these data have been collected, I explained how the annotation of spontaneous narratives of 162 participants has been carried out. Then, I provided a step-by-step overview of the statistical computation conducted through the software R. As for the qualitative considerations, I explained how I collected new evidence (i.e. elicited judgments and narration tasks).

In the next chapters, I will show that quantitative and qualitative approaches are not mutually exclusive and that it is possible to combine different sources of data and different methodological procedures in order to shed light on the distribution of nominal modifiers in LIS. Once the significant predictors are identified and the direction of the effect of each level is understood, the formal linguist steps in. What s/he has to do is explain the attested data and provide an account of the observed variation. This will be done in the next chapter.

4. Universal 20 in LIS: a case of sociolinguistic variation

4.1. Introduction

As a minority language, LIS shows the influence of several social factors such as the lack of formal recognition by Italian authorities, the paucity of bilingual programs for deaf pupils at school, the absence of a written form, the constant pressure from the dominant spoken language, etc. Given this sociolinguistic context, it is not surprising that the process of language standardization has not been accomplished yet. Although LIS signers coming from various geographical areas or social background do not have difficulty understanding each other, recent studies showed a considerable degree of variation in several linguistic domains like syntax, phonology, lexicon, and prosody (Cardinaletti, Cecchetto & Donati, eds., 2011; Conte, Santoro, Geraci & Cardinaletti, 2010; Geraci, Battaglia, Cardinaletti, Cecchetto, Donati, Giudice & Mereghetti, 2011). Variation has also been found in the distribution of interrogative signs contained in wh-questions, namely a domain in which order permutations were not expected (Branchini, Cardinaletti, Cecchetto, Donati & Geraci, 2013).

Despite the fact that elicited data show a relatively stable order in the structure of nominal expressions (cf. Section 2.4.2.), corpus data have been analyzed in order to obtain a clearer picture of the dimensions of variation in this domain. Like previous corpus works on LIS, this research is based on the Labovian first wave sociolinguistic approach to language variation (Eckert, 2012, see Section 1.4.1.), and its empirical base is the largest corpus of LIS currently available (Geraci et al., 2011).

Two quantitative studies have been conducted for the purposes of this dissertation. The first study addresses the distribution of modifiers with respect to the head noun. The second study focuses on the duration of nominal modifiers. The results emerging from the statistical analysis conducted on both studies are presented in Section 4.2. Methodological issues (e.g. data collection, protocol of annotation, statistical data

processing) are not discussed here because they have already been extensively addressed in Chapter 3. Results from both studies show that linguistic and social factors are crucial to understand the syntactic distribution of modifiers and their duration. Although I am mainly interested in the syntactic distribution of modifiers and its consequences for Universal 20, results from the second study strengthen those of the first one and, in a sense, give further support to their analysis. The interpretation of the findings from both studies is presented in Section 4.3.[1]

4.2. Analysis

In this section, the results emerging from the statistical analysis conducted on both quantitative studies are presented. For each study, the formula corresponding to the model resulting from the step-up and step-down procedure is reported and discussed. In particular, I focus on the significant factors showing an effect on the position of modifiers in LIS nominal expressions and their duration.

4.2.1. The distribution of nominal modifiers

The dependent variable investigated in the first study is the position of nominal modifiers with respect to the head noun. Four options were in principle expected: i) prenominal modifier (modifier>noun); ii) postnominal modifier (noun>modifier); iii) repetition of the modifier (modifier>noun>modifier); iv) repetition of the head noun (noun>modifier>noun). An example for each of these options is represented with glosses in (1), where productions taken from the corpus are provided. The relevant nominal expressions are indicated by square brackets and are boldfaced.

(1) a. modifier>noun

TRY GO-FOR-IT, FIRST STEP [**BEAUTIFUL EXPERIENCE**] GO-AHEAD

1 The research work reported in this chapter has been refined thanks to useful comments and observations coming from the audiences of the conferences FEAST (Warsaw, June 2012) and TISLR (London, July 2013).

'Give it a try, go for it. After the first step you'll see it's a beautiful experience and you'll go ahead with it!'

b. noun>modifier

LEAVE PARIS GO-TO [**EXPERIENCE BEAUTIFUL**]

'I left and went to Paris, it was a beautiful experience'

c. modifier>noun>modifier

NOW ANYMORE, GO-BACK IX_LOC NO, MUST [**ANOTHER JOB ANOTHER**]

'At that time I didn't go back there, I had to find another job'

d. noun>modifier>noun

THERE-IS [**BED ACCOMMODATION TWELVE ACCOMMODATION**]

'It has twelve berths'

The modifiers in (1) (BEAUTIFUL, ANOTHER, TWELVE) directly modify the head noun. Remember from Section 3.3.2. that absence of prosodic break between modifier and head noun and spreading of NMMs are considered as indicators of constituency so that the modifier should not be analyzed as a DP-external item (e.g. predicate, parenthetical repetition, etc.).

The overall distribution of modifiers with respect to the head noun is illustrated in Figure 1. The most frequent order observed is noun>modifier (1216 tokens, 60%). The second most frequent option is the order modifier>noun (692 tokens, 34%). The repetition cases are not frequent in the corpus: 4% for the duplication of the modifier (88 tokens) and 2% for the duplication of the head noun (27 tokens).

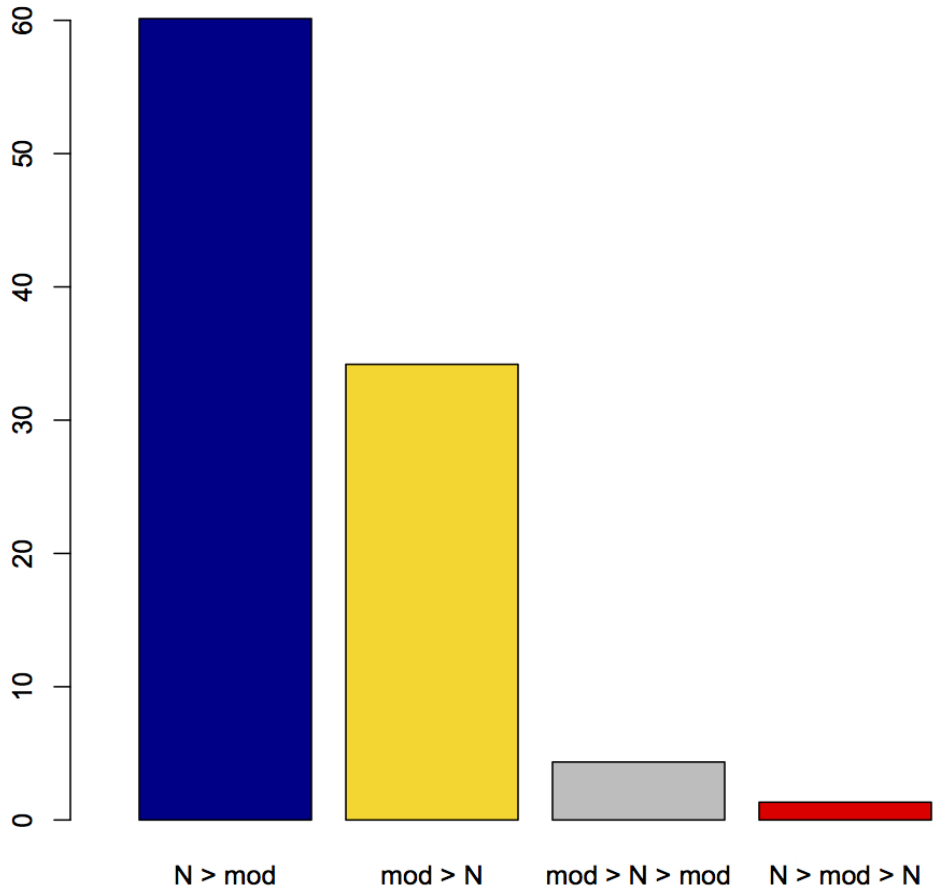


Figure 1: General distribution of nominal modifiers in the corpus

Although interesting and never described in the literature of LIS before, the cases of repetition are too few in order to conduct a systematic quantitative analysis and are therefore excluded from the dataset.[2] The new dataset containing only the two most frequent order options has a total of 1908 tokens. It includes 1216 postnominal modifiers (64%) and 692 prenominal modifiers (36%). This dataset provided the empirical basis for the binomial statistical analysis.

A preliminary observation that can be drawn is that the percentage of prenominal modifiers (36% in the dataset without folding cases) is too large to be considered as epiphenomenal or a marginal aspect of LIS. Rather, it calls for a more serious

2 It might be interesting to analyze these repetition cases in future research. Repetition is a well-attested phenomenon in LIS and is present in other syntactic domains of this language. Geraci, Cecchetto and Zucchi (2008) briefly discuss repetitions in sentential complement constructions. Geraci and Bayley (2011) and Branchini, Cardinaletti, Cecchetto, Donati, and Geraci (2013) discuss repetitions of wh-signs (see Volterra, 1987 for a description of morphological reduplication in LIS).

investigation.

A close inspection of order distribution with respect to the type of modifiers classified according to Cinque's hierarchy reveals the interesting pattern illustrated in Figure 2.

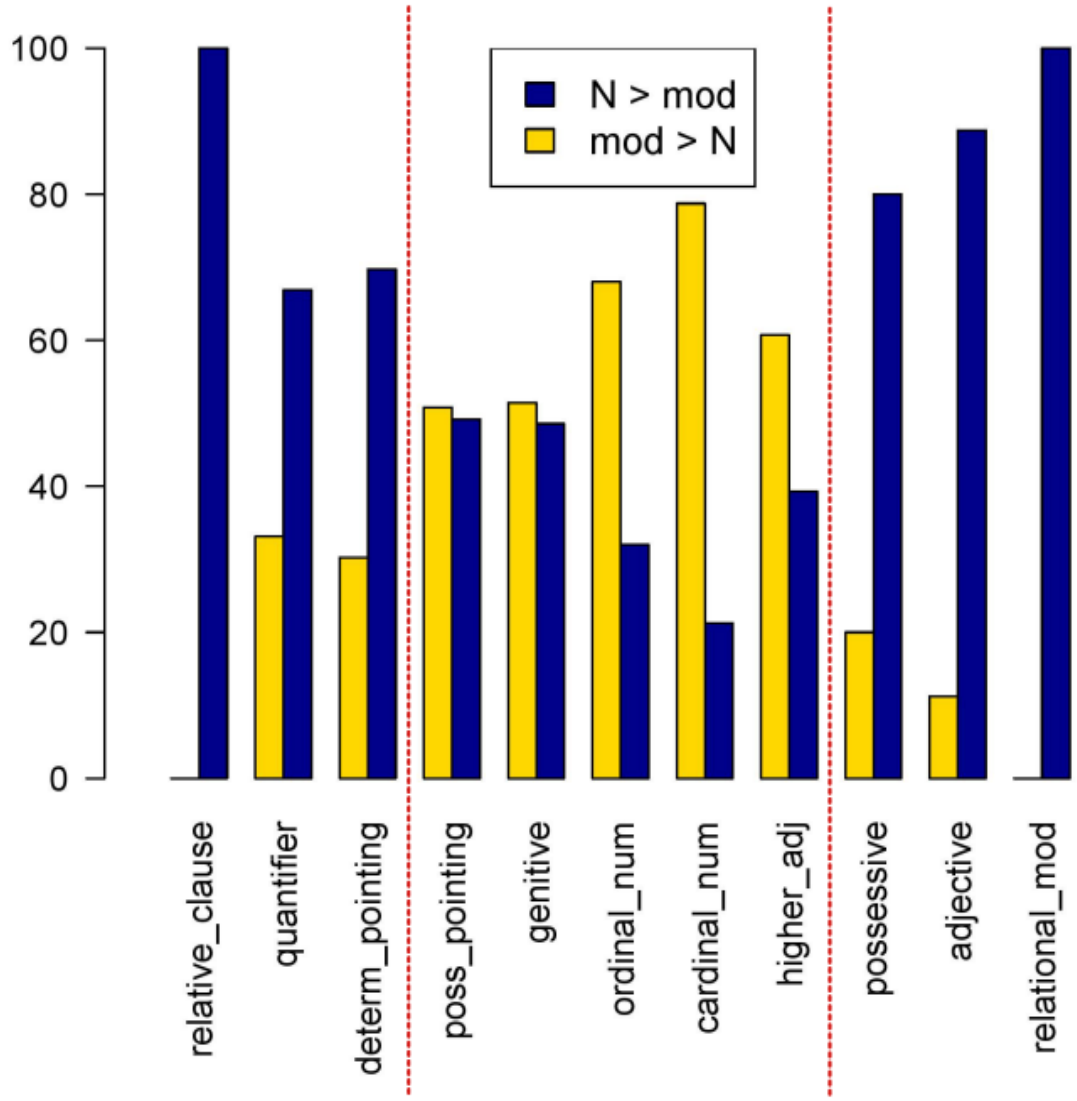


Figure 2: Distribution of nominal modifiers according to their typology

In Figure 2, the different types of modifiers are ordered from left-to-right following Cinque's (2012) proposal. Modifiers occurring to the left are hierarchically higher than those coming to the right. As discussed in Section 2.3.1., this hierarchy has been enriched by including some nominal modifiers not considered in Cinque (2012), namely genitives, possessive modifiers,[3] higher adjectives, and relational modifiers.

3 The position of possessive modifiers in the structure of LIS nominal expressions will be tackled in

In the corpus data analyzed for this dissertation, I found three different signs expressing the notion of possession, namely pointing signs with a possessive meaning and two non-pointing signs with a possessive meaning.[4] Pointing possessives exhibit the index finger extended toward the location associated with the possessor. One of the pointing possessives found in the corpus is reported in the example in (2) and is shown in Figure 3.

(2) PASSION IX-1_POSS WHICH?

'What is my passion?'

Figure 3: IX-1_POSS, pointing possessive

In most of the occurrences found in the corpus, the possessor is the signer himself, therefore the pointing sign is directed toward the signer's chest. Phonologically, pointing signs with a possessive meaning are articulated exactly like determiner-like pointing signs. Sentential prosody and the overall syntactic configuration help distinguish between the two types of indexes. Whether the two types of pointing signs occupy the same syntactic position is left as open issue.

The other two types of possessives have a different phono-morphological makeup. The

the discussion in Section 4.3.1.

4 As for the attributive/predicative status of possessives (cf. Padden, 1988), all the items annotated in the corpus are cases of DP-internal attributive possessives.

first one displays an index finger handshape with wrist rotation toward the location of the possessor. It is exemplified in (3) and shown in Figure 4.

(3) MUM POSS-1_WRIST NAME S-I-L-V-I-A

'My mum's name is Silvia'

Figure 4: POSS-1_WRIST: possessive with wrist rotation

The other non-pointing possessive is characterized by an open-hand handshape facing the location of the possessor and moving toward it.

(4) PROBLEM FAMILY POSS-1_OPEN OBLIGED MOVE IX_LOC ITALY IX_LOC

'Because of my family's problems I had to move here (to Italy)'

Figure 5: POSS-1_OPEN: possessive with open-hand handshape

From a quantitative point of view, the two non-pointing possessives show a similar syntactic distribution in that they are both predominantly postnominal, as shown in Table 1.

	Poss>N		**N>Poss**	
Pointing possessive	92	51%	89	49%
Non-pointing possessive with wrist rotation	15	19%	62	81%
Non-pointing possessive with open-hand handshape	7	20%	28	80%

Table 1: Distribution of possessives in the corpus

The distributional patterns in Table 1 suggest splitting the group of possessives into two categories, namely pointing possessives and non-pointing possessives. Pointing possessive modifiers may distribute both before (51%) and after (49%) the head noun without any strong preference. Interestingly, in the corpus they never co-occur with genitives. On the contrary, non-pointing possessives preferably follow the noun (around 80%), thus confirming the split between the two categories. In some cases, in the corpus

both non-pointing possessives with wrist rotation and non-pointing possessives with open-hand handshape are found in co-occurrence with genitives. This is shown in the two examples reported in (5) (cf. possessor doubling construction in Section 2.3.1.).

(5) a. TEACHER METHOD POSS-3_WRIST DEAF SUITABLE

'The teacher's method was suitable for deaf (students)'

b. PE[5] […] UNCLE DREAM POSS-3_OPEN

'That was my uncle's dream'

It is tempting to motivate the contrast between pointing possessives and non-pointing possessives by assuming that they occupy two different syntactic positions. Referring to Alexiadou et al.'s (2007) analysis implemented in Cinque's (2005, 2010, 2012) framework, the two types of non-pointing possessives are considered strong possessives occurring in a low structural position, whereas pointing possessives are considered deficient possessives occupying a higher structural position. For the time being, I do not speculate whether they correspond to weak or clitic possessives. More data are needed to investigate their nature in detail.[6] Concretely, for the purposes of this dissertation, the two types of possessives that mainly occur in postnominal position have been collapsed into one category. These modifiers, which are referred to as non-pointing possessives, occupy one of the projections in the lower portion of the structure (see below). Pointing possessives display a different syntactic distribution. Although they show a balanced pattern, I treat this type of possessives as occupying one of the projections in the middle portion of the structure (see below). In this respect, the central area of the DP contains modifiers that are clearly not postnominal.

5 The sign conventionally labeled as PE is realized with extended index finger and a special downward rotation of the wrist. In the example reported in the text, the sign PE functions as an anaphoric demonstrative. Since it is not discussed in detail here, the reader is referred to Cecchetto, Geraci and Zucchi (2006), Branchini (2007), and Brunelli (2011).

6 On the basis of analyses on spoken languages, the tripartition including clitic, weak, and strong elements has been proposed by Cardinaletti (1998) for possessive modifiers and by Cardinaletti and Starke (1999) for personal pronouns. Bertone (2007) studies the pointing signs in LIS that function as personal pronouns. Bertone adopts Cardinaletti and Starke's (1999) classification and proposes that the three classes of pronouns are identifiable in LIS on the basis of duration: clitic indexes last 1/12 sec., weak indexes 1/6 sec., and strong indexes 1/2 sec (also see Bertone & Cardinaletti, 2011).

Turning back to Figure 2, it is clear that modifiers do not distribute uniformly. Modifiers occurring in the lowest and highest positions of Cinque's hierarchy tend to occur in postnominal position (blue bars), while modifiers occurring in the central part of Cinque's hierarchy tend to be more prenominal (yellow bars). This fact better qualifies the preliminary observation that a 36% of prenominal modifiers could not be considered like a marginal phenomenon of LIS. Rather, this situation calls for a more systematic and principled explanation.

In order to better capture this distributional variation, the different types of modifiers illustrated in Figure 2 are collapsed into three groups corresponding to three sections of the hierarchical structure: i) modifiers in the higher portion of the structure showing a preference for the postnominal position; ii) modifiers in the intermediate portion of the structure showing a preference for the prenominal position; iii) modifiers in the lower portion of the structure showing a preference for the postnominal position.[7] The collapsed levels and the corresponding categorization are illustrated in (6). From a statistical point of view, the predictor Modifier type becomes a variable with three levels (high, middle, and low modifiers).

(6) a. High modifiers: relative clauses, quantifiers, determiner-like pointing signs[8]

b. Middle modifiers: possessive pointing signs, genitives, cardinal numerals, ordinal numerals, higher adjectives

c. Low modifiers: possessives, adjectives, relational modifiers

To illustrate, examples for each type of modifier in each group are provided below. When attested, modifiers are shown both in prenominal and postnominal position.

7 In principle, the levels contained in the variable Modifier type could be collapsed differently. More specifically, the different types of modifiers could be classified into two categories, namely peripheral portion (high and low modifiers) and central portion (intermediate modifiers) of the hierarchy. However, from the theoretical perspective that will be adopted in the discussion, it is important to maintain a tripartite classification (high, middle, and low modifiers) because the analysis of the data proves to be more coherent. The procedure in which similar levels are not combined into one single level is discussed in Tagliamonte (2006: 150).

8 The pointing signs included in the corpus have been classified according to two categories, namely determiner-like and possessive pointing signs. The two classes have been distinguished on the basis of the different annotation (cf. Section 3.3.1.).

Examples of high modifiers are provided in (7), middle modifiers in (8), and low modifiers in (9). The relevant nominal expressions are indicated by square brackets and boldfaced.

(7) a. N>Relative Clause [Old signer from Rome]

[**ONE KNOW INSIDE OFFICE**] … IX-1 ASK PLEASE $_{2}$EXPLAIN$_{1}$

'I asked a person I know in the office to give me a clarification'

b. Quantifier>N [Old signer from Bologna]

FINGERSPELLING [**ALL DEAF**] KNOW WELL

'All Deaf people were proficient at fingerspelling'

b'. N>Quantifier [Middle-aged signer from Salerno]

[**HEARING ALL**] COMPLAIN

'All hearing people were complaining (about that)'

c. Determiner-like pointing>N [Middle-aged signer from Lamezia]

IX-1 SCHOOL INSTITUTE IX-1 INSIDE, [**IX_DET FATHER**] FOUR YELL

'When I was at the institute, my father used to yell at the four of them'

c'. N>Determiner-like pointing [Young signer from Bologna]

IX-1 SECRETARY, [**BOSS IX_DET**] $_{3}$GIVE-ORDER$_{1}$ LIKE-NEG

'As a secretary, I didn’t like when my boss gave me orders'

(8) a. Possessive pointing>N [Middle-aged signer from Brescia]

[**IX-1_POSS HUSBAND**] VENETIAN

'My husband is Venetian'

a'. N>Possessive pointing [Old signer from Milan]

[**HOUSE IX-1_POSS**] CLOSE

'My house was close by'

b. Genitive>N [Young signer from Ragusa]

IX-1 BECOME DEDICATED LEARN [**DEAF SIGN**] IX-1 LEARN

'I dedicated myself to learn Deaf people's sign language'

b'. N>Genitive [Old signer from Turin]

PERSON++ HEARING … KNOW [**PROBLEM DISABLED**]

'Hearing people are aware of disabled people's problems'

c. Cardinal>N [Old signer from Milan]

(IX-1) TAKE [**ONE GOLD**], [**TWO SILVER**], [**THREE BRONZE**]

'I won one gold, two silver, and three bronze (medals)'

c'. N>Cardinal [Young signer from Bologna]

IX-1 [**PASSION TWO**] SNOWBOARD TRAVEL

'I have two passions: snowboarding and traveling'

d. Ordinal>N [Old signer from Florence]

[**SECOND HUSBAND**] CHILD THREE

'I had three children with my second husband'

d'. N>Ordinal [Old signer from Lamezia]

[**CHILD FIRST**] HEARING, SECOND DEAF

'My first child is hearing, the second one is deaf'

e. Higher adjective>N [Old signer from Lamezia]

[**OTHER FRIEND DEAF**] COMMUNICATION THERE-IS

'I am in touch with other Deaf friends'

e'. N>Higher adjective [Middle-aged signer from Rome]

WANT NEG IX-1 [**FRIEND OTHER**] NEG

'S/he didn't want me to have other friends'

(9) a. Possessive>N [Middle-aged signer from Florence]

CHILD HEARING, [**POSS-1_WRIST MOTHER-IN-LAW**] HAPPY

'My mother-in-law rejoiced after I gave birth to a hearing child'

a'. N>Possessive [Middle-aged signer from Ragusa]

[**FAMILY POSS-3_WRIST**] [**FAMILY POSS-1_WRIST**] TOGETHER GO-TO ROME

'His family and my family went to Rome'

b. Adjective>N [Middle-aged signer from Florence]

[**BIG FESTIVAL**] SOMEONE GO-ONSTAGE

'At this big festival, someone appeared onstage'

b'. N>Adjective [Old signer from Ragusa]

WAIT-FOR [**PRESIDENT NEW**]

'We are looking forward to our new president'

c. N>Relational modifier [Young signer from Trani]

SOON INSIDE TYPE VOLUNTARY-WORK, [**SERVICE CIVIL**]

'Soon I began with some sort of voluntary work, civil service'

Following the procedure illustrated in Section 3.4.4., the position of modifiers with respect to the head noun is analyzed employing mixed-effects models (Baayen, Davidson & Bates, 2008). Since the dependent variable is binomial (noun>modifier vs. modifier>noun order) a logistic model was applied (Jaeger, 2008). A step-up and step-down procedure with participant and item introduced as random factors has been conducted considering linguistic and social factors as potential predictors of the position of modifiers. The list of social and linguistic predictors and their respective levels has already been shown in Sections 3.2.1. and 3.3.2. For the sake of convenience, the complete list is reported unitarily in Table 2. As explained in Section 3.4.3., for some variables the levels showing similar distribution have been collapsed. Both the original variables and the recoded variables containing collapsed levels are shown in the table

below.

Linguistic predictors and their levels	**Linguistic predictors and their levels after collapsing**
Modifier_type: quantifier, pointing, ordinal numeral, cardinal numeral, possessive, genitive, higher adjective, adjective, relational modifier, relative clause	Modifier_type2: high (relative clause, quantifier, determiner_pointing), middle (possessive_pointing, genitive, ordinal numeral, cardinal numeral, higher adjective), low (possessive, adjective, relational modifier)
Inbetween_signs: adjacent signs, one intervening sign, two intervening signs	Inbetween_signs2: adjacent signs, non-adjacent signs
Number_of_modifier: one, two, three or more modifiers	Number_of_modifier2: single modifier, multiple modifiers
Clause_type: main clause, argument clause, adjunct clause	-
Social predictors and their levels	**Social predictors and their levels after collapsing**
Gender: female, male	-
Age: range 18 – 81	
Age_group: young, middle-aged, old	-
Geographical_provenance: Turin, Milan, Brescia, Bologna, Florence, Rome, Salerno, Trani, Lamezia, Ragusa	-
Residence_status: urban, rural	-
Family: hearing, Deaf parents, Deaf relatives	-
Job: white collar, blue collar, professional, student, unemployed	Job2: working (white and blue collar, professional), non-working (student and unemployed)
School_qualification: primary, middle, high	-
Register: low, medium, high	Register2: low (low, medium), high

Table 2: List of predictors and levels for the study on position. Dependent variable: Position2 (levels: prenominal, postnominal)

The step-up and step-down procedure converged over the same model, which consisted of two linguistic (Modifier type and In-between signs), and three social (Age, Family, and Register) fixed effects. A significant interaction was found between Modifier type and Gender, although Gender was not selected among the significant parameters during the step-up and step-down procedure.[9]

The formula of the best model resulting from the step-up and step-down procedure is reported in (10).

(10) bestmodel_position = glmer (Position2 ~ rcs (Age, 3) + Modifier_type2 * Gender + Register2 + Family + Inbetween_signs2 + (1|Gloss) + (1|Name), family = binomial, na.action = na.omit, data = DP)[10]

For some predictors, the collapsed configuration (i.e. 'Modifier_type2', 'Inbetween_signs2', 'Register2') has been selected because of its better explicatory strength. The predictor Age has been included in the model within the function 'rcs ()', which stands for 'restricted cubic spline' and is used to capture non-linear distributions. The numerical coefficient included in this function indicates the number of knots contained in the spline. Knots are generally used to smooth splines. In this case, a spline with three knots was applied to the continuous variable Age. The number of knots has been established by comparing different options (i.e. models with a different number of knots) with 'anova ()'.

Outliers have been removed in order to exclude atypical responses, as illustrated in the formula in (11). The function 'subset ()' has been added to exclude those data points that are distant more than 2 points with respect to the standard deviation. Overall, the data points that have been excluded are 74.

9 In the statistical analysis, Gender is maintained as fixed effect because it proved to be significant in the interaction with Modifier type. According to Baayen (2008: 165), the model should include both the interaction and the single main effects involved in the interaction.

10 The function 'na.action = na.omit' is used to treat possible empty cells in the dataset.

(11) bestmodel_position_subset = glmer (Position2 ~ rcs (Age, 3) + Modifier_type2 * Gender + Register2 + Family + Inbetween_signs2 + (1|Gloss) + (1|Name), family = binomial, na.action = na.omit, data = DP, subset = abs (scale (resid (bestmodel_position))) < 2)

The parameters estimated by the final model are listed in Table 3. The reference level of the dependent variable is the order modifier>noun. As already explained in Section 3.4.4., the direction of the effects can be observed by looking at the estimate in the second column of the table. Values reported in this column indicate how likely it is to find a modifier in postnominal position (i.e. the opposite of the reference value) according to the levels of the fixed factors included in the model. When significant ($p<0.05$, marginally significant if $p<0.1$), positive values imply that modifiers are more likely to appear postnominally, whereas negative values indicate that modifiers are more likely to appear prenominally.

Fixed factors	Estimate	Std. Error	Z value	Pr (>|z|)
(intercept)	1.48345	0.60059	2.470	0.0135 *
Mod.type: high	2.29701	0.36505	6.292	3.13e-10 ***
Mod.type: low	3.05460	0.30455	10.030	< 2e-16 ***
In-between signs: adjacent	-0.93421	0.22755	-4.105	4.03e-05 ***
Age: age	-0.02874	0.01319	-2.180	0.0293 *
Age: age'	0.02452	0.01350	1.816	0.0693 .
Gender: male	0.36750	0.21462	1.712	0.0868 .
Register: low	-0.16753	0.17390	-0.963	0.3353
Family: Deaf parents	-0.42874	0.18317	-2.341	0.0192 *
Family: Deaf relatives	-0.29798	0.17968	-1.658	0.0972 .
Mod.type * Gender (high * male)	-0.57504	0.31927	-1.801	0.0717 .
Mod.type * Gender (low * male)	-0.84143	0.35519	-2.369	0.0178 *

Table 3: Mixed-Effect Analysis. Parameters Estimated by the Final Model and their Statistical Significance Analysis. Model's goodness of fit values: AIC 1594, BIC 1669 and logLik -783.

In the remainder of the section, I analyze the effect of the significant predictors included in the final model. I do not discuss the effect of Gender and Register which proved to be non-significant predictors (see Table 3). However, it is important to clarify the reason why these two predictors have been introduced in the formula of the final model. Gender has not been selected as a simple main effect, but it has been considered in its interaction with Modifier type. The effect of this interaction will be later discussed in this section. Register is maintained in the formula in (11) because it has an overall contribution to the goodness of fit of the general model (i.e. it is selected as significant predictor in the step-up and step-down procedure).

As for the linguistic factors, the direction of the effect of Modifier type indicates that modifiers sitting in the higher and lower portion of Cinque's (2012) hierarchy are more likely to occur in postnominal position. This tendency is shown in Figure 6.

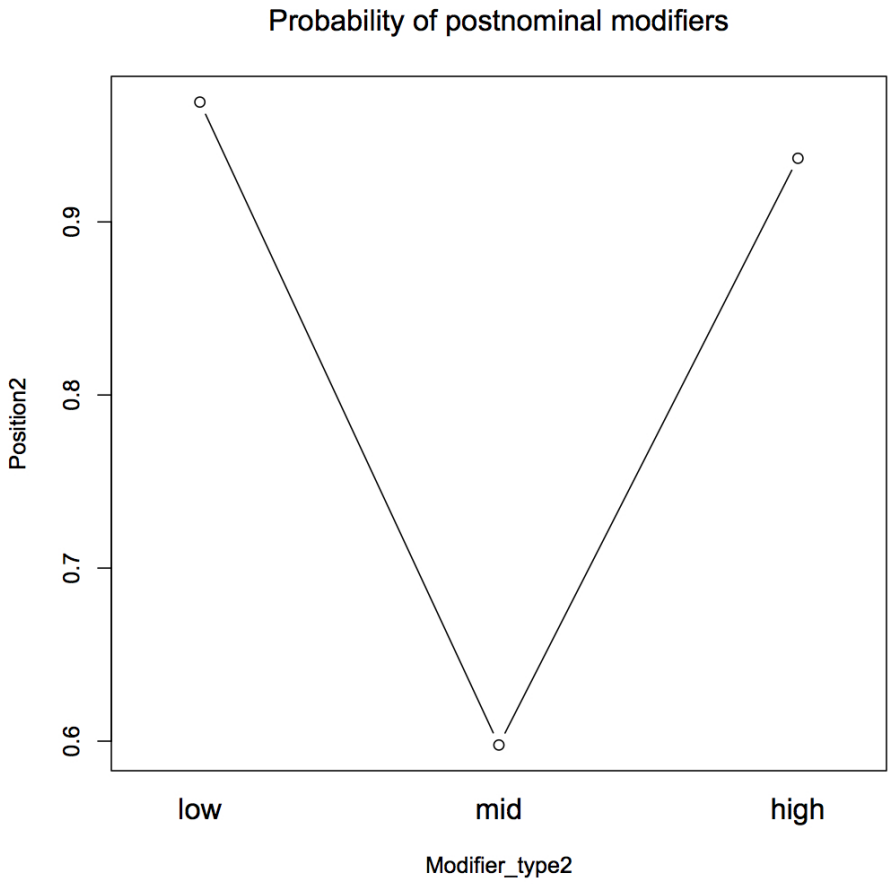


Figure 6: Probability for modifiers to occur in postnominal position according to their typology

The effect of In-between signs indicates that there is a significant difference between modifiers that are adjacent and modifiers that are non adjacent to the head noun (p = 4.03e-05). As shown in Figure 7, the former are less likely to occur in postnominal position.

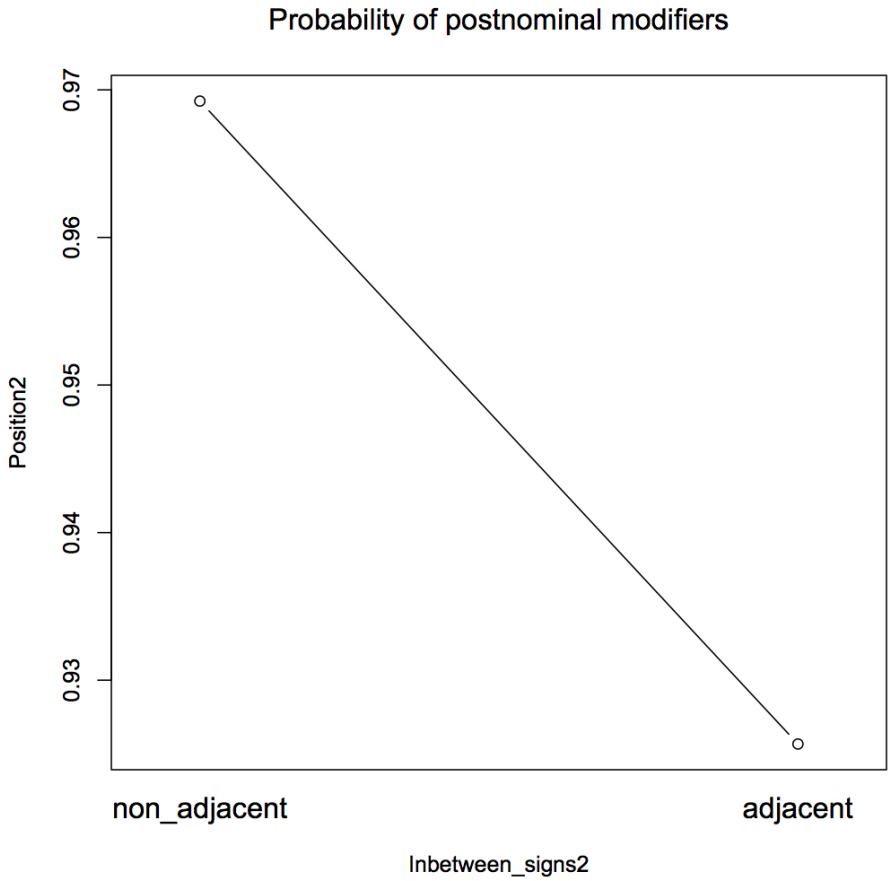


Figure 7: Probability for modifiers to occur in postnominal position according to their adjacency to the noun

As for the social factors, the continuous variable Age shows a significant effect on the position of modifiers. As illustrated in Figure 8, the signers born before 1945 and the signers born after 1965 are more likely to produce postnominal modifiers.

Figure 8: Probability for modifiers to occur in postnominal position according to signers' age

Another social variable that exerts an influence on the dependent variable is Family. As shown by the values reported in Table 3, signers with Deaf parents show a significantly different behavior with respect to signers coming from hearing families. Specifically, they are less likely to produce postnominal modifiers. Signers with a Deaf relative are somehow in between the other two categories, as shown in Figure 9.

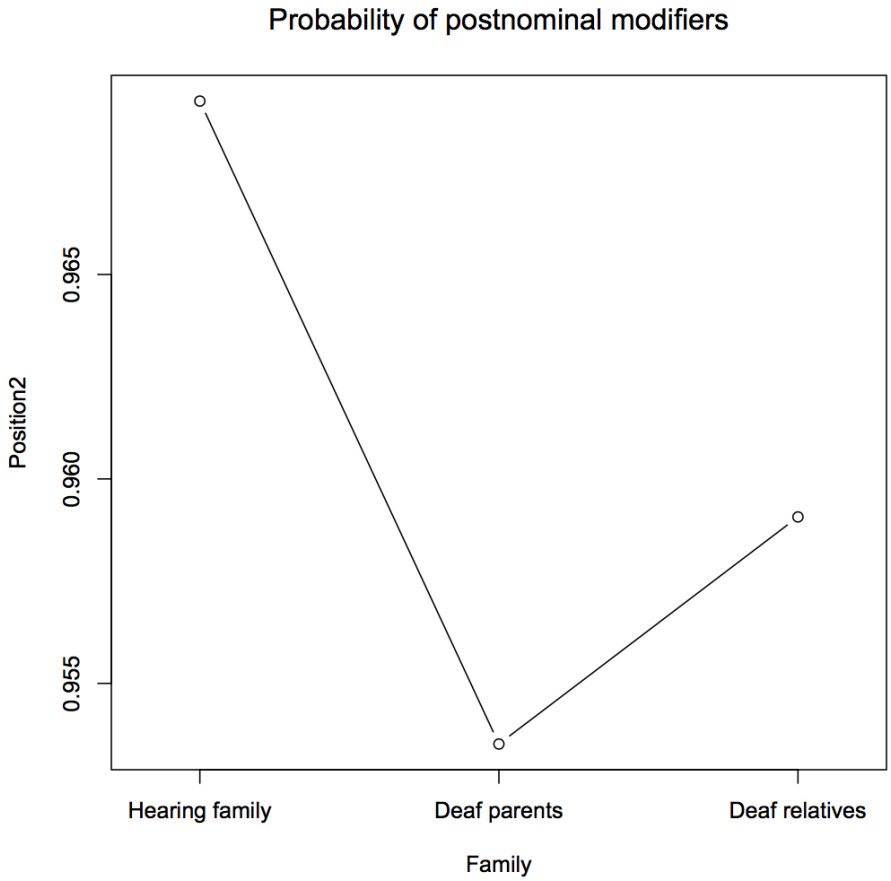


Figure 9: Probability for modifiers to occur in postnominal position according to signers' family

A significant interaction is found between Modifier type and Gender. Women show a more extreme pattern than men because they are more likely to produce modifiers belonging to the central part of Cinque's hierarchy in prenominal position and modifiers sitting in the peripheral projections in postnominal position. This pattern is shown in Figure 10.

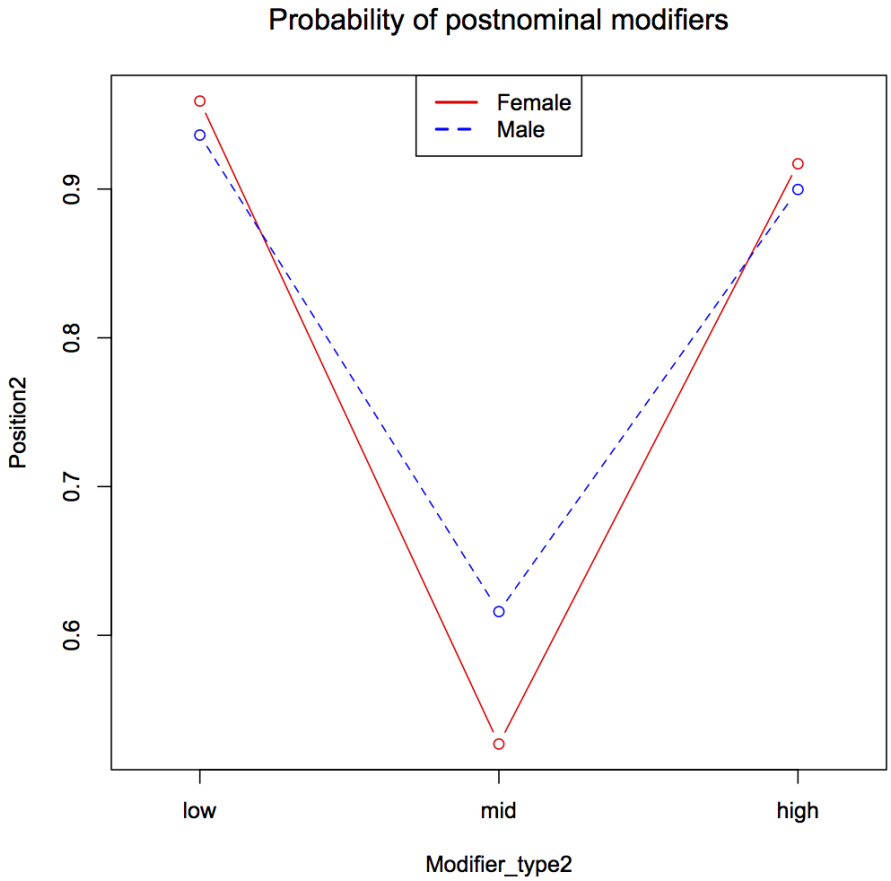


Figure 10: Probability for modifiers to occur in postnominal position according to the interaction between modifier type and gender

I will discuss the impact of these results for the syntax of LIS nominal expressions in Section 4.3.

4.2.2. The duration of nominal modifiers

In the second quantitative study, potential differences in the duration of nominal modifiers were investigated. In addition to a strong lexical effect, the syntactic effect of the positioning of modifiers vis-à-vis the noun and the effect of social factors were expected.

Indeed, all prosodic factors being equal, cardinal numerals and pointing signs (of all morphosyntactic categories) are expected to be shorter than the other modifiers for articulatory reasons. Pointing signs are known to have a short duration, especially when

used as clitic or weak elements (Bertone, 2007; Bertone & Cardinaletti, 2011).[11] Digits (intending cardinal numerals from '0' to '10') are expected to be quite shorter than other signs (Geraci, 2009),[12] therefore an effect of cardinals is also assumed.

As for the position of modifiers (this was the dependent variable of the previous study), I predict that modifiers at the final boundary of the prosodic constituent are more likely to be produced with long duration because of phrase-final lengthening (see Branchini et al., 2013 for a similar effect of duration on repeated wh signs and on the aspectual marker DONE).

Finally, as for social factors, Age is expected to exert an influence on the duration of modifiers. Because of physiological reasons the older the participant, the more likely s/he is to produce modifiers with slower movements, and hence longer duration. In addition to this, however, other social factors (Gender, Family, and Geographical provenance) may possibly contribute to the variation of modifier length. The effects of these social factors are expected to be coherent with those emerged as significant in the first study (cf. Section 4.2.1) and hence to reinforce the explanation provided for them.

The logarithmic conversion of duration used to normalize the distribution (Baayen, 2008) was analyzed employing mixed-effects models. A step-up and step-down procedure was then conducted adopting the same criteria of inclusion of significant parameters as in the first study. Participant and item were introduced as random factors and the analysis has been carried out considering both linguistic and social factors as potential predictors of the duration of modifiers.

The complete list of predictors and their levels is provided in Table 4. Both the original variables and the recoded variables containing collapsed levels are shown in the table below.

11 Bertone (2007: 156-157) claims that some (weak) demonstratives show such a short duration that they are barely perceivable. According to Bertone and Cardinaletti (2011), pronouns can fall into three categories, namely clitic, weak, and strong pronouns. Clitic pronouns are produced very quickly showing a duration of less than 200 milliseconds. The duration of weak pronouns is usually in the range of 200-300 milliseconds. Strong pronouns are articulated in a slower way so that their duration exceeds 300 milliseconds. Details on the duration of pointing signs have been collected by counting the frames between the start and end point. The start point corresponds to the moment in which the index configuration is headed toward a point in the signing space. The end point corresponds to the moment in which the pointing sign ends before the hand starts articulating the following sign.

12 Geraci (2009) shows that digits, contrary to other signs, do not have the movement component but are basically represented with a hold. If a signer is asked to produce a digit in isolation, s/he will articulate it without adding any particular path movement (apart from interpolation movements).

Linguistic predictors and their levels	**Linguistic predictors and their levels after collapsing**
Position2 (binomial): prenominal, postnominal	-
Modifier_type: quantifier, pointing, ordinal numeral, cardinal numeral, possessive, genitive, higher adjective, adjective, relational modifier, relative clause	Modifier_type2: card_point (det. pointing, poss. pointing, cardinal), other (quantifier, ordinal, possessive, genitive, higher adjective, adjective, relational modifier, relative clause)
Inbetween_signs: adjacent signs, one intervening sign, two intervening signs	Inbetween_signs2: adjacent signs, non-adjacent signs
Number_of_modifiers: one, two, three or more modifiers	Number_of_modifiers2: single modifier, multiple modifiers
Clause_type: main clause, argument clause, adjunct clause	-
Social predictors and their levels	**Social predictors and their levels after collapsing**
Gender: female, male	-
Age: range 18 – 81	
Age_group: young, middle-aged, old	-
Geographical_provenance: Turin, Milan, Brescia, Bologna, Florence, Rome, Salerno, Trani, Lamezia, Ragusa	Geographical_provenance2: BrTuSa (Brescia, Turin, Salerno), other (Milan, Bologna, Florence, Rome, Trani, Lamezia, Ragusa)
Residence_status: urban, rural	-
Family: hearing, Deaf parents, Deaf relatives	-
Job: white collar, blue collar, professional, student, unemployed	Job2: working (white and blue collar, professional), non-working (student, unemployed)
School_qualification: primary, middle, high	-
Register: low, medium, high	Register2: low (low, medium), high

Table 4: List of predictors and levels for the study on duration. Dependent variable:

logduration.

The step-up and step-down procedure converged over the same model, which consisted of two linguistic (Position and Modifier Type) and five social (Age, Register, Family, Gender, and Geographical provenance) fixed effects. A significant interaction was found between Modifier Type and Gender.

The formula of the best model resulting from the step-up and step-down procedure is reported in (12). Unlike the previous study on the position of modifiers, this study revolves around a continuous dependent variable (i.e. duration). As a consequence, it requires the function 'lmer ()' instead of the function 'glmer ()'.

(12) bestmodel_duration = lmer (logduration ~ Age + Position2 + Register2 + Family + Modifier_type2 * Gender + Geographical_provenance2 + (1|Gloss) + (1|Name), data = DP)

For some predictors, the collapsed configuration (i.e. 'Position2', 'Modifier_type2', 'Register', 'City2') has been selected because of the better explicator strength. Age has not been included in a rcs function because a regular linear distribution was expected without any particular alteration.[13] The duration of the nominal modifiers was expected to increase along with the participants' age for physiological reasons. Outliers have not been removed because it has been observed that this operation does not improve the goodness of fit of the model.

The parameters estimated by the final model are listed in Table 5.[14] The estimate values in the second column show how likely it is to find modifiers with long duration (positive estimate values) or short duration (negative estimate values).

13 In the mixed-effects model, both Age and rcs (Age) have been tested. Since the rcs function did not improve the goodness of fit of the model, it has not been further considered.

14 The threshold of significance for t-values is |2| (Baayen, 2008: 70). This means that if the t-value is an absolute value greater than 2, then it can be considered statistically significant.

Fixed factors	Estimate	Std. Error	t value
(intercept)	5.193515	0.110685	46.92
Position: postnominal	0.260137	0.035385	7.35
Mod.type: card_point	-0.222185	0.055770	-3.98
Age	0.001954	0.001595	1.23
Register: low	-0.037696	0.053411	-0.71
Gender: female	0.110799	0.052499	2.11
Family: Deaf parents	-0.160507	0.055877	-2.87
Family: Deaf relatives	-0.065167	0.056589	-1.15
Geographical provenance: other	0.214169	0.046608	4.60
Mod.type * Gender (card_point * f)	-0.135397	0.060605	-2.23

Table 5: Mixed-Effect Analysis. Parameters Estimated by the Final Model and their Statistical Significance Analysis. REML criterion at convergence: 2937.9.

As in the study concerning the distribution of modifiers, the analysis here revolves around the significant predictors included in the final model. I do not discuss the effect of Age and Register because the difference between their respective levels proved to be statistically non-significant (see Table 5). These two fixed factors have been maintained in the formula in (12) because they have an overall contribution to the goodness of fit of the general model.

Looking at the linguistic factors included in the final model, Position emerges showing a strong effect (t = 7.35). The direction of the effect indicates that prenominal modifiers are shorter than postnominal ones. This tendency is shown in Figure 11.

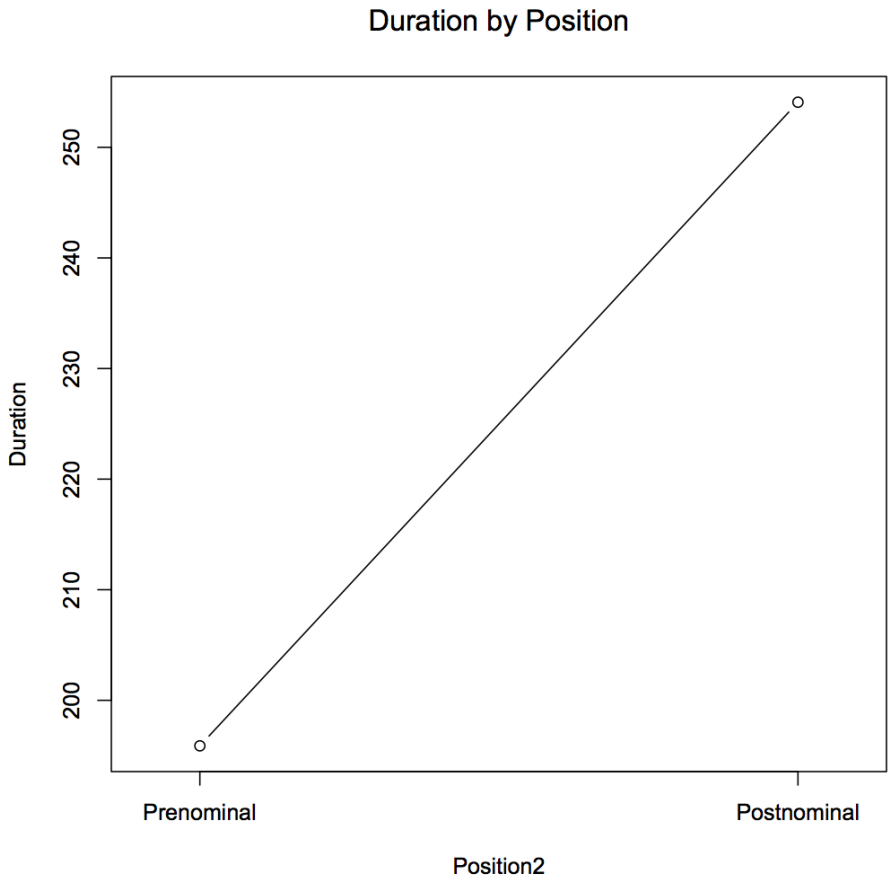


Figure 11: Duration of modifiers according to their position with respect to the head noun

The other significant linguistic variable is Modifier type (t = -3.98). Pointing signs (both possessive and determiner-like pointing signs) and cardinal numerals behave differently than the other types of modifiers. As expected, they are produced with a shorter duration. This tendency is illustrated in Figure 12.

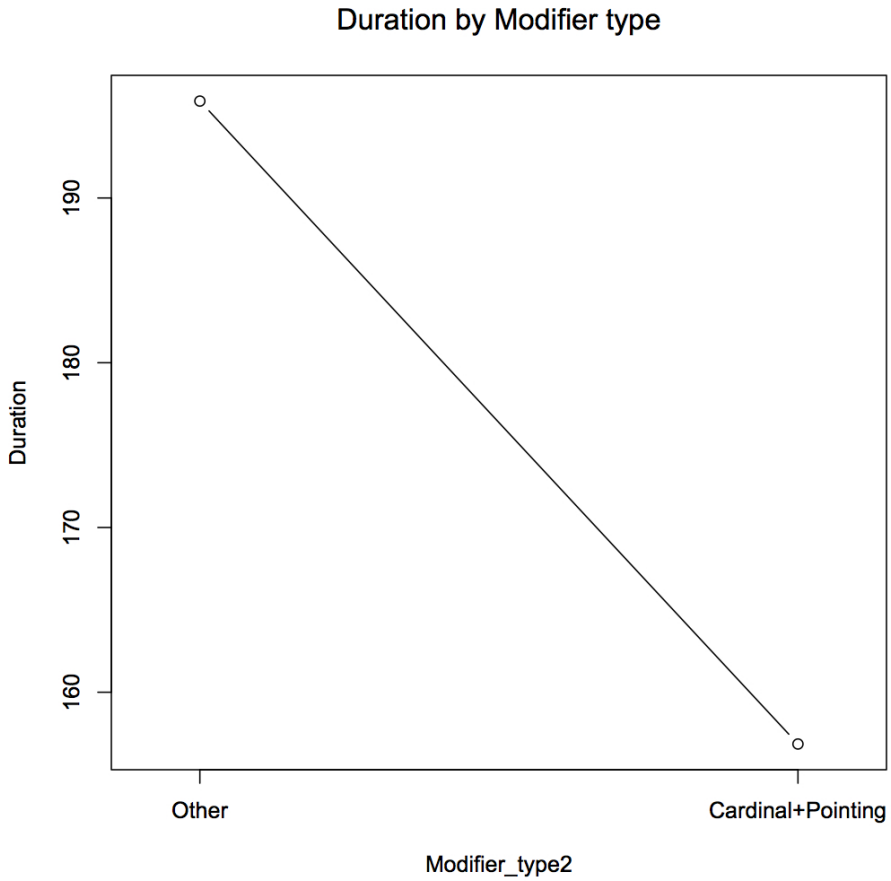


Figure 12: Duration of modifiers according to their typology

One of the social predictors that exert an influence on the dependent variable is Gender (t = 2.11). As shown in Figure 13, women are overall slower signers than men (but see also the effect of the interaction with the Modifier type parameter).

Figure 13: Duration of modifiers according to signers' gender

With respect to the family background, significant differences are found. As shown by the values reported in Table 5, signers with Deaf parents show a significantly different behavior with respect to signers coming from hearing families. Specifically, they are faster in producing nominal modifiers. Signers with a Deaf relative are somehow in between the other two categories, as illustrated in Figure 14.[15]

15 This result is similar to the one found in the study on the position of modifiers thus confirming that signers from Deaf parents behave differently than signers from hearing families and that signers with at least one Deaf relative show an intermediate pattern.

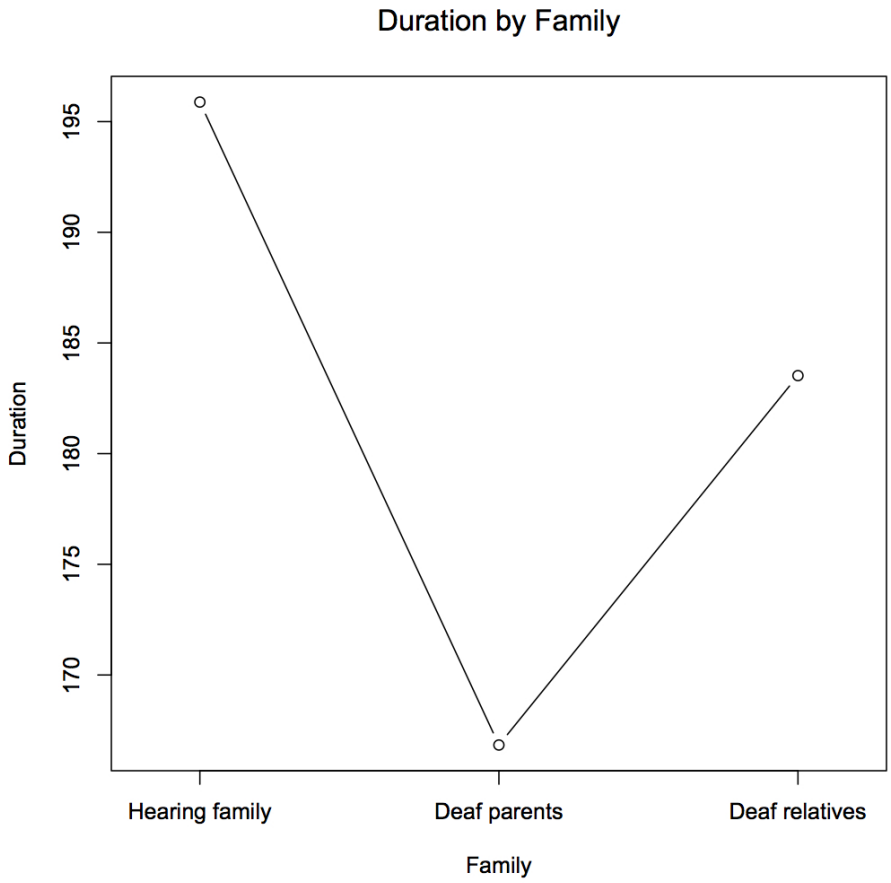


Figure 14: Duration of modifiers according to signers' family

The effect of Geographical Provenance is particularly strong (t = 4.60). As shown in Figure 15, signers from Turin, Brescia, and Salerno are faster than signers from the other cities.

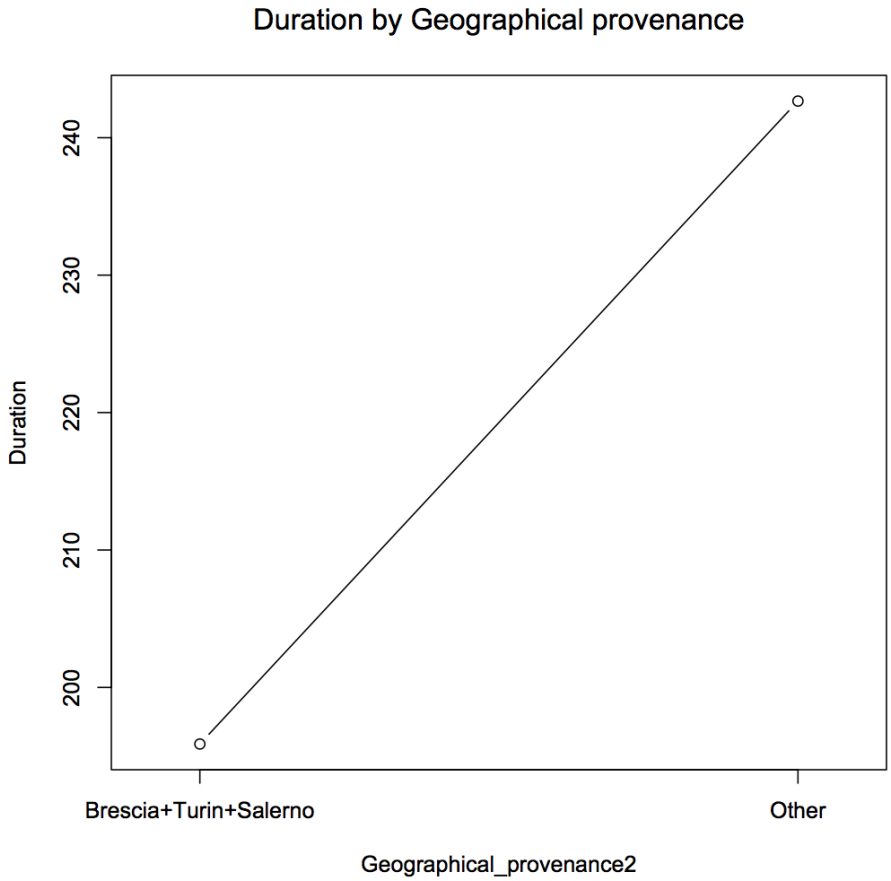


Figure 15: Duration of modifiers according to signers' geographical provenance

The interaction between Modifier type and Gender shows that despite women are overall slower signers than men (see also Figure 13), they are faster than men when pointing signs and cardinal numerals are considered (t = -2.23). This tendency is illustrated in Figure 16. As in the study on the distribution of modifiers, women display the most extreme pattern: they are slower than men in producing the signs with longer duration and faster than men in producing the signs with shorter duration, namely cardinals and pointing signs.

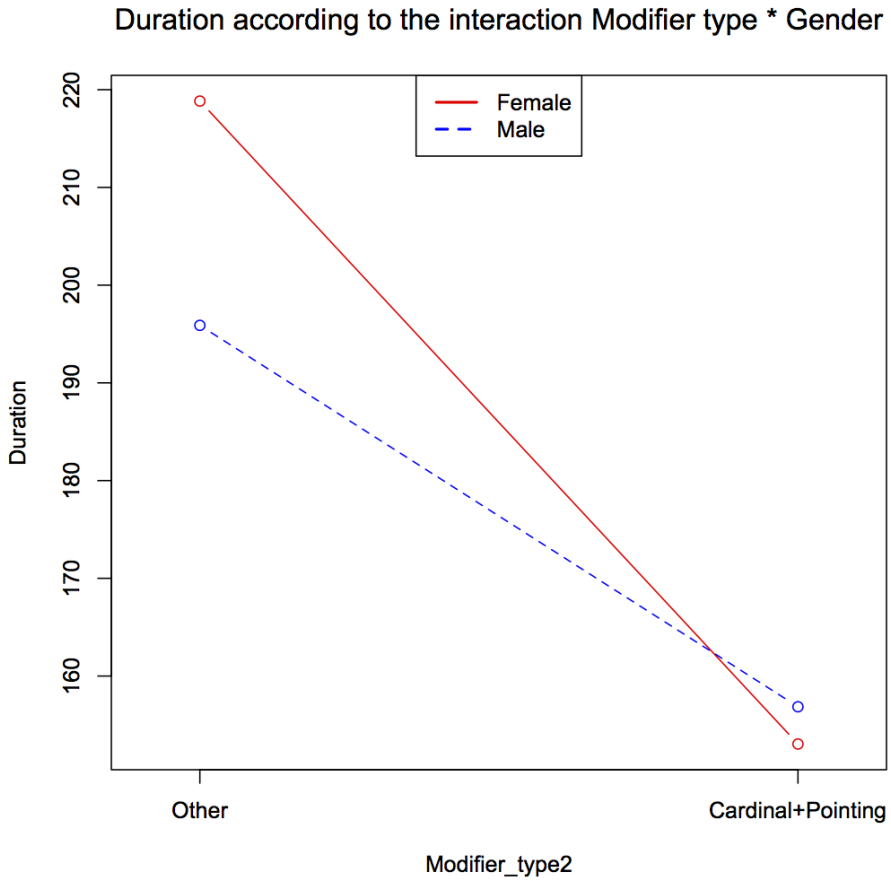


Figure 16: Duration of modifiers according to the interaction between Modifier type and Gender

I will discuss the impact of these results for the syntax of LIS nominal expressions in the next section.

4.3. Discussion

The two quantitative studies presented in this chapter show that the syntactic distribution of nominal modifiers and their duration are highly constrained both by linguistic and social factors. The statistical results are analyzed in this section.

4.3.1. The distribution of nominal modifiers

Our quantitative study investigating the position of modifiers in LIS showed that variation in their distribution is much higher than previously noted (Bertone, 2007, 2009; Branchini, 2007; Brunelli, 2011). At the macroscopic level, four options have been considered for data annotation: modifiers can be i) prenominal, ii) postnominal, iii) repeated before and after the noun, or iv) the noun can be repeated before and after the modifier. The two latter options are only marginally present in the corpus (around 6%), thus they were not considered in the statistical analysis.

The statistical analysis revealed two linguistic, two socio-linguistic effects,[16] and one interaction, as shown in (13).

(13) a. Linguistic effects: Modifier type, In-between signs

b. Socio-linguistic effects: Age, Family

c. Interaction: Modifier type * Gender

The main effect of Modifier type clearly indicates the presence of hierarchical constraints. Indeed, the factor itself is directly derived by Cinque's proposal of a fine-grained hierarchical structure of the DP layer. For statistical reasons due to the efficiency of the model, the various levels of the original configuration have been collapsed into three categories (low, middle, and high modifiers) corresponding to the edges and the central part of Cinque's hierarchy (cf. Section 2.3.1.).

Three further considerations are needed: the first one concerns how it is technically possible to derive such ordering options and which pattern of the Universal 20 can capture them; the second one concerns how to manage possessive markers which were not considered in Cinque's original proposal; and the third one concerns why all this syntactic machinery is needed.

As for the first consideration, the most frequent patterns resulting from corpus data are:

16 In the following analysis, I do not discuss the effect of Gender and Register. Gender is not discussed because it was not selected among the significant parameters during the step-up and step-down procedure. It was included in the final model because it significantly interacts with Modifier type. Register contributes to the goodness of fit of the model, but the difference between its levels (low vs. high) has not proved to be significant (see Table 3).

i) N>Dem; ii) Num>N; and iii) N>Adj. In principle, these patterns can be captured by two different orders, either Num>N>Dem>Adj[17] or Num>N>Adj>Dem. In order to establish which typological configuration reflects the most frequent pattern in LIS, a close inspection of the relative order between determiner-like pointing signs and adjectives is required. The observation of the annotated complex nominal expressions involving two or more than two modifiers reveals that the pattern N>Adj>Dem is more frequent (18 occurrences) than N>Dem>Adj (4 occurrences).[18] Three examples of N>Adj>Dem order found in the corpus are reported in (14).

(14) a. PERSON HEARING IX

'the/this hearing person'

b. NEPHEW DEAF IX

'the/this Deaf nephew'

c. ASSOCIATION NEW IX

'the/this new association'

These data suggest that the most frequent macro-typological order characterizing the nominal domain in LIS is the one reported in (15).

(15) Num>N>Adj>Dem

Deriving basic word order from partial orders is only possible on the assumption that transitivity holds across partial orders. One potential problem of applying transitivity of

17 According to Cinque (2005), the order Num>N>Dem>Adj is unattested. Cinque claims that it cannot be derived because: i) if the noun raises without pied-piping, the wrong order of Merge must be assumed (Num>Dem>Adj>N); ii) if the noun raises via pied-piping of the 'whose picture' type, the wrong order of Merge must be assumed (Num>Adj>Dem>N); iii) if the noun raises via pied-piping of the 'picture of whom' type, again the wrong order of Merge must be assumed (Dem>Adj>Num>N).

18 The four cases of N>Dem>Adj reveal a predicative usage of the adjective which is probably contained in a reduced RC. Thus, these are cases of indirect modification (for more details, see Section 2.3.1.). Notice that if they were genuine cases (i.e. not included in RCs), they would be underivable under Cinque's analysis (2005) (cf. footnote 17).

order relations to human language is that of underestimating the role of other factors that might influence order relations. Indeed, the manifestation of a word order pattern could be determined by semantic reasons (e.g. scope and sensitivity to definiteness) that operate on a relatively rigid syntactic structure. This is a general problem that emerges in rigidly applying Cinque's approach to word order. In the specific case of nominal modification in LIS, I maintain the assumption that transitivity holds across partial orders. However, this assumption is less relevant in this study since we are already in an environment where a high level of variability is attested. Therefore, we already know that there are optional triggers of word order change. Indeed, in Chapter 5, I will show that definiteness is a key factor in determining the order of cardinals with respect to the nominal head.

The pattern shown in (15) can be captured by the options provided by Cinque's (2005) theory. Simple NP movement derives the noun>modifier order for the lower projections of the hierarchy. The modifier>noun order for projections sitting in the central part of the hierarchy is derived by simple merge of these projections without any syntactic movement. The noun>modifier order of the higher part of the hierarchy is derived by pied-piping of the 'picture of whom' type. The syntactic derivation needed to obtain the pattern in (15) is illustrated in Figure 17.

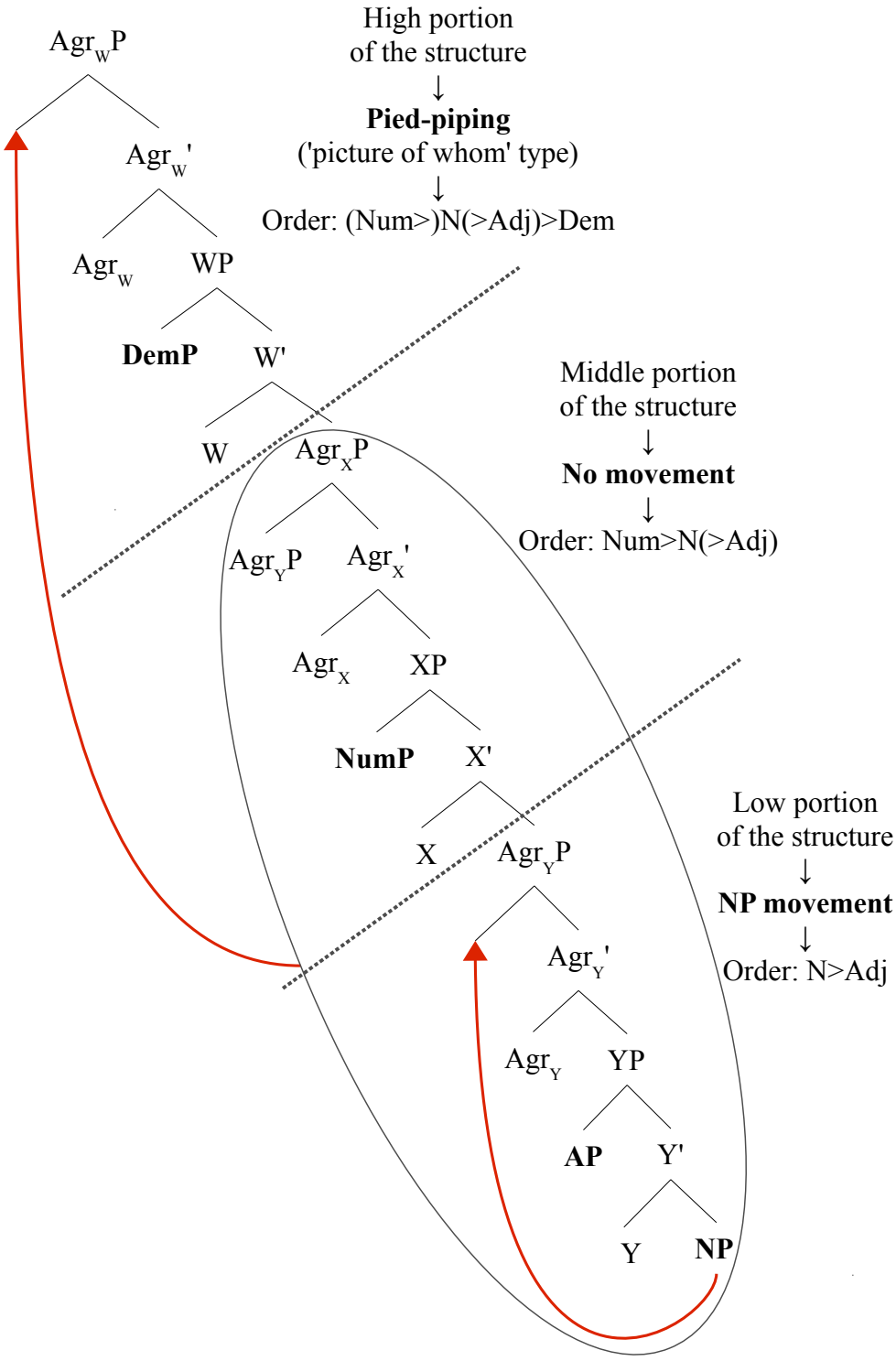


Figure 17: Syntactic derivation corresponding to Num>N>Adj>Dem order

The structure and syntactic movements illustrated in Figure 17 account for the most frequent pattern. Still, the other attested options can all be derived under Cinque's

(2005) framework and, crucially, no violations to this approach are found. All in all, it appears that LIS grammar has not rigidly set word order parameters since it allows for optional movements, even in intra-signer contexts.

The order reported in (15) is different from the one observed in Bertone (2007, 2009) and Brunelli (2011) (cf. Section 2.4.2.). I suspect that the informants involved in those studies are signers of a stricter variety of LIS in which modifiers are all postnominal.[19] Notice further that the informants involved in Bertone's (2007, 2009) and Brunelli's (2011) works would be included in the group of younger signers, namely those showing an overall preference for postnominal modifiers in the corpus (see Figure 8).

As for the second consideration, I implemented Cinque's hierarchy by adding two separate projections hosting the two lexically different types of possessives: pointing possessives and non-pointing possessives. This is in line with the adaptation of the analysis of Alexiadou et al.'s (2007) illustrated in Section 2.3.1. according to which different syntactic positions account for the distribution of different types of possessive modifiers. The reason to treat the two types of LIS possessive markers differently is not just that of introducing redundant taxonomy in an already dense DP cartography, rather it is motivated by the different syntactic distribution. Pointing possessives behave like modifiers sitting in the central part of the hierarchy, while the other possessives behave like modifiers sitting at the edge of the hierarchy. I suggest that the issue on their position in the extended projection of the NP can be addressed by placing pointing possessives at the junction between the higher and the central portion of the hierarchy and non-pointing possessives at the junction between the central part and the lower edge of the hierarchy. Possessives are thus in crucial positions in LIS, signaling the three different areas of the hierarchy: the two peripheries and the central area (see Figure 2).

As for the third consideration, I speculate that the presence of special cues may guide the acquisition process of the DP structure in children exposed to LIS. Concretely, when the grammar of a language does not codify specific linear orders in terms of categorical restrictions, as in the case of the nominal structure in LIS, learners need clear cues signaling the various levels of the hierarchy.[20] In the case of LIS, the alternation of

19 I had the chance to work with the main informant of Bertone (2007) and Brunelli (2011), who indeed uses strictly postnominal modifiers.

20 Notice that we are dealing with a language showing a high degree of variation. Word order relations provide positive evidence for a certain type of syntactic derivation which does not result from a stable

different syntactic operations (simple NP-movement, modifier merge with no movement and pied-piping) are precisely these cues. The two projections hosting possessive signs sitting at the junctions among the different areas of the hierarchy mark the cues more visibly and, in principle, might facilitate the process of language acquisition providing positive evidence for such an articulated structure.

The effect of In-between signs is a linear effect. Its direction reveals that non-adjacent modifiers are more likely to be postnominal than adjacent ones. To illustrate this general trend, consider quantifiers. Their distribution with respect to the noun when they are adjacent and when they are not is represented in the contingency table below, and real corpus examples are provided in (16).

	Quantifier > N		N > Quantifier	
Adjacent	48	35%	88	65%
Non-adjacent	3	17%	15	83%

Table 6: Distribution of quantifiers according to Adjacency

(16) a. Quantifier>N

[ALL DEAF] SIGN, IX-1 BE-USED-TO NEG

'All deaf people were signing and I was not used to it'

b. N>Quantifier

CLEAN **[BEDROOM ALL]**, EVERYWHERE

'I used to clean all bedrooms, everywhere'

c. Quantifier>X>N

BUT LEAVE **[ALL IX-1_POSS FRIEND]**

'But I had to leave all my friends'

d. N>X>Quantifier

grammaticalization process. On this line of reasoning, word order relations provide reliable cues that guide children during their acquisition process.

HI **[FRIEND DEAF ALL]**

'Hi to all (my) deaf friends'

We already know that quantifiers usually occur in postnominal position (cf. Figure 2). Results in Table 6 show that this pattern is even stronger when there is an intervening sign between the quantifier and the noun.[21]

The tendency to place stacked modifiers in LIS in postnominal rather than in prenominal position may reflect the influence of underlying processing mechanisms. Martin and Romani (1994:509) working on English show that stacked adjectives occurring in prenominal position, as in (17)a, are more challenging than stacked adjectives (inside a relative clause) appearing in postnominal position, as in (17)b.

(17) a. She bought the soft, large pillows at the department store

b. She bought the pillows that were soft and large at the department store

According to the authors, prenominal adjectives cannot be integrated in sentence structure until the noun is processed (delayed integration) and therefore imply a higher memory load. On the contrary, postnominal modifiers are not so demanding for working memory because the head noun has already been processed and they can be immediately integrated (immediate integration). This is supported by the finding that patients with short-term memory impairments perform better with immediate integration than with delayed integration. Turning to LIS, Geraci, Gozzi, Papagno, and Cecchetto (2008) show a short-term memory difference between signs and words (a fact already known for other sign languages as well, see Klima & Bellugi, 1979). Specifically, signs are particularly challenging for short-term memory. Given this limited capacity, it is at least conceivable that signers avoid those configurations leading to a short-term memory overload. Indeed, stacked nominal modifiers are preferably

21 When a nominal expression contains two modifiers and one of these is non-adjacent to the noun, in principle we expect two possibilities: either the two modifiers belong to the same Modifier type group, or the two modifiers are postnominal and belong to two different Modifier type groups (i.e. high and low modifiers). If that was the case, then Modifier type would interact with In-between signs. This interaction did not emerge as significant in the statistical analysis.

produced in postnominal position so that they can be immediately attached to the nominal head with lower processing costs.

Once looking at the internal distribution of nominal expressions containing two modifiers, an interesting pattern emerges: in 68% of the cases, both modifiers are adjacent to the noun. This configuration can be obtained only by placing one modifier before and one modifier after the head noun (modifier>noun>modifier).[22] As before, this tendency may be explained by processing considerations. Hawkins (2004) claims that there is a correlation between data variation and grammar conventions (Performance-Grammar Correspondence Hypothesis). More specifically, preferences in performance reflect the mechanisms on which the grammar of the language is built. In the spirit of Hawkins (2004), the fact that nominal expressions containing two different modifiers are frequently arranged according to the modifier>noun>modifier order can be seen as an attempt of LIS to maximize the syntactic difference among modifiers.[23]

Turning now to the significant social predictors, the effect of Age reveals a threefold partition: i) signers born before 1945 showing a slight preference for postnominal modifiers; ii) signers born between 1945 and 1965 producing modifiers preferably in prenominal position; iii) signers born after 1965 showing a significant preference for postnominal modifiers (cf. Figure 8). The fact that the generation of signers born between 1945 and 1965 behaves differently with respect to the other two groups may be due to historical circumstances. During the Second World War and right afterwards, some Institutes for Deaf students temporarily shut down. In some cases, students were sent home for precaution or because of structural damage.[24] Since most Institutes were located in urban areas they had to face the threat of bomb blasts. For instance, in 1942 Prinotti Institute in Turin was damaged by firebombs. In 1943, the State Institute for Deaf students in Milan was destroyed by a bomb attack. In 1943, Smaldone Institute in

22 Notice that the most frequent configuration emerging when two modifiers co-occur within the same nominal expression, namely modifier>noun>modifier, represents a violation of Dryer''s (2011: 6) principle of Intracategorial Harmony ("the Dem, Num, and Adj tend to all occur on the same side of the N").

23 Notice that the least frequent orders (noun>modifier>modifier and modifier>modifier>noun) bear a resemblance to clausal center embedding since in both cases a linearly closer modifier intervenes between a more external one and the modified element. Similarly to center embedding (see Cecchetto, Geraci & Zucchi, 2006), it can be argued that the configuration including stacked modifiers at the nominal level is disfavored because of its high processing costs.

24 These historical details have been provided by Luca Des Dorides (p.c., 2014).

Bari was severely damaged by the explosion of a ship in the harbor.[25] The devastating impact of the Second World War on some of the buildings devoted to the education of Deaf students probably had a disaggregating effect on the Deaf community undermining the socialization process among signers. The conjecture I propose here combines this disaggregating effect with a linguistic process of re-creolization. The linguistic implication of such historical circumstances was that the signers who grew up in the postwar period were exposed to LIS under different conditions with respect to the other generations. As already discussed, the main tendency in this population is to produce modifiers in prenominal position. It is worth noting that this configuration obtained by absence of NP movement corresponds to the unmarked pattern in Cinque's (2005) proposal. This is compatible with the idea that the language underwent a process of semi-pidginization (at least in the nominal domain). After 1965, a more stable socio-economic environment encouraged the Deaf community to reaggregate activating a process of re-creolization. During the postwar period, two different substrata corresponding to two different generations of signers converged so that a pattern of postnominal modifiers emerged (similarly to the prewar period). Whether this re-creolization process represents the intermediate stage between pidgin and creole or simply consists in the original LIS variety used by older signers re-emerged after particular historical events is left as open issue.

In the group of predictors that has been considered for the analysis, there was another factor associated with signers' age, namely Age group. This factor has been conceived as a predictor reflecting the special education policies that have been adopted in Italy in the last decades (i.e. residential schools for old signers, transition period for middle-aged signers, and mainstream education for young signers). Most importantly, the step-up and step-down procedure maintained Age in the model but excluded Age group as non-significant predictor. This reveals that signers' age has an effect on the distribution of modifiers only if the subdivision into age groups is based on historical circumstances. On the contrary, different education policies did not affect the syntactic distribution of modifiers.

The effect of Family can be interpreted in terms of different degrees of acquisition. Signers with Deaf parents are slightly more variable, probably showing more

25 http://www.storiadeisordi.it/ (20 August, 2014)

competence in managing the syntactic repertoire, as typical of native signers/speakers of a language. This would imply that different semantic nuances are associated with different order options. Further investigations in this area are needed in order to get a more precise picture. Signers from hearing families are more regular in preferring postnominal modifiers. These signers behave like advanced learners of a second language, who have mastered the basic rule of the language without being completely acquainted with the exceptions (see Littlewood, 1984: ch.3). Their preference for postnominal modifiers is therefore an overgeneralization of the general tendency. This preference is not due to the possible interference from spoken Italian at home or in the domestic environment because nominal modifiers in Italian are almost always prenominal (except RCs and some adjectives). Signers coming from families with at least one Deaf relative show an intermediate behavior, thus reinforcing the idea that the effect of Family correlates with the different degrees of language acquisition.

Finally, the interaction between Modifier type and Gender shows an interesting pattern displayed by women. They follow the most frequent option depending on the structural position of the modifier. Namely, women are more likely to produce postnominal modifiers if they are at the peripheries of the hierarchy, and they are more likely to produce prenominal modifiers if they are in the central portion of the hierarchy. This behavior of the female population (i.e. following the most frequent option) is also found in the literature of spoken languages (Labov, 2001). Labov (2001: 266) refers to the preference for women to less stigmatized and more prestigious variants. Information about prestige is not available, however, the most frequent options can be considered as the less stigmatized in a sociolinguistic environment where there is no standard form, no written form, and no official recognition of the language.

4.3.2. The duration of nominal modifiers

The second quantitative study showed that the duration of modifiers in LIS is also affected by linguistic and social factors. The statistical analysis revealed two linguistic, three socio-linguistic effects,[26] and one interaction, as shown in (18).

26 In the following analysis, I do not discuss the effect of Age and Register. These two factors contribute to the goodness of fit of the model, but the difference between their respective levels has not proved to be statistically significant (see Table 5).

(18) a. Linguistic effects: Modifier type, Position

b. Socio-linguistic effects: Family, Geographical provenance, Gender

c. Interaction: Modifier type * Gender

As for the linguistic predictors, Modifier type showed a significant effect. In particular, the short duration of two classes of modifiers, namely pointing signs and cardinals, confirms the initial expectation. Signers are particularly fast in producing pointing signs probably because in narrative discourse they often use weak indications (i.e. pointing signs with short duration) for referentiality purposes. Cardinals were articulated with short duration because the majority of cardinals (around 70%) in the corpus were digits from '0' to '10' and these are exactly the cardinals expected to be produced without path movement.

As previously expected, the effect of Position on modifier duration can be easily explained in terms of phrase-final lengthening. Postnominal modifiers are likely to appear at the end of a prosodic constituent and therefore be subject to final lengthening (for an effect of final lengthening in LIS wh-questions see Branchini et al., 2013).[27]

With respect to the social factor Family, signers with Deaf parents are faster than signers with Deaf relatives who in turn are faster than signers from hearing families. The three groups show differences in the level of fluency. Again, this is compatible with the overall picture of L1/L2 acquisition differences (Sankoff & Laberge, 1973). This interpretation supports the finding resulted from the study on the distribution of modifiers. The fact that the difference is stronger in the prosodic domain than in the syntactic one is also expected given that the phonological critical period seems to be quite shorter than the syntactic one (Guasti, 2007: 271).

Another social factor that showed a significant effect is Geographical provenance. The collapsed configuration of the variable categorized the ten cities of the corpus into two

27 Notice that it is not always the case that typically short modifiers (pointing signs and digits) occur prenominally more often than long modifiers do. Indeed, digits are usually prenominal, but the majority of pointing signs occurs in postnominal position. A possible interaction between Modifier type and Position has been controlled for, but no significant results have been found.

groups: on the one hand Turin, Brescia, and Salerno, on the other hand the other seven cities. The significant difference between these two groups is not immediately identifiable. Considering the residential schools funded in Turin, Brescia, and Salerno[28] does not help clarify why signers coming from these cities are faster in producing nominal modifiers. However, the effect of Geographical provenance can be accounted for by considering two possible sources of explanation: i) the LIS varieties in Turin, Brescia, and Salerno are characterized by some prosodic constraints inducing signers to produce modifiers with shorter movement; ii) the lexical variants used in these three cities are articulated with shorter movement and hence result in shorter duration. In order to test these two hypotheses, temporal spans relative to three distinct modifier types are extracted and compared with each other. In particular, pointing signs, cardinals (with a special divide between digits from '2' to '9' and cardinals over '9' because the latter group of cardinals may be characterized by lexical variation), and adjectives are considered. Pointing signs and digits are realized uniformly throughout the country and may provide evidence for the first hypothesis, while adjectives in some cases exhibit local variants and may provide evidence for the second hypothesis. The average duration of these types of modifiers in the two groups of cities is shown in the second and third columns of Table 7. The forth column shows the result of the t-test and indicates whether the difference between the two groups of cities is significant or not.

28 The historic residential schools in Turin are Institute for the deaf mutes (funded in 1835 and still operating in a different location), Prinotti Institute (funded in 1881 and closed in 1993), and Magarotto Institute (funded in 1954 and still operating). The residential schools in Brescia are Pavoni Institute (funded in 1842 and closed in 1980) and Mompiano School for Deaf students (funded in 1856 and now operating as mainstream school). In Salerno, the Smaldone Institute was funded in 1907 and is still operating.

Modifier type	Turin, Brescia, Salerno	Milan, Bologna, Florence, Rome, Trani, Lamezia, Ragusa	t-test
Duration of pointing signs (mean)	4.72637	4.878557	t = -2.4375 p = 0.01532
Duration of cardinals bet. '2' and '9' (mean)	4.685606	5.073186	t = -3.2646, p = 0.001822
Duration of cardinals over '9' (mean)	5.220937	5.580338	t = -2.3315, p = 0.0245
Duration of adjectives (mean)	5.392604	5.584618	t = -2.9128 p = 0.003868

Table 7: Comparing the duration of different types of modifiers in two different groups of cities

The p-values reported in Table 7 show that the duration of all the types of modifiers considered is significantly different between the two groups of cities. The fact that pointing signs and cardinals (regardless of the divide between digits from '2' to '9' and cardinals over '9') are produced with shorter duration in Turin, Brescia, and Salerno indicates that in these three cities there are prosodic constraints influencing the nominal domain. Thus, for these two types of modifiers the first hypothesis is confirmed. The fact that signers in these three cities are faster in producing adjectives may be due to local variants requiring shorter movement. Looking at the most frequent adjective in the corpus, namely DEAF (97 occurrences), the prediction is confirmed. Indeed, the canonical sign for 'deaf' requires the index finger to touch the ear and then the chin. Signers from Turin, Brescia, and Salerno reduced the internal movement of the sign touching only an intermediate point between ear and chin. This movement reduction results in an overall short duration (duration mean 5.099376) which is lower than the average duration of the same sign in the other cities (duration mean 5.551195). Therefore, lexical variants may be responsible for the difference in duration, thus confirming the second hypothesis.

A gender distinction has proved to be significant both as single factor and as factor included in the interaction with Modifier type. Although women are overall slower signers than men, they are faster in producing pointing signs and cardinal numeral. This effect can be explained in terms of lexical choices. While LIS may offer different lexical variants for the same adjective, which could be in principle subject to gender variation as future research may reveal, pointing signs and cardinals from '0' to '10'' (which represent the majority of the cardinal numerals in the corpus) are almost invariant in LIS. Once measured on equal ground, namely on pointing signs and cardinals, women are faster than men.

4.4. Conclusion

The two corpus studies presented in this chapter show that even in languages with a considerable word order variation like LIS, the basic tenets of Universal 20 in its most recently revised form are solid. This finding is of particular relevance both at the empirical and the theoretical level.

At the empirical level, a classic typological research question was addressed from a different perspective. The main aim of the study was to capture the pattern of nominal modification by looking at variation inside a single language rather than by looking at crosslinguistic variation. This novel approach has proven useful because it provided the same regularities as predicted by Universal 20. While LIS would have been classified among languages without a specific order under gross macroscopic classifications, this chapter shows that its behavior is consistent with the Universal 20, namely that of a mixed language (LIS most frequent option is Num>N>Adj>Dem). However, LIS regularities do not have the form of rigid categorical rules. Rather, they have the form of statistical tendencies. This is an expected result if we consider that LIS has no standard form, no written form, and no official status in Italy. To put it in Hawkins (2004) terms, we are witnessing a stage of language evolution in which the most common pattern is going to be internalized as a rule by the grammar (a similar process seems to be ongoing in other domains of LIS, like in question formation, see Geraci & Bayley, 2011). This work reveals that the process of stabilization is put forward by younger generations of

signers; while native and native-like signers are still preserving a larger repertoire of syntactic constructions. The findings of the syntactic study are somehow corroborated by the findings of the study on modifier duration. Women also show a distinct behavior in selecting the most frequent options. The significant linguistic factors also provide a key contribution to the empirical import of these findings. Specifically, they show the effects of a highly structured syntax of the nominal domain in shaping the variability of LIS, and they also show processing effects, which are compatible with the idea that processing factors contribute to the final shape of grammatical rules (Hawkins, 2004).

At the theoretical level, these two quantitative studies show that the generalizations emerging from Universal 20 and its formal derivation are not just adequate descriptions of crosslinguistic observations. Rather, they tap into the deep cognitive mechanisms which are behind abstract language structures and provide an adequate explanation to linguistic facts. This conclusion is further strengthened by the choice of studying a signed language. Indeed, the final results show that the abstract syntactic mechanisms underlying the derivation of Universal 20 do not depend on the specific modality of transmission of a signed or spoken language.

5. The syntax of cardinal numerals in LIS

5.1. Introduction

The two quantitative studies presented in Chapter 4 investigated a considerable amount of linguistic data coming from 162 native and native-like signers and provided a clearer picture of the significative mechanisms governing the syntax of nominal modification in LIS. Some questions about the fine-grained structure of nominal expressions are still open and qualitative investigation is needed to provide an answer to them. In particular, each modifier included in the hierarchy of the extended projection of the NP should be examined in depth so that an accurate description of the phenomena under investigation can be provided.

The main aim of this chapter is to investigate the syntax of LIS cardinal numerals. It presents the research methods and procedures that might be applied to the investigation on other modifiers populating the DP cartography. With respect to the content of the chapter, the distribution of cardinal numerals functioning as modifiers in the nominal domain will be addressed. In the literature of LIS, there has been as yet no systematic examination on their categorical status and on their distribution within the DP. So far there has been little discussion on whether the sign ONE has an ambiguous status functioning as cardinal and as indefinite determiner. Moreover, the existing accounts do not take into consideration a particular type of construction in which special cardinals are included, namely Measure Phrases. Given the limited research on this research area, this chapter is intended to stimulate the debate on the nature and distribution of cardinals.

The linguistic data shown and discussed in this chapter come from the LIS corpus and from elicitation sessions (for the methodology, see Chapter 3). The two sources of data present different pictures on the distribution of cardinals leading to a contradictory scenario: according to elicited data cardinals are postnominal modifiers, whereas corpus data show a preference for cardinals to appear prenominally. One of the main

advantages of adopting a qualitative methodology is the possibility to obtain negative evidence. Indeed, acceptability judgments have been collected in order to understand the grammaticality status of unattested productions. On the basis of the available data, a syntactic account of the distribution of cardinals in LIS is provided.

The chapter is organized as follows. Section 5.2 presents the cardinal system in LIS and provides an overview of the formational strategies employed in order to express cardinals. Section 5.3 summarizes what has been documented so far on the distribution of cardinals and presents an unexpected discrepancy between elicited and corpus data. Section 5.4 is intended to provide a better understanding of corpus data. In this section, I argue that we should not consider all tokens under the same umbrella because of the presence of two special cases, namely the sign ONE and cardinals included in Measure Phrases. Then, the argumentation moves into a more qualitative direction. Additional evidence is shown, and new findings help shed light on the issue. Section 5.5 provides a formal proposal explaining the data and accounting for word order issues involving cardinals.[1]

5.2. The cardinal system in LIS

The cardinal system in LIS involves the use of both hands and is a base-10 system. According to Zeshan, Escobedo Delgado, Dikyuva, Panda, and De Vos (2013), both deaf and hearing individuals show a preference in using 10 as numeral base. This preference is motivated by the anatomical configuration of human hands.

In LIS, cardinals from '1' to '10' are realized by extending the corresponding number of fingers, as shown by the handshapes reported in Figure 1.[2] Despite the iconic reasons that may motivate this type of realization, it is worth noting that this is not always the

1 The research work reported in this chapter has been refined thanks to useful comments and observations coming from the audiences of the conferences SignNonmanuals (Klagenfurt, April 2014) and FEAST (Venice, June 2014). Part of it has been discussed in Mantovan, Geraci, and Cardinaletti (2014).

2 The explanatory drawings included in this section are taken from Radutzky (2001). They refer to the most widespread variants in Italy. In some areas of the country different signs may be used (especially signs for tens and numbers from '11' to '19'). This variation concerns the lexicon of the language and hence is not relevant for the purposes of the discussion. The data recorded during the elicitation sessions come from two signers from Southern Italy and one signer from Northern Italy (for more details see Section 3.2.2.).

case in the world's sign languages. For example, ASL signers use one hand and ten different handshapes to express cardinals from '1' to '10'.

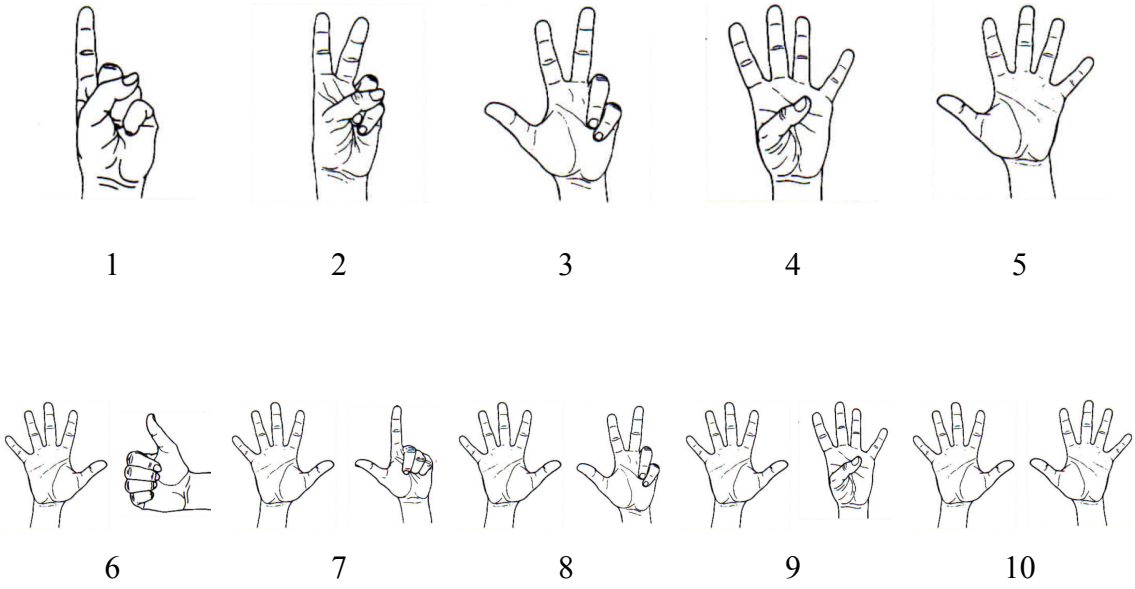


Figure 1: LIS cardinals from '1' to '10'

Some LIS signers produce the sign ONE by extending the thumb (S handshape) instead of the index finger[3] and the sign TWO by extending thumb and index finger (L handshape) instead of index and middle finger. These two alternatives (S and L handshapes) are obligatorily articulated by the dominant hand in cardinals SIX and SEVEN, respectively. Notice also that from '6' to '10', the non-dominant hand always shows the 5 handshape.[4]

Zeshan et al. (2013) present the formational strategies that are commonly adopted to express cardinal numerals in the world's sign languages. Their classification includes the following strategies: multiplicative, additive, subtractive, digital, and spatial modification. I apply the classification of Zeshan et al. (2013) to the cardinal system in LIS. In order to express quantities above '10', LIS adopts three formational strategies: i)

3 For more on the phonological characteristics of the sign ONE, see Section 5.4.1.

4 With respect to cardinals from '6' to '10', sign languages show interesting variation. As previously discussed, in LIS both hands are used. French Sign Language (LSF) is similar to LIS with the exception of cardinal '10' which is not produced by extending ten fingers. It is a two-handed sign involving the F handshape with the loop of the F representing cardinal '0'. ASL signers use one hand and differentiate cardinals from '1' to '5' and from '6' to '10' by selecting different fingers. Cardinals from '6' to '10' in Argentine Sign Language (LSA) are one-handed signs and touch the body.

multiplicative; ii) digital; iii) additive. In the remainder of this section, each strategy is discussed in turn, and some illustrative examples from LIS are provided.

The multiplicative strategy is a case of simultaneous morphology. It combines digits handshapes with particular hand or wrist movements. In cardinals from '11' to '15', for example, handshapes from '1' to '5' realized with contralateral palm orientation are associated with a repeated downward twist of the wrist on the lateral plane. Figure 2, which represents the sign THIRTEEN, shows this formational mechanism.

Figure 2: LIS sign for '13' (multiplicative strategy)

Cardinals from '16' to '19' are similarly produced except that the palm orientation is toward the signer and the repeated downward twist of the wrist is on the vertical plane.[5]

The multiplicative strategy is used also to express thousands. In these cardinals, palm orientation is toward the interlocutor while handshapes from '1' to '10' move downward on the lateral plane. Figure 3 shows how THREE-THOUSAND is realized.

5 Another way to express cardinals from '16' to '19' involves the additive strategy. These variants are produced first by signing TEN with outward palm orientation, then by moving both hands in opposite directions and showing how many units are involved.

Figure 3: LIS sign for '3000' (multiplicative strategy)

Tens are produced with the same palm orientation. Handshapes from '2' to '9' are combined with finger bending. In some cases (for example, '40'), there are two possibilities: bending all fingers, or bending only the index finger.[6]

Another possible option is to have a complex movement by combining two different types of movements. This is the case when we look at hundreds which are produced by directing the relevant handshape toward the interlocutor. The complex movement involves an outward shift toward the ipsilateral side of the signing space and simultaneous finger bending. Figure 4 shows the sign THREE-HUNDRED.

6 Finger bending in tens is probably motivated by the fact that index finger toward thumb is reminiscent of the shape of ZERO.

Figure 4: LIS sign for '300' (multiplicative strategy)

The digital strategy is used for cardinals from '21' to '99' (with the exclusion of tens). It consists in producing individual digits in the order in which they appear in writing. For example, the sign for '34' is realized by juxtaposing THREE and FOUR, as illustrated in Figure 5. The transitional movement from one sign to the other may imply a slight outward shift toward the ipsilateral side of the signing space.

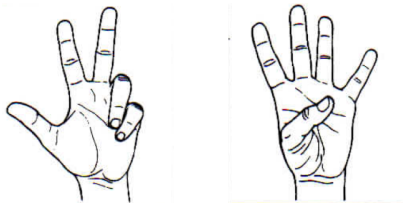


Figure 5: LIS sign for '34' realized by THREE and FOUR (digital strategy)

Finally, the additive strategy is used for numbers above '100'. It combines two or more signs, the sum of which equals to the quantity expressed by the cardinal. To illustrate this point, I consider the sign for '3409'. It is formed by three signs ordered from the highest to the lowest: THREE-THOUSAND, followed by FOUR-HUNDRED, followed by NINE.

According to Zeshan et al. (2013), sign languages may also employ the subtractive

strategy and spatial modification. The subtractive strategy consists in indicating a certain quantity and then the quantity that needs to be subtracted to obtain the target cardinal. The subtractive strategy is illustrated in the example in Figure 6 taken from Mardin Sign Language (Zeshan et al., 2013: 378).

Figure 6: Substractive strategy in Mardin Sign Language

Spatial modification is not a very common strategy. It consists in enlarging the size of the cardinal and it is used to express augmentative forms (in Figure 7 below this is indicated by the superscript '^{aug}'). In Alipur Sign Language, the sign HUNDRED is realized with the palms facing each other at a relatively close distance. When the distance between the palms is larger, the sign means '1000'. When the distance is even larger, the meaning is '100.000'. The three examples are illustrated in Figure 7 (Zeshan et al., 2013: 381).

Figure 7: Spatial modification in Alipur Sign Language ('100', '1000', '100.000')

LIS displays neither subtractive strategy nor spatial modification.

Like many other sign languages, LIS allows for numeral incorporation. This is a morphological process that incorporates the cardinal handshape into a sign. This process involves cardinals from '1' to '5' and in some cases even from '1' to '10'. The signs that can be combined with cardinals are plural personal pronouns (e.g. IX-2, IX-3pl), nouns referring to time (e.g. MONTH, YEAR, HOUR)[7] and a few other signs (e.g. FLOOR, MARK).

5.3. The cardinals puzzle

This section summarizes what has been documented in the relatively recent literature with respect to the distribution of cardinals within the nominal domain. To date, there is no extensive study on this issue. However, the three works on broader research topics in LIS presented in Section 2.4.2. allude to the position of cardinals. On the basis of elicited data, Bertone (2007), Branchini (2007), and Brunelli (2011) claim that cardinals occur in postnominal position. This is illustrated in the examples in (1) (Bertone, 2007: 84), in (2) (Branchini, 2007: 50), and in (3) (Brunelli, 2011: 61). Both noun and cardinal are boldfaced and the relevant nominal expression is indicated by square brackets.

7 This type of numeral-incorporated signs will be taken into consideration and discussed from a syntactic perspective in Section 5.5.2.

(1) [**BOOK** NEW **TWO** IX_DEM] MINE

'These two new books are mine'

(2) [**CHILDREN THREE** IX_DEM] ICE-CREAM LIKE

'These three children like ice cream'

(3) IX-1 [**BOOK THREE**] EXIST

'I have three books'

A slightly different perspective is offered by Cecchetto, Geraci, and Zucchi (2009). They claim that the order N>Card is the prevalent one, as exemplified in (4)a (2009: 84). However, (4)b (2009: 84) shows that also Card>N is possible, even if such order is employed to a lesser extent. This word order variation is neither further examined nor motivated by the three authors.

(4) a. STUDENT THREE ARRIVE DONE

b. THREE STUDENT ARRIVE DONE

'Three students arrived'

With respect to the distribution of cardinals, the quantitative analysis on the position of modifiers presented in Section 4.2.1. reveals a surprising result. Indeed, corpus data show that in spontaneous narratives the majority of cardinals appears before the noun, contrary to what has been found in previous literature (Bertone, 2007; Branchini, 2007; Brunelli, 2011). The distribution of cardinals obtained in this study is reproduced in Table 1. The case of Card>N emerges as the preponderant order option (79% of occurrences).[8]

8 The repetition cases, Card>N>Card and N>Card>N, are not reported in the table because they represent only a marginal portion of the whole amount of cardinal numerals found in corpus data (4% and 2%, respectively).

Word order	n	%
Card>N	278/353	79%
N>Card	75/353	21%

Table 1: *Distribution of cardinal numerals in corpus data*

This study on the position of cardinals with respect to the noun has been inspired by an unexpected finding: two different types of data show two opposite patterns. On the one hand, according to the existing literature based on elicited data, the most common or even the only possible order is N>Card. On the other hand, corpus data give evidence for the prevalence of the reversed order, namely Card>N.

Given this starting point, the research question that should be addressed is the following: why do we observe such a discrepancy? In what respect is Card>N different from N>Card (and vice versa)? The nature of cardinal numerals is worth being investigated in order to find out what drives their distribution. The two orders may involve either two distinct syntactic positions or a unique one. They may involve some semantic nuances. For example, the canonical order in Russian is Card>N, as in (5)a (Yadroff & Billings, 1998: 319). The reversed order (N>Card) is possible but it involves an approximative reading, as in the colloquial form in (5)b and in the standard form in (5)c (Yadroff & Billings, 1998: 319).

(5) a. On otdal svoj laptop [za pjat’ knig]
he traded own laptop for five books-GEN.pl
'He traded his laptop (computer) for five books' (no inversion)

b. On otdal svoj laptop [za knig pjat’]
he traded own laptop for books-GEN.pl five
'He traded his laptop (computer) for approximately five books' (colloquial)

c. On otdal svoj laptop [knig za pjat’]
he traded own laptop for books-GEN.pl for five

'He traded his laptop (computer) for approximately five books' (standard)

The first quantitative study in Section 4.2.1. presented the sociolinguistic facts that are relevant to the distribution of cardinals with respect to the head noun. In the next sections, I will use qualitative data in order to better understand this distribution. Once the data will be cleared of two potential confounders, a crucial effect of definiteness will be used to explain the difference between prenominal and postnominal cardinal.

5.4. Understanding the puzzle

The rationale used to tackle the distribution of cardinals in LIS starts from corpus data and then considers also elicited data. First, two subsets of cardinals are discussed, namely the sign ONE and cardinals within Measure Phrases (e.g. THREE WEEK). The former represents a case of ambiguity (i.e. the sign for ONE can be used as cardinal and as indefinite determiner), whereas the latter is characterized by an exceptional distribution. For these reasons, I argue that these two subsets should not be included in the refined analysis presented here. If they were, they could contaminate the rest of the data and lead to wrong or at least spurious generalizations. Then, on the basis of new grammatical data and negative evidence, the study is headed toward a more fine-grained analysis of cardinals.

5.4.1. First confounder: the case of ONE

The sign ONE appears as something different from the other items, as is also clear according to frequency data. The six cardinal numerals most frequently produced in the corpus are listed and sorted by frequency in Table 2. Data are presented as numerical values in column 2 and as percentages in column 3.

Gloss	n	%
ONE	101/353	29%
TWO	49/353	14%
THREE	32/353	9%
FOUR	13/353	4%
FIVE	16/353	5%
SIX	9/353	3%

Table 2: Most frequent cardinals in corpus data

Table 2 shows that the sign ONE is the most frequent item with 101 occurrences. Looking at these frequency data, ONE is likely to involve a more complex picture. To better understand what influences the use of ONE and to check whether or not it has to be considered as a special case, both linguistic and sociolinguistic data should be looked at. According to Table 2, ONE outnumbers TWO, THREE, and other cardinals. Is this fact homogeneous within the corpus?

The interesting result is that signers' age plays a crucial role in the investigation of the peculiarities of the sign ONE. The relevant effect is Age group, namely the nominal variable dividing participants into three groups (young, middle aged, old signers) with the purpose of taking into consideration the impact of different education policies.[9] In particular, the occurrences of ONE and TWO (i.e. the second most frequent cardinal) have been compared with respect to the three age groups by applying the chi-square test.[10] The difference has been proved to be significant thus revealing that old and middle-aged signers do not produce ONE and TWO like young signers. Table 3, Table 4, and Table 5 show frequency data related to old, middle-aged, and young signers, respectively.

9 A possible correlation between the continuous variable Age and the production of the sign ONE has also been tested. From a statistical perspective, a significant effect of the predictor did not emerge. Even considering the age grouping introduced in Section 4.2.1. (signers born before 1945, between 1945 and 1965, and after 1965) did not lead to significant results.

10 Pearson's Chi-squared test with Yates' continuity correction
Old and middle-aged signers vs. young signers
X-squared = 4.1846, df = 1, p-value = 0.04079

Gloss	n	%
ONE	45/144	31%
TWO	15/144	10%
THREE	11/144	8%
FOUR	7/144	5%
FIVE	7/144	5%
SIX	3/144	2%

Table 3: Frequency data of cardinals produced by old signers (tot: 144)

Gloss	n	%
ONE	41/131	31%
TWO	19/131	15%
THREE	12/131	9%
FOUR	3/131	2%
FIVE	6/131	5%
SIX	4/131	3%

Table 4: Frequency data of cardinals produced by middle-aged signers (tot: 131)

Gloss	n	%
ONE	15/78	19%
TWO	15/78	19%
THREE	9/78	12%
FOUR	3/78	4%
FIVE	3/78	4%
SIX	2/78	3%

Table 5: Frequency data of cardinals produced by young signers (tot: 78)

Information on the three different age groups may reveal that some diachronic change has happened or is happening. Looking at frequency data reported on the above three tables, we can see that cardinals from '2' to '6' are produced in a rather homogeneous way. The three age groups do not differ considerably in this respect. The first striking result is that old and middle-aged signers produce the sign ONE more frequently than young signers (31% of the total number of cardinals, as opposed to 19%). If the use of ONE is compared to the use of TWO, age group differences are quite clear. In the old population, the discrepancy between the frequency of the sign ONE and the frequency of the sign TWO is 21% (see Table 3). In the middle-aged population, it is 16% (see Table 4). Interestingly, in the young population the production of ONE and the production of TWO show exactly the same frequency (see Table 5). Why these three groups behave so differently is an issue that needs to be addressed.

On the one hand, old and middle-aged signers probably treat ONE not only as a simple cardinal. Participants belonging to these two age groups are more likely to use the sign ONE both as cardinal and as indefinite article. On the other hand, young signers are more likely to use ONE as pure cardinal. Notice that this diachronic change shows that LIS is departing from Italian in this respect.

Looking at corpus data in detail, ONE emerges as an item conveying two different linguistic functions: it can be used either as cardinal numeral or as indefinite determiner. The two functions associated with ONE are shown in the following two examples. Both of them are produced by old signers. In (6), an old signer from Rome explains how to

move the king in a chess game. This example shows how ONE can be used as cardinal.

(6) KING ONE STEP ONE CL […] ONE STEP

'The king moves one step at a time, one step'

In (7), an old signer from Florence introduces herself. This example shows how ONE can be used as indefinite determiner.

(7) IX-1 BE-BORN ONE TOWN IX_i C??? PROVINCE PERUGIA IX_i

'I was born in a town, C???, in the province of Perugia'

In both cases, the context is crucial in order to detect the status of ONE and understand its meaning. However, in the corpus, there are a few ambiguous cases in which it is not clear whether ONE functions as a cardinal or as a determiner. The example in (8), produced by a young signer from Rome, shows a case of ambiguous ONE.

(8) THERE-IS ONE DEAF THERE-IS, MEET

'I met a/one deaf person'

Once counting the ambiguous cases of ONE and the unambiguous ones (i.e. determiner vs. cardinal ONE) in corpus data, the distinct usages of this sign can be evaluated with respect to the three age groups. Table 6 shows these data.

	Old signers		Middle-aged signers		Young signers	
	n	%	n	%	n	%
Determiner ONE	12	27%	10	24%	1	6%
Cardinal ONE	16	35%	17	42%	10	67%
Ambiguous status	17	38%	14	34%	4	27%
Total	45	100%	41	100%	15	100%

Table 6: Usages of ONE in the three different age groups

The data reported in Table 6 provide quantitative support to the hypothesis according to which some diachronic change is happening. ONE is evolving from being ambiguous between determiner and cardinal to the unambiguous status of cardinal.

If ONE as indefinite determiner is probably fading away, we expect young signers to express indefinite nominal expressions through other linguistic means. Some of them may still use ONE, and this can be due to the influence of older signers. The fact that one case of Determiner ONE was found in the young population (see Table 6) confirms that. In this particular case, indefiniteness is conveyed by producing ONE with tremoring motion (i.e. slight trembling movement of the forearm and hand). The higher the degree of indefiniteness, the more visible the tremoring motion. This sign is used in ASL with the same linguistic function, as reported by MacLaughlin (1997). This special use of ONE in LIS has been noted first by Bertone (2007) and is exemplified in (9) (Bertone, 2007: 146). According to Bertone, determiner ONE with tremoring motion is never articulated in a particular point in space, but rather in an unmarked location. It is usually associated with a facial expression denoting uncertainty which consists in pulling the mouth ends down. This can be combined with a shrug, as similarly reported for LSC (see Barberà, 2012: 237)

ind

(9) BOOK ONE $_2$GIVE$_1$

'Give me a book'

Determiner ONE with tremoring motion has been found in a few cases in the corpus. The example in (10) shows the one produced by the young signer. The non-manual component is the same described by Bertone.

ind

(10) IX-1 ASK ONE FRIEND PLEASE $_2$HELP$_1$

'I'd ask a friend to help me'

However, contrary to Bertone's data, in (10) ONE precedes the noun. The correlation between the prenominal position and indefiniteness will be further discussed in Section 5.4.3.

Another way to express indefiniteness is to produce a bare NP and to emphasize the relevant NMMs of uncertainty. This option is shown in (11) (young signer from Trani).

ind

(11) PERSON TRUST DIFFICULT VERY

'It's very difficult to trust someone'

As observed before, the sign ONE sometimes has a quantificational reading, but in some other cases it can be used as indefinite determiner. To understand whether ONE is completely ambiguous or not, some distributional and phonological aspects should be looked into.

Bertone (2007) claims that determiner ONE differs from cardinal ONE because it is associated with tremoring motion and with particular NMMs expressing uncertainty, namely spread lips with the corners down in a reversed U-shape. What has been found in the corpus is that ONE can be used to express indefiniteness regardless of the

presence of tremoring motion, as in the example in (7). Determiner ONE is usually produced in a steady position or with a slight movement upward, probably in correlation with upward head tilt and shrug of the shoulders. Also, cardinal ONE can be produced with very little to no movement, especially when it is in prenominal position. Alternatively, the index finger can be slightly shifted in a forward position of the signing space.

As for the non-manual components, the particular reversed U-shape mouth gesture observed by Bertone is often found both with cardinal ONE and determiner ONE. This may reflect the mouthing derived by the Italian equivalent 'uno'.[11] However, a difference in terms of NMMs has always been detected. As illustrated in Figure 8, cardinal ONE is not associated with particular NMMs. On the contrary, determiner ONE often co-occurs with backward-tilted head and slightly raised eyebrows (see Figure 9). These NMMs are used to convey the notion of indefiniteness (this aspect will be further discussed in Section 5.4.3).

Figure 8: NMMs associated with cardinal ONE [Old signer from Turin]

Figure 9: NMMs associated with determiner ONE [Middle-aged signer from Turin]

Looking at palm orientation, cardinal ONE is usually articulated so that the palm faces the signer (cf. Figure 8). This differs from determiner ONE, which can be produced with

11 Notice that also Italian 'uno' is ambiguous between indefinite determiner and cardinal numeral (not to mention the additional pronominal use in, for example, 'uno mi ha chiesto il nome', transl. 'one asked me my name'). This last use is not possible in LIS, though.

the palm facing the signer or with contralateral orientation (cf. Figure 9).[12] This pattern is similar to the one discussed for HKSL (for further details see Tang & Sze, 2002). With respect to the tendency concerning palm orientation in LIS, apparent counterexamples in which the palm of cardinal ONE does not face the signer can be explained in terms of anticipatory assimilation. The example in (12) is taken from the production of a middle-aged signer from Rome. The sign WEEK, which has outward palm orientation, induces the prenominal cardinal ONE to be articulated with contralateral orientation for convenience's sake.

(12) REFECTORY EAT FINISHED, REFECTORY ARRANGE TURN, ONE WEEK IX-1, THEN WEEK IX-3

'After we finished eating at the refectory, we took turns arranging things, one week it was up to me, then it was up to someone else'

Corpus data reveal another interesting phonological fact: ONE can be realized either with the extended thumb (54 occurrences), or with the extended index finger (47 occurrences).[13] The two handshapes are shown in Figure 10 and Figure 11, respectively.

12 The patterns concerning the palm orientation of determiner ONE and cardinal ONE have been identified according to corpus data. Negative evidence is needed in order to confirm these data.

13 As shown in Figure 1, Radutzky (2001) only reports the sign ONE realized with the extended index finger.

Figure 10: ONE with extended thumb [Middle-aged signer from Turin]

Figure 11: ONE with extended index finger [Middle-aged signer from Salerno]

Nevertheless, this difference is not relevant for the distinction between cardinal ONE and determiner ONE. It seems to be rather an instance of diatopic variation since the index finger is preferred by Southern signers, whereas the thumb handshape is more used in the North. Signers from Central Italy tend to use both hand configurations. These patterns are shown in Table 7.

Signers' geographical provenance	**Extended thumb**		**Extended index finger**	
	n	**%**	**n**	**%**
Northern Italy (Turin, Milan, Brescia, Bologna)	44/51	86%	7/51	14%
Central Italy (Florence, Rome)	9/24	37,5%	15/24	62,5%
Southern Italy (Trani, Salerno, Lamezia, Ragusa)	3/26	12%	23/26	88%

Table 7: Variation of the handshape of ONE according to signers' geographical provenance

To shed light on the status of ONE, it is also worth looking at its distribution with respect to the head noun.

Word order	n	%
ONE>N	93/101	92%
N>ONE	8/101	8%

Table 8: Distribution of ONE in corpus data

Table 8 shows that in most occurrences ONE appears before the noun (92%). When it is found in this position, it is often ambiguous between the determiner and the cardinal status. In those few cases in which ONE follows the noun (8%), it is associated with the quantificational reading, it does not combine with the typically indefinite NMMs, and it is produced with palm orientation toward the signer.

To sum up, we saw that the sign ONE is subject to both diatopic and diachronic variation. Its status is ambiguous because it can be used as either a cardinal or an indefinite determiner. In some cases, context may help identify the real function of ONE. In other cases, some distributional and phonological aspects may provide some clue to disambiguate the sentence. However, it is not always easy to draw a neat distinction between the two functions.

Because of its twofold nature and a number of ambiguous cases, the sign ONE should not be considered in the study on the distribution of cardinals. Raw data and percentages referring to the distribution of cardinals without the sign ONE are reported in the last two columns of Table 9.

Word order	General distribution		Distribution without ONE	
Card>N	278/353	79%	184/252	73%
N>Card	75/353	21%	68/252	27%

Table 9: Distribution of cardinals with and without ONE

After excluding the occurrences of ONE, which amount to about one third of all tokens,

there is no substantial change in the distribution. The difference between the two most common orders gets reduced from 58% to 46%. Nevertheless, the prevalence of Card>N is maintained.

5.4.2. Second confounder: Measure Phrases

Another special case that requires additional analysis is the distribution of cardinals included in Measure Phrases (hereafter MPs). This construction represents an important part of the cardinal corpus (140 out of 353 cardinals). The nouns that occur in MPs are measure nouns referring to time, capacity, weight, length, temperature, and currency.

The distribution of cardinals with respect to measure nouns is illustrated in Table 10. For each case, an example is provided in the last column of the table.

Measure noun	Card>N	Example
YEAR	77/77	TWO YEAR
MONTH	13/13	ELEVEN MONTH
WEEK	10/10	THREE WEEK
DAY	17/17	FIVE DAY
EVENING	1/1	ONE EVENING
HOUR	4/4	SEVEN HOUR
MINUTE	4/4	FIVE MINUTE
SECOND	1/1	FIVE SECOND
TIME/INSTANCE	2/2	THREE TIME
KILOGRAMME	1/1	SEVENTY KILOGRAMME
LITRE	2/2	ONE LITRE
BARREL	1/1	SEVEN BARREL
KILOMETRE	2/2	THIRTY KILOMETRE
METRE	2/2	FOUR-HUNDRED METRE
CREDIT	1/1	SIXTY CREDIT
LIRA	1/1	FIFTEEN-THOUSAND LIRA
DEGREE	1/1	FORTY DEGREE
Total	140/140	

Table 10: Distribution of cardinals included in MPs

According to the results reported in the table, cardinals consistently precede measure nouns showing a categorical distribution.

However, the fact that postnominal cardinals have not been found in the MPs annotated in the corpus does not necessarily mean that, in these constructions, the order N>Card is ungrammatical. As a matter of fact, unattested does not mean ungrammatical. From a methodological point of view, it is worth noting that corpus data do not provide negative

evidence and do not always show all possible grammatical options. With such kind of data, it is not possible to ask specific questions on what has been signed. This is the reason why, once the analysis on corpus data is done, it is useful to involve some native signers. During fieldwork, I asked for some acceptability judgments in order to test whether a measure noun can be followed or not by a cardinal. An example is reported in (13).

(13) a. IX-1 REPEAT++ TWO-HUNDRED-THOUSAND TIME

'I repeated it two-hundred-thousand times'

b. * IX-1 REPEAT++ TIME TWO-HUNDRED-THOUSAND

On the basis of acceptability judgments, the only option that emerges is Card>N. The reversed order is judged as ungrammatical.

The fact that cardinals contained in MPs show a categorical distribution (i.e. they can occur only before the noun, while the reversed order is not possible) indicates that this subset of cardinals behaves differently with respect to the other cardinals. Therefore, I claim that they should be treated as a special category and should be removed from the counting. The case of cardinals included in MPs represents a puzzle within the puzzle and requires an independent explanation. This is addressed in Section 5.5.2.

While removing the tokens corresponding to cardinals within MPs, it is important to consider that some instances of cardinal ONE may occur in MPs. Obviously, these overlapping cases have already been removed.

Table 11 shows the distribution of cardinals with respect to the head noun in the three stages: i) the distribution of cardinals with both confounders; ii) the distribution without ONE; iii) the distribution without ONE and without MPs.

Word order	General distribution		Distribution w/o ONE		Distribution w/o ONE and w/o MPs	
Card>N	278/353	79%	184/252	73%	68/136	50%
N>Card	75/353	21%	68/252	27%	68/136	50%

Table 11: Distribution of cardinals without the two confounders

Once removed the two special cases, the distribution of the cardinals coming from corpus data look like as in the last column of Table 11: the percentage of postnominal cardinals is identical to that of prenominal ones showing a perfectly balanced situation (50% prenominal cardinals and 50% postnominal cardinals). The picture that emerges is even more intricate with respect to the initial one, since it shows an apparently uncontrolled variability and is still in contrast to the pattern discussed in previous literature.

5.4.3. New insights from new data

The data variability described in the previous section is an issue that needs to be addressed. In particular, what is important to understand is whether this is an uncontrolled phenomenon or rather a phenomenon constrained by specific factors. In order to understand in what respect Card>N is different from N>Card and vice versa, some additional data have been collected.

First of all, a picture-based narration task has been defined. The material used as stimulus is a wordless comic strip illustrated in Ohser (2000). The story is represented in Figure 12. Ohser's illustrations are generally self-explanatory and do not give rise to interlinguistic influences since they do not contain any written text. Then, the three LIS native signers presented in Section 3.2.2. were asked to generate a narrative based on this story.

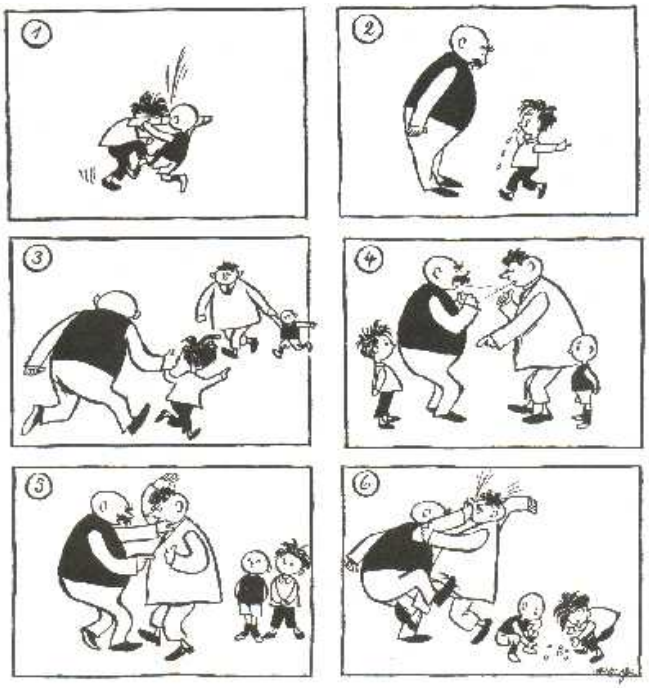


Figure 12: Extract from "Vater und Sohn" (Ohser, 2000)

The story illustrated in Figure 12 is interesting because it triggers the production of a cardinal in two different contexts. In the first panel, two children are represented for the very first time. Being first-mentioned referents, they are expected to be introduced in the discourse by an indefinite nominal expression (e.g. 'two children were fighting on the street'). On the contrary, the two children represented in the fifth panel are pre-established referents, therefore they are expected to be referred to by using a definite nominal expression (e.g. 'the two children were looking at their fathers').

Data annotation has been conducted by using the annotation software ELAN (Johnston & Crasborn, 2006). Manual and non-manual features have been carefully annotated on separate tiers. The coding scheme associated to the NMMs relevant for this study is provided in (14). The duration of NMMs has been measured as the time interval intervening between start and end points.

(14) a. NM-Head: left, right, raised, down, forward, back

b. NM-Eyebrows: lowered, raised

c. NM-Body: left, right, down, forward, back

d. NM-Eyes: blink, squint, close, wide, trackhands eye-gaze

The data collected during the narration task reveal that the position of cardinals may be influenced by information structure. When the children are first mentioned (see first panel in Figure 12), both orders (Card>N and N>Card) have been produced, while in further mentioning (see fifth panel in Figure 12) only the N>Card order has been found. This seems to suggest that new-discourse information (i.e. first-mentioned referents) is compatible with both orders, whereas old-discourse information (i.e. already-mentioned referents) is compatible with the postnominal cardinal only. Evidence coming from acceptability judgments is needed in order to confirm this pattern.

The informants' assessment of their own productions confirms the pattern. When explicitly asked about the order possibilities in the two distinct contexts, they answered that the new-information situation allows for both sign order options, as exemplified in (15). On the contrary, in the old-discourse context only the N>Card order is possible, as illustrated in (16).

(15) Indefinite nominal expressions (new-information context)

a. TWO CHILD

b. CHILD TWO

'Two children'

(16) Definite nominal expressions (old-information context)

a. * TWO CHILD

b. CHILD TWO

'The two children'

It is worth noting that the relative order of the cardinal with respect to the noun is not enough to distinguish the two discourse functions, as the sequentially identical data in

(15)b and (16)b demonstrate.

Some further hints come from NMMs. If the signer is dealing with a new referent, the prenominal or postnominal cardinal is usually accompanied by backward-tilted head and raised eyebrows (see Figure 13). If the referent has already been mentioned in the discourse, then the postnominal cardinal is compatible with squinted eyes, lowered eyebrows and chin down (see Figure 14).

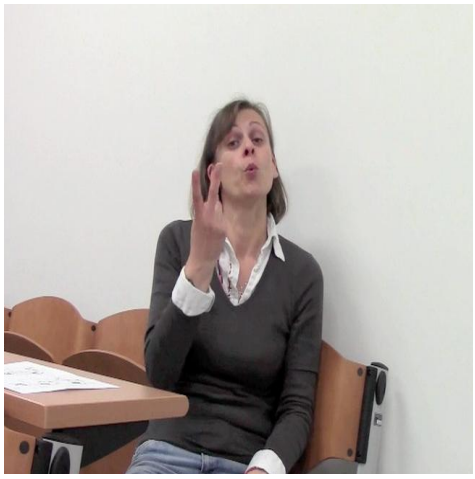


Figure 13: TWO as new-discourse information

Figure 14: TWO as old-discourse information

Another difference that has been noted between the two patterns is that only already-mentioned referents can be followed by a classifier sign. The ordering characterizing this construction is given in (17).

(17) N>Card>CL (only in definite nominal expressions)

Two classifiers that can be used in this type of construction are shown in Figure 15. Both of them are whole entity classifiers.

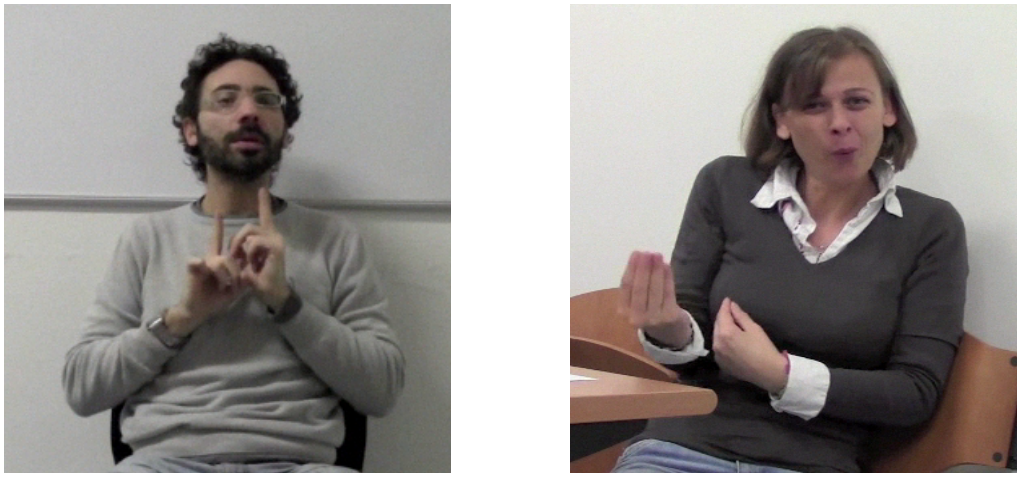


Figure 15: Classifier signs following cardinals in definite nominal expressions

The classifier locates the referent in a specific point in space, so it is not unexpected that this sign occurs only in a definite environment.[14] Moreover, notice that classifiers seem to require a discourse antecedent and as such are independently expected to refer to discourse-old information. Therefore, the presence of these classifiers provides independent evidence of the definiteness of the relevant nominals.

In the light of what has been found in the qualitative investigation involving the three LIS informants, let us now turn back to corpus data and check whether the pattern is consistent or not.

This assessment should address the following issues: i) which pattern do signers display when they deal with first-mentioned referents and with already-mentioned referents in spontaneous narration?; ii) according to what has been found in elicited data, prenominal numerals cannot be contained in definite nominal expressions: is this prediction confirmed by corpus data?

With respect to the first issue, I show the production of two signers who first introduce a new referent and then refer back to the same referent later on in the discourse. The first pair of sentences is produced by a young signer from Florence and is shown in (18).

(18) a. HEARING IX-3 ACQUIRE **[TWO LANGUAGE]**

14 According to Bertone (2006: 169), the location of the classifier plays an important role in determining the nature of the related head noun in terms of definiteness in LIS. In other words, the classifier encodes the referentiality of the noun it is associated with.

'(My) hearing child is acquiring two languages'

b. GIVE$_3$ **[LANGUAGE TWO]** IMPORTANT

'It is important to give him/her the two languages'

This signer talks about her life as deaf mother of a hearing child. With respect to the language acquisition of her child, the signer says that s/he is acquiring two languages. The cardinal reported in (18)a is in prenominal position (TWO LANGUAGE) and conveys a new-discourse information. Later, in her spontaneous narration, she explains that her hearing child is acquiring both Italian and LIS. The cardinal in (18)b is in postnominal position (LANGUAGE TWO) and it is associated to old-discourse information. As for the NMMs, the cardinal TWO in (18)a is produced with neutral facial expression, whereas the cardinal TWO in (18)b is realized with squinted eyes. The two cardinals look different also with respect to space location. The former is not directed toward any particular point in space, whereas the latter is produced with a forward movement toward a point in the signing space.

The second case relevant to the discussion was produced by a middle-aged signer from Florence and is reported in (19).

(19) a. LATER IX-1 TRY APPLY RELOCATE FLORENCE IX_LOC CEAUX **[TWO SISTER]** INSIDE

'Later I took a shot at applying for relocation in Florence where I had two sisters'

b. BEFORE [**SISTER TWO** CL] THINK DIFFERENT TURN

'Before, with my two sisters at home, my mental approach was different, we took turns'

I n (19)a, the signer introduces her two sisters as new referents. Here the cardinal is produced in prenominal position (TWO SISTER). In (19)b, she mentions the same referents again and produces the N>Card order (SISTER TWO). Interestingly, the

different NMMs correlate with the different discourse interpretations: neutral NMMs with new-discourse information and chin down together with lowered eyebrows with old-discourse information. The definite cardinal construction in (19)b includes a classifier sign which locates the two sisters in a specific point in the signing space (as we saw with elicited data in Figure 15).

After having discussed the two pairs in (18) and (19), it is possible to conclude that corpus data are consistent with the sign order patterns illustrated in (15) and (16). Indeed, cardinals included in constructions with indefinite reading can be found either before or after the noun and cardinals included in constructions with definite reading are found postnominally.

Turning to the second issue that needs to be double-checked in the corpus, I examined the case of prenominal cardinals. According to the acceptability judgments in (15) and (16), the order Card>N is associated with first-mentioned referents only. After careful examination, the result is that out of 67 prenominal cardinals in the corpus, none of them is connected to pre-established referents. Therefore, corpus data confirm the pattern emerged in elicited data in that cardinals included in constructions with definite reading are not found in prenominal position.

5.5. Analysis

After considering the data from both a quantitative and a qualitative perspective, a syntactic analysis accounting for the distribution of cardinals in LIS is presented in this section. In order to unravel the cardinals puzzle, two research questions need to be answered: i) why do indefinite cardinal constructions allow two orders?; ii) why do cardinals with definite interpretation occur in postnominal position only? Besides the cardinals puzzle, the puzzle within the puzzle is also examined here. Indeed, an account of the special distribution of MPs is provided.

In order to derive the different sign order options in LIS, I adopt Cardinaletti and Giusti's (2006) proposal, which has already been illustrated in Section 2.3.2., and I implement it within Cinque's framework (cf. Section 2.3.1.). In a nutshell, Cardinaletti and Giusti

(2006) propose that cardinals sit in two distinct projections, namely in a DP-external and a DP-internal projection. In article languages, when cardinals are not preceded by a definite determiner they are claimed to function as quantifiers and occupy the head of the QP. When cardinals are preceded by a definite determiner they are claimed to function as quantitative adjective and occupy the specifier position of a lower DP-internal projection.

In the spirit of Cardinaletti and Giusti (2006), I propose that when cardinals in LIS are included in a construction with indefinite reading they function as pure quantifiers since they actually quantify the referents introduced in the discourse. When cardinals appear in a construction with definite reading they do not count anything and are rather nominal modifiers used to refer back to an already mentioned and counted entity. I assume that some instances of pointing signs would correspond to the functional equivalent of the definite determiner attested in article spoken languages.[15] More systematic studies are needed in order to explain the functions of determiner-like pointing signs in LIS. However, as discussed in Section 5.4.3, definiteness is marked by prosodic and/or morphosyntactic strategies. The former are specific NMMs (squinted eyes, lowered eyebrows, chin down), the latter consist in a DP-internal classifier or an additional morpheme inducing the cardinal to move slightly forward and downward in a specific point in space.

5.5.1. Explaining the distribution of cardinals

Considering cardinals with definite reading, recall that they can occur in postnominal position only, as shown in the prototypical example in (16)b (CHILD TWO). I propose to derive this order with the structure represented in Figure 16. The cardinal is low in the structure and it is located in the projection of the quantitative adjective (QuantAP). Then, the nominal element (CHILD) moves passing the cardinal and goes into the specifier of a higher projection. For concreteness, I assume this to be the agreement projection dominating the projection that hosts the cardinal. In order to account for the

15 With respect to the pointing signs co-occurring with nouns in LIS, Bertone (2007: 158) claims that these elements function as demonstratives rather than articles. In line with Bertone's view, Brunelli (2011: 56) claims that LIS does not display definite articles. A different perspective is offered in Bertone (2011: 126) where DP-internal pointing signs are considered demonstratives when they are compulsory and articles when they are optional. Pointing signs are discussed in this dissertation in the methodology chapter (Chapter 3).

ungrammaticality of the Card>N order with definite interpretation (see Section 5.4.3), I claim that the NP-movement across the cardinal is obligatory.

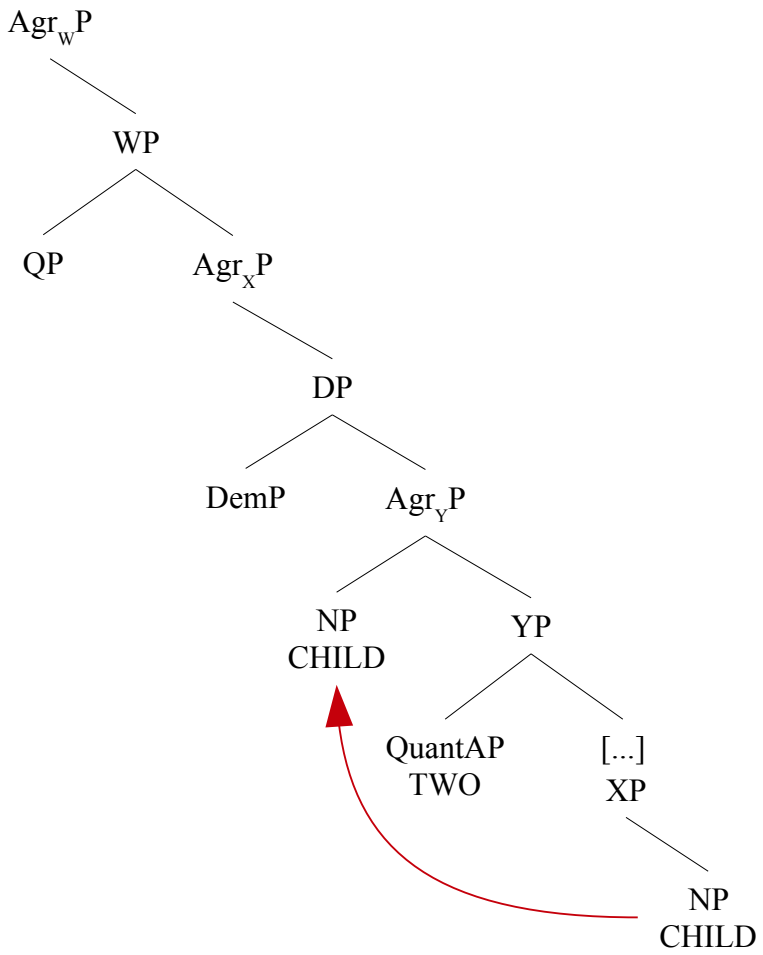


Figure 16: Structural representation of definite cardinal constructions

The NP movement illustrated in Figure 16 is obligatory. For concreteness, it might be argued that in this construction the noun is attracted toward a hierarchically higher position so that definiteness features can be checked.

The same type of NP movement is found with ordinary adjectives like DEAF, RED, etc., which occur in postnominal position most of the times (almost 90% of occurrences), as illustrated in Table 12.

Word order	n	%
Adj>N	53/470	11%
N>Adj	417/470	89%

Table 12: Distribution of adjectives in corpus data

The similar distribution shown by cardinals associated with definite reading and ordinary adjectives gives support to the idea that this group of cardinals is to be treated as a particular class of adjectives.

Turning to the construction with indefinite interpretation, the derivation should capture both the prenominal and postnominal cardinal. I propose to derive the two orders in (15) (TWO CHILD and CHILD TWO) with the structure represented in Figure 17. The cardinal occupies a high position in the structure and it is located in the projection of the quantifier (QP) which sits in the specifier of a functional projection above the DP.[16] The nominal element (CHILD) may raise across the cardinal targeting the specifier of a higher agreement projection.

16 In Cinque's framework (2005, 2012), quantifiers are not found in the backbone of the extended projection of the NP, but they sit in the specifier of a dedicated functional projection. Differently from Cinque (2005), Cardinaletti and Giusti (2006) analyze QP as a projection that directly dominates the extended projection of the NP and selects the DP as its complement. What is important to note at this point of the discussion is that the quantifier occupies a DP-external position.

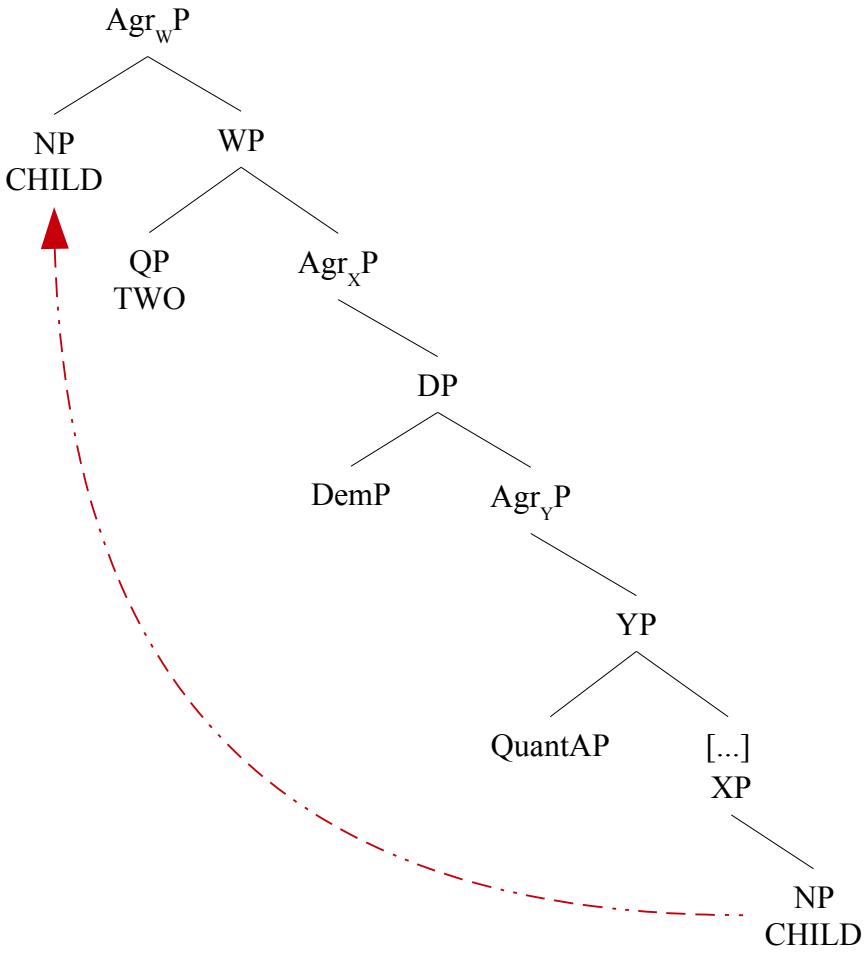


Figure 17: Structural representation of indefinite cardinal constructions

The NP movement across the cardinal is optional and it is indicated by a dashed arrow in Figure 17. According to my informants, the possibility to find the cardinal either before or after the noun does not give rise to two different semantic connotations. The optionality of this syntactic movement accounts for the possibility of having two sign orders. Notice that in this construction the noun does not have a definite interpretation and thus it does not have to raise along the structure for feature checking. This might provide an explanation for the optionality of the NP movement. What exactly triggers this movement is left as open issue, though.

If cardinals with indefinite reading are to be treated like quantifiers, then a similar distribution between the two types of modifiers should be expected. Table 13 shows how quantifiers are distributed with respect to the head noun in corpus data.

Word order	n	%
Quant>N	51/154	33%
N>Quant	103/154	67%

Table 13: Distribution of quantifiers in corpus data

The percentages in Table 13 indicate that quantifiers prefer the postnominal position. It is worth noting that these results may be contaminated by the fact that the category of quantifiers is an umbrella category containing different types of elements, namely universal, existential, and distributive quantifiers. The three different classes of quantifiers may have different distributions and hence different positions in the syntactic structure. In particular, in the light of Cinque's (2005, 2012) framework, one of these three classes of quantifiers may sit in a projection which is low enough to be located in the central portion of the hierarchy of LIS nominal expressions. If this conjecture is correct, cardinals may sit in that projection. A good candidate for this hypothesis could be the class of existential quantifiers (also called vague numerals) like 'many', 'few', etc. The internal subcategorization of quantifiers could help clarify whether this type of quantifiers and cardinal numerals show a similar syntactic behavior. If this is the case, the parallelism between cardinals and existential quantifiers could provide additional support for arguing the existence of two distinct syntactic positions.

Giusti (1994) observes that the alternation between prenominal and postnominal order is found in different languages. For example, Hebrew existential quantifiers may precede or follow the DP, as in (20).

(20) a. [$_{QP}$ meat [$_{DP}$ yeladim]] ?ohavim le-saxek

few boys love to-play

b. [$_{QP}$ [$_{DP}$ yeladim] [$_{Q'}$ meatim [$_{DP}$ ~~yeladim~~]]] ?ohavim le-saxek

boys few-MASC.PL love to-play

These data provide empirical support for the hypothesis that quantifiers and cardinals with quantificational reading occur in a DP-external position.

The fact that order differences correlate with definite/indefinite interpretation is not a novelty in the literature. Similarly to LIS, Shupamem (a Bantu language spoken in South Western Cameroon) displays this correlation. The position of the cardinal with respect to the noun reveals its semantic interpretation. When the cardinal follows the noun it yields a definite reading, whereas when it precedes the noun it is associated with an indefinite reading. The two cases are exemplified in (21) (adapted from Vázquez-Rojas, 2011: 235).

(21) a. pɛʔ pón
two child.PL
'two children'

b. pón pí pà:
child.PL AGR two
'the two children'

When the NP has a definite interpretation (see (21)b), the cardinal occurs in postnominal position and, interestingly, an agreement marker ('pí') appears between noun and cardinal. In LIS, too, the cardinal follows the noun when a definite interpretation is involved. As previously discussed, definiteness is transmitted through NMMs, a DP-internal classifier, or an extra morpheme inducing the cardinal to move toward a point in space. Such morpheme bears a resemblance to Shupamem agreement marker on the cardinal and might be revealing of a concordance phenomenon.

5.5.2. Explaining the exceptionality of Measure Phrases

Explaining the distribution of cardinals contained in MPs implies finding an account for

their categorical distribution.

My proposal is that the measure noun (KILO, METER, etc.) is not the real head of the NP, but rather together with the cardinal it constitutes a complex element that quantifies another nominal element. It is this other nominal element that is the real NP of the construction. So, in LIS the partitive-like element (which corresponds to the partitive PP in English and Italian) is the real head of the entire nominal expression. This head may remain silent as in (22), or it can be lexically expressed as in (23) (SAUSAGE).

(22) IX-1 RUN [TWO KILOMETER Ø]

'I ran two kilometers'

(23) [TWO KILOGRAMME SAUSAGE] IX-1 EAT DONE

'I ate two kilograms of sausages'

Evidence for this analysis is provided by the pair in (24). The sentence in (24)a shows the order Card>Measure N>Partitive N. Crucially, in (24)b, the partitive noun is moved to the left of the cardinal leaving the measure noun in situ.

(24) a. [TWO KILO SAUSAGE DELICIOUS] IX-1 BUY DONE

b. ? [SAUSAGE TWO KILO DELICIOUS] IX-1 BUY DONE

'I bought two kilos of delicious sausages'

Since by assumption DP-internal movements are only possible if the moved constituent contains the noun (cf. Section 2.2.1.), we have clear evidence that SAUSAGE in (24)b is the real head noun. This is so because this is the only moving element. Notice, indeed, that the quality adjective DELICIOUS has been stranded. The fact that the adjective and the MP are left in situ is evidence that the measure noun is not the head noun of the whole construction.

As a matter of fact, the measure noun must remain adjacent to the cardinal. Besides, no

lexical material can be inserted between these two elements, as shown in (25),[17] and no order permutations within the MP are allowed.

(25) a. * TWO DELICIOUS KILO SAUSAGE IX-1 BUY DONE

b. TWO KILO SAUSAGE DELICIOUS IX-1 BUY DONE

'I bought two kilos of delicious sausages'

I would like to speculate that the strict order Card>Measure N is derived by the syntactic spec-head configuration that characterizes the two elements. Syntax may be responsible for the distributional restrictions of such construction. The special relationship between cardinals and measure nouns in some cases results in incorporation.[18] An example is the MP THREE-MONTH, shown here in Figure 18. Other examples of overt incorporation are TWO-DAY, THREE-HOUR, FOUR-YEAR. In this respect, I take measure nouns as behaving like functional classifier-like elements.

Figure 18: MP incorporation THREE-MONTH

I propose to derive the two orders in (24) (TWO KILO SAUSAGE DELICIOUS and SAUSAGE TWO KILO DELICIOUS) with the structure represented in Figure 19. The

17 Although adjectives in LIS usually follow the noun (cf. Chapter 4, Figure 2), some value adjectives like BEAUTIFUL may sometimes precede the noun as shown in Chapter 4, Example (1)a.

18 Whether numeral incorporation is a syntactic or a phono-morphologic phenomenon (or maybe both) is an intriguing issue which could be usefully explored in further research.

cardinal sits in the specifier of the QP, as already argued above, while the measure noun (MeasN) sits in its head. Under this configuration incorporation may apply. The different ordering options are derived by moving the real head noun (SAUSAGE) obligatorily across the quality adjective (DELICIOUS) and optionally across the MP (TWO KILO).

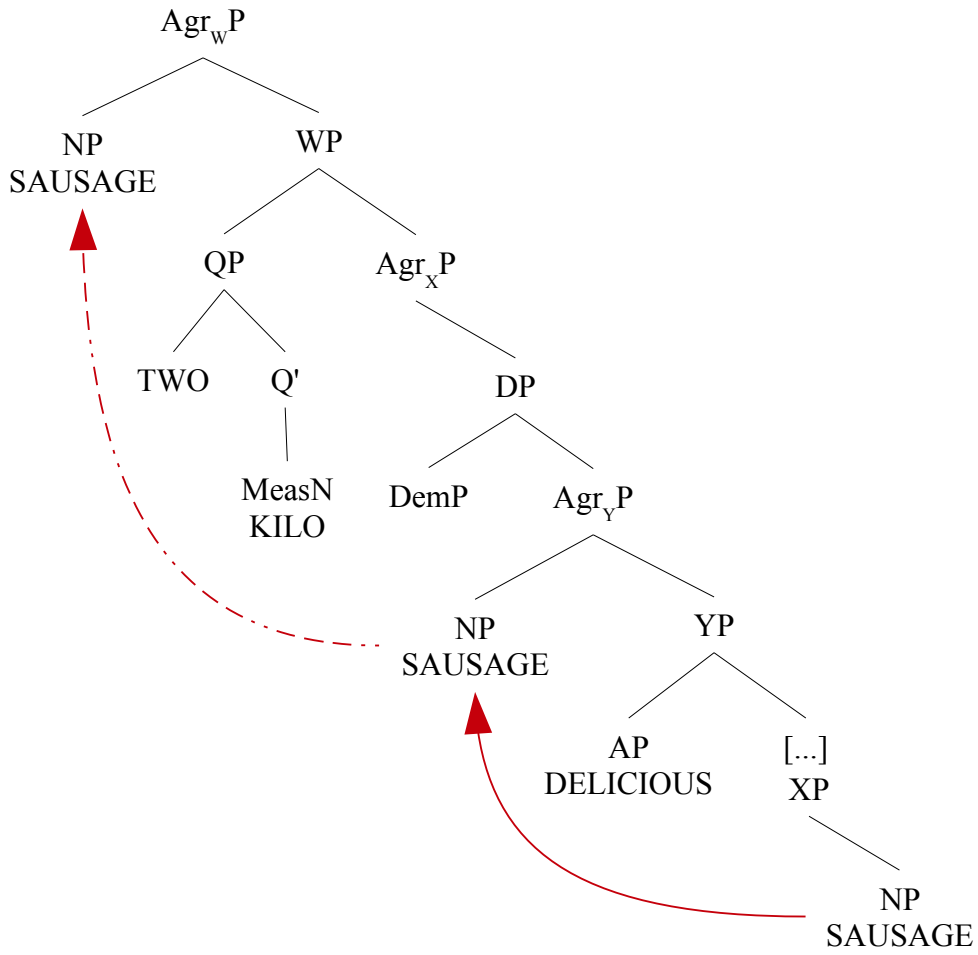


Figure 19: Structural representation of MPs

5.6. Conclusion

In the study described in this chapter, quantitative and qualitative data have been combined with the purpose of capitalizing on the advantages of each source.

Methodologically, I started with an in-depth analysis of corpus data. Then, I collected new data (direct elicitation and acceptability judgments) in order to test some specific linguistic contexts and to collect negative evidence. In the light of this qualitative part of the study, I reconsidered corpus data and checked for consistency. Such data variety has provided new insights into the distribution of cardinals in LIS.

Going back to the initial puzzle, it is now possible to depict a somewhat clearer picture. What has been previously claimed in the literature (N>Card) is still valid, but it offers an incomplete picture: prenominal cardinals are also possible.

In order to investigate the distribution of cardinals, two particular cases need to be analyzed separately, namely ONE (ambiguous between cardinal and determiner) and cardinals occurring in MPs (items with categorical distribution).

All other tokens show an apparently uncontrolled variation. However, this variability is far from being random, rather it seems to be sensitive to a fine-grained distinction between definite and indefinite nominal expressions, in addition to the sociolinguistic effects already discussed in Chapter 4 influencing the preferences when optionality is available. Cardinals can occur in either indefinite or definite constructions. These two linguistic environments give rise to some significant interpretational, distributional, and phonological differences. Indeed, the distinction between the definite and indefinite reading is marked by sign order manipulation (prenominal/postnominal cardinal), morphosyntactic properties (presence/absence of a classifier), and prosodic features (presence/absence of specific NMMs). In order to account for these two different patterns in a principled way, I claimed that cardinals in LIS can occupy two different syntactic positions (in the spirit of Cardinaletti & Giusti, 2006): they can be either quantifiers or quantitative adjectives. Under the proposed analysis, also the categorical behavior of cardinals included in MPs can be accounted for.

Conclusions

This dissertation investigated the syntax of nominal modification in LIS. The major contribution of this investigation consists of three types of innovations: theoretical, methodological, and empirical ones.

The main theoretical innovation consists in showing that language universals are generalizations that can be applied not only to languages displaying a categorical behavior, but also to languages with a considerable degree of intralinguistic variation. From a typological perspective, Universal 20 appears as a categorical pattern that classifies the codified aspects of languages. However, LIS does not reveal the presence of a categorical pattern; rather, it shows a particularly varying pattern where nominal modifiers are rather flexibly distributed with respect to the noun. Given these two divergent aspects (categorical vs. flexible), a problematic issue has emerged: how should LIS be classified? The combination of three apparently irreconcilable approaches to the facts of human language (linguistic typology, generative linguistics, and sociolinguistics) was used to tackle this theoretical challenge and provide an answer for it. This dissertation showed that LIS has a prevalent order (Num>N>Adj>Dem) and some alternative options. Typological universals and formal derivations can be applied to languages showing intralinguistic variation as long as none of the attested orders violates the constraints at their basis. Crucially, all the attested word order permutations in LIS can be derived by the formal analysis of Universal 20 proposed by Cinque's (2005) work. This fact proves that the constraints observed in crosslinguistic variation are also applicable to language-internal variation. In a sense, the validity of Universal 20 (i.e. only some word order combinations are allowed within the nominal domain) is now recast so that language-internal variation falls within its scope and languages instantiating variation need no longer be considered mixed languages (i.e. languages without order). Indeed, it seems that the generalization behind Universal 20 is capturing quite a deep organizational fact of human language along the lines proposed in Cinque's

formal account.

The second theoretical innovation concerns the fine-grained structure of nominal expressions. By definition, Universal 20 only accounts for the gross structure of the DP. Recent works by Cinque (2005, 2010, 2012) showed that the picture is even more complicated and can be represented through a well-defined hierarchical structure in which semantic aspects may play a role. The results of the investigation that has been carried out in this dissertation show that Cinque's hierarchy can also be applied to languages with a considerable amount of intralinguistic variation. In the case of LIS nominal domain, two key points have been identified in the structure. They are defined by the projections hosting possessive signs. Structurally, these modifiers are crucial because they divide the hierarchy into three main parts highlighting the different syntactic strategies needed to derive DPs in LIS. Indeed, the lower portion of the hierarchy undergoes NP movement, the central portion is characterized by absence of movement, and the higher portion is subject to a syntactic derivation through pied-piping. All in all, these findings enhance our understanding on the structure of LIS nominal expressions and on the mechanisms contributing to its final shape.

The methodological innovation consists in combining quantitative with qualitative procedures. The main advantage of examining corpus data through a systematic statistical analysis is that linguistic facts can be evaluated on a large scale obtaining a more accurate picture of the language and its internal variation. The main advantage of developing qualitative considerations on elicited data is that specific syntactic constructions can be analyzed in detail and native signer's judgments can help discern what is allowed and disallowed by language grammar. Capitalizing on the advantages of both research sources has proved to be particularly fruitful for the purposes of this dissertation.

The empirical innovation is represented by the enlargement of the empirical basis of Universal 20 through the observation of a sign language. According to the results of the studies presented in Chapter 4, the pattern resulting from the most frequent combinations of signs within LIS nominal expressions is Num>N>Adj>Dem. Although rare, this word order is attested in spoken languages (a.o. Hebrew, Basque, Celtic, Indonesian, Wolof, and a number of creole languages). This overall pattern is not rigid

at all, since alternative combinations are also possible. Most importantly, all of them are compatible with the predictions deriving from Universal 20.

However, this dissertation goes further on analyzing the fine-grained structure of LIS nominal modification. The study on the distribution of modifiers revealed that the variability attested in LIS is syntactically constrained by the same fine-grained structure as discussed in Cinque (2005, 2012). The existence of different orders can be explained in terms of processing costs of syntactic operations. More visible portions of the structure, namely the two peripheral edges, are more likely to undergo syntactic movement. On the contrary, less visible portions of the structure, specifically the central projections, are more prone to absence of NP movement. The observed variation is also influenced by social factors. In particular, signers from hearing families are less variable because they tend to over-generate the most frequent pattern, namely postnominal modifiers. They show fewer options like very fluent second-language learners. By contrast, native signers show more variability since they master a larger repertoire of syntactic constructions. The effect of signers' age indicates that the language is evolving since the postnominal pattern is being internalized as a rule of LIS grammar.

The study on the duration of modifiers provides empirical support to the results of the other corpus study and, crucially, reveals that the syntactic structure of nominal expressions is also visible to the prosodic domain as reflected in final lengthening.

The innovative combination of different methodologies provided new insights on cardinal numerals introducing an important addition to the picture of the nominal modification in LIS. Although predominantly prenominal in the corpus (as shown in the most frequent pattern above), cardinals are claimed by other authors to consistently follow the noun (Branchini, 2007; Bertone, 2007, 2009; Brunelli, 2011). The detail-oriented study on cardinals revealed that their apparently uncontrolled variation is sensitive to an even finer-grained distinction between definite and indefinite nominal expressions. Cardinals contained in definite constructions follow the noun, whereas those in indefinite constructions can either follow or precede the noun. The different syntactic behavior characterizing the two types of cardinals is captured by assuming that they occupy two different positions in the hierarchy (as proposed by Cardinaletti & Giusti, 2006). Furthermore, this study has highlighted the fact that the sign ONE and the

cardinals included in Measure Phrases should be separately analyzed. The sign ONE can be ambiguously used as cardinal and as indefinite determiner, whereas cardinals in Measure Phrases show a categorical distribution since they can only appear before the measure noun. These two potential confounders should not be included in the study of cardinals so that they do not contaminate the rest of the data leading to spurious conclusions.

With respect to the other classes of nominal modifiers, many questions concerning their distribution remain unanswered. For instance, quantifiers have proved to be predominantly postnominal in corpus data, still there is a considerable amount of prenominal quantifiers that cannot be simply ignored. As suggested in Section 5.5.1., it may be possible that different types of quantifiers (universal, existential, distributive) display different distributional patterns that are obscured when these subclasses are all considered under the same umbrella. Further experimental investigations are also needed to provide an in-depth analysis of the syntactic behavior characterizing the different classes of possessives. In particular, specific studies should evaluate whether the syntactic analysis outlined in this dissertation can effectively account for their distribution. Similarly to quantifiers and possessives, also the other modifier types would deserve a deeper analysis. A number of possible future investigations using the experimental setup presented in this dissertation for cardinals are recommended in order to advance our knowledge of the mechanisms driving nominal modification in LIS.

References

Abels, Klaus & Ad Neeleman. 2006. *Linear asymmetries and the LCA*. Manuscript, http://ling.auf.net/lingbuzz/000279. (20 August, 2014).

Abels, Klaus & Ad Neeleman. 2009. Universal 20 without the LCA. In José M. Brucart, Anna Gavarró & Jaume Solà (eds.), *Merging features: Computation, interpretation, and acquisition*, 60-79. Oxford: Oxford University Press.

Abner, Natasha. 2012. *There once was a verb: the predicative core of possessive and nominalization structures in American Sign Language*. Los Angeles, CA: University of California, doctoral dissertation.

Abney, Steven P. 1987. *The English Noun Phrase in its sentential aspect*. Cambridge, MA: MIT, doctoral dissertation.

Alexiadou, Artemis, Liliane Haegeman & Melita Stavrou. 2007. *Noun phrase in the generative perspective* (Studies in Generative Grammar 71). Berlin & New York, NY: Mouton de Gruyter.

Aronoff, Mark, Irit Meir & Wendy Sandler. 2005. The paradox of sign language morphology. *Language* 81(2). 301-344.

Baayen, Harald R. 2008. *Analyzing linguistic data. A practical introduction to statistics using R.* Cambridge, UK: Cambridge University Press.

Baayen, Harald R., Doug J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. In *Journal of Memory and Language* 59. 390-412.

Baker, Mark C. & Jim McCloskey. 2007. On the relationship of typology to theoretical syntax. *Linguistic Typology* 11(1). 285-296.

Barberà, Gemma. 2012. *The meaning of space in Catalan Sign Language (LSC). Reference, specificity and structure in signed discourse.* Barcelona: Universitat Pompeu

Fabra, doctoral dissertation.

Barbiers, Sjef. 2005. Word order variation in three-verb clusters and the division of labour between generative linguistics and sociolinguistics. In Leonie M.E.A. Cornips & Karen P. Corrigan (eds.), *Syntax and variation. Reconciling the biological and the social*, 233-264. Amsterdam: John Benjamins Publishing Company.

Barrett, Rusty. 2008. Linguistic differentiation and Mayan language revitalization in Guatemala. In *Journal of Sociolinguistics* 12(3). 275–305.

Berruto, Gaetano. 1997. *Fondamenti di sociolinguistica*. Bari: Laterza.

Bertone, Carmela. 2007. *La struttura del sintagma determinante nella Lingua dei Segni Italiana (LIS)*. Venezia: Ca' Foscari University, doctoral dissertation.

Bertone, Carmela. 2009. The syntax of noun modification in Italian Sign Language (LIS). *University of Venice Working Papers in Linguistics* 19. 7-28.

Bertone, Carmela. 2011. *Fondamenti di grammatica della Lingua dei Segni Italiana*. Milano: Franco Angeli.

Bertone, Carmela & Anna Cardinaletti. 2011. ll sistema pronominale della lingua dei segni italiana. In Anna Cardinaletti, Carlo Cecchetto & Caterina Donati (eds.), *Grammatica, lessico e dimensioni di variazione nella LIS*, 145-160. Milano: Franco Angeli.

Bošković, Željko. 2009. More on the No-DP analysis of article-less languages. *Studia Linguistica* 63. 187-203.

Boster, Carole Tenny. 1996. On the quantifier-noun phrase split in American Sign Language and the structure of quantified noun phrases. In William Edmondson & Ronnie B. Wilbur (eds.), *International Review of Sign Linguistics,* 159-208. Mahwah, NJ: Lawrence Erlbaum.

Branchini, Chiara. 2007. *On relativization and clefting in Italian Sign Language (LIS)*. Urbino: University of Urbino, doctoral dissertation.

Branchini, Chiara, Anna Cardinaletti, Carlo Cecchetto, Caterina Donati & Carlo Geraci. 2013. Wh-duplication in Italian Sign Language (LIS). *Sign Language and Linguistics* 16(2). 157-188.

Branchini, Chiara & Caterina Donati. 2009. Relatively different: Italian Sign Language relative clauses in a typological perspective. In Anikó Liptàk (ed.), *Correlatives cross-linguistically,* 157-191. Amsterdam: John Benjamins Publishing Company.

Brunelli, Michele. 2011. *Antisymmetry and sign languages. A comparison between NGT and LIS*. Utrecht: LOT Publications.

Bybee, Joan L. & Suzanne Fleischman. 1995. *Modality in grammar and discourse*. Amsterdam & New York, NY: John Benjamins Publishing Company.

Cardinaletti, Anna. 1998. On the deficient/strong opposition in possessive systems. In Artemis Alexiadou & Chris Wilder (eds.), *Possessors, predicates, and movement in the Determiner Phrase*, 17-53. Amsterdam: John Benjamins Publishing Company.

Cardinaletti, Anna & Giuliana Giusti. 1992. Partitive ne and the QP-hypothesis. In Elisabetta Fava (ed.), *Proceedings of the XVII meeting of Generative Grammar*, 121-141. Turin: Rosenberg & Sellier.

Cardinaletti, Anna & Giuliana Giusti. 2006. The syntax of quantified phrases and quantitative clitics. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell companion to syntax*, vol. V, 23-93. Oxford: Blackwell Publishing.

Cardinaletti, Anna & Michal Starke. 1999. The typology of structural deficiency: A case study of the three classes of pronouns. In Henk Van Riemsdijk (ed.), *Clitics in the Languages of Europe*, 145-234. Berlin: Mouton de Gruyter.

Cardinaletti, Anna, Carlo Cecchetto & Caterina Donati (eds.). 2011. *Grammatica, lessico e dimensioni di variazione nella LIS*. Milano: Franco Angeli.

Cecchetto, Carlo, Carlo Geraci & Sandro Zucchi. 2006. Strategies of relativization in Italian Sign Language. *Natural Language & Linguistic Theory*, 24(4). 945-975.

Cecchetto, Carlo, Carlo Geraci & Sandro Zucchi. 2009. Another way to mark syntactic dependencies: The case for right-peripheral specifiers in sign languages. *Language* 85(2). 278-320.

Cecchetto, Carlo, Serena Giudice & Emiliano Mereghetti. 2011. La raccolta del Corpus LIS. In Anna Cardinaletti, Carlo Cecchetto & Caterina Donati (eds.), *Grammatica, lessico e dimensioni di variazione nella LIS*, 55-68. Milano: Franco Angeli.

Charachidzé, Georges. 1989. Ubykh. In George B. Hewitt (ed.), *The indigenous languages of the Caucasus: The North West Caucasian Languages*, vol. 2, 357–459. Delmar, NY: Caravan Books.

Cheng, Lisa Lai-Shen & Norbert Corver. 2013. *Diagnosing syntax*. Oxford: Oxford University Press.

Chomsky, Noam. 1957. *Syntactic structures*, The Hague & Paris: Mouton.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In Roderick Jacobs & Peter Rosenbaum (eds.), *Reading in English transformational grammar*, 184-221. Waltham: Ginn.

Chomsky, Noam. 1971. *Problems of knowledge and freedom*. New York, NY: Pantheon.

Chomsky, Noam. 1980. *Rules and representations*. New York, NY: Columbia University Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam. 1986. *Knowledge of language: its nature, origin and use*. New York, NY: Praeger.

Chomsky, Noam. 1988. *Language and problems of knowledge: the Managua lectures*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by step: Essays on Minimalist syntax in honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*, 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and beyond. The cartography of syntactic structures*, vol. 3, 104-131. New York, NY: Oxford University Press.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*.

New York, NY: Oxford University Press.

Cinque, Guglielmo. 2003. Greenberg's Universal 20 and the Semitic DP. In Lars-Olof Delsing, Cecilia Falk, Gunlög Josefsson & Halldór Á. Sigursson (eds.), *Grammatik i fokus/Grammar in Focus*. *Festschrift for Christer Platzack 18 November 2003*, vol. 2, 243-251. Lund: Wallin & Dalholm.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistics Inquiry* 36. 315-332.

Cinque, Guglielmo. 2007. A note on linguistic theory and typology. *Linguistic Typology* 11. 93–106.

Cinque, Guglielmo. 2010. *The syntax of adjectives*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 2012. A partial map of extended functional projection of the NP. *Class lecture in advanced syntax*, a.y. 2011/2012, Ca' Foscari University of Venice.

Cinque, Guglielmo. 2013a. Cognition, typological generalizations, and Universal Grammar. *Lingua* 130. 50-65.

Cinque, Guglielmo. 2013b. Word order typology: a change of perspective. In Theresa Biberauer & Michelle Sheehan (eds.), *Theoretical approaches to disharmonic word orders*, 47–73. Oxford: Oxford University Press.

Cinque, Guglielmo. in preparation. *On the movement account of Greenberg's Universal 20: refinements and replies*. Manuscript, University of Venice.

Cinque, Guglielmo & Luigi Rizzi. 2008. The cartography of syntactic structures. In Vincenzo Moscati (ed.), *Working papers on language and cognition (CISCL)* 2. 43–59.

Coene, Martine & Yves D'Hulst. 2003. Introduction: The syntax and semantics of noun phrases. Theoretical background. In Martine Coene & Yves D'Hulst (eds.), *From NP to DP. The Syntax and Semantics of Noun Phrases*, vol. 1, 1-33. Amsterdam: John Benjamins Publishing Company.

Conte, Genny, Mirko Santoro, Carlo Geraci & Anna Cardinaletti. 2010. Why are you raising your eyebrows? *Language Resources and Evaluation Conference (LREC)*. 53-56.

Coseriu, Eugenio. 1973. *Lezioni di linguistica generale*. Torino: Boringhieri.

Crasborn, Onno & Han Sloetjes. 2008. Enhanced ELAN functionality for sign language corpora. *Language Resources and Evaluation Conference (LREC)*. 39-43.

Croft, William. 1990. *Typology and Universals*. Cambridge, UK: Cambridge University Press.

Croft, William. 2003. *Typology and Universals*. Cambridge, UK: Cambridge University Press.

Cysouw, Michael. 2010. Dealing with diversity: Towards an explanation of NP-internal word order frequencies. *Linguistic Typology* 14(2-3). 253-286.

Daniel, Michael. 2011. Linguistic typology and the study of language. In Jae Jung Song (ed.), *The Oxford handbook of linguistic typology*, 43-68. Oxford: Oxford University Press.

Dryer, Matthew S. 2006. *On Cinque on Greenberg's Universal 20*. Paper presented at Max-Planck- Institute für evolutionäre Anthropologie, Leipzig.

Dryer, Matthew S. 2011. On the order of demonstrative, numeral, adjective, and noun: an alternative to Cinque. Handout of a paper presented at the University of Pavia in June 2011. http://exadmin.matita.net/uploads/pagine/1898313034_cinqueH09.pdf. (20 August, 2014).

Dryer, Matthew S. 2013a. Order of subject, object and verb. In Matthew S. Dryer & Martin Haspelmath (eds.), The World Atlas of Language Structures (WALS) online. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/81. (20 August, 2014).

Dryer, Matthew S. 2013b. Polar questions. In: Matthew S. Dryer & Martin Haspelmath (eds.), The World Atlas of Language Structures (WALS) online. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/116. (20 August, 2014).

Dryer, Matthew S. & Martin Haspelmath (eds.). 2013. The World Atlas of Language Structures (WALS) online. Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info. (20 August, 2014).

Eckert, Penelope. 2012. Three waves of variation study: The emergence of meaning in

the study of sociolinguistic variation. *Annual Review of Anthropology* 41(1). 87-100.

Fischer, Susan D. 1978. Sign Language and creoles. In Patricia Siple (ed.), *Understanding language through sign language research*, 309-331. New York, NY: Academic Press.

Flydal, Leiv. 1951. Remarques sur certains rapports entre le style et l'état de langue. In *Norsk Tidsskrift for Sprogvidenskap* 16. 240-257.

Fukui, Naoki. 2001. Phrase structure. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 374-406. Malden: Blackwell Publishing.

Geraci, Carlo. 2009. Epenthesis in Italian Sign Language. In *Sign Language & Linguistics* 12(1). 3-51.

Geraci, Carlo & Robert Bayley. 2011. Chi, cosa, dove, perché, quando: la distribuzione dei segni wh- in LIS. In Anna Cardinaletti, Carlo Cecchetto & Caterina Donati (eds.), *Grammatica, lessico e dimensioni di variazione nella LIS*, 125-141. Milano: Franco Angeli.

Geraci, Carlo, Carlo Cecchetto & Sandro Zucchi. 2008. Sentential complementation in Italian Sign Language. In Michael Grosvald & Dionne Soares (eds.), *Western Conference On Linguistics (WECOL).* 46-58.

Geraci, Carlo, Marta Gozzi, Costanza Papagno & Carlo Cecchetto. 2008. How grammar can cope with limited short-term memory: Simultaneity and seriality in sign languages. *Cognition* 106. 780-804.

Geraci, Carlo, Katia Battaglia, Anna Cardinaletti, Carlo Cecchetto, Caterina Donati, Serena Giudice & Emiliano Mereghetti. 2011. The LIS corpus project. A discussion of sociolinguistic variation in the lexicon. *Sign Language Studies* 11(4). 528-574.

Geraci, Carlo, Robert Bayley, Chiara Branchini, Anna Cardinaletti, Carlo Cecchetto, Caterina Donati, Serena Giudice, Emiliano Mereghetti, Fabio Poletti, Mirko Santoro & Sandro Zucchi. 2010. Building a corpus for Italian Sign Language: Methodological issues and some preliminary results. *Language Resources and Evaluation Conference (LREC).* 98-101.

Giusti, Giuliana. 1993. *La sintassi dei determinanti*. Padova: UNIPRESS.

Giusti, Giuliana. 1994. L'ordine NQ in lingue QN. In Roberto Dolci & Giuliana Giusti (eds.), *Studi di grammatica tedesca e comparativa*, 49-66. Venezia: Centro Linguistico Interfacoltà.

Giusti, Giuliana. 1996. Is there a FocusP and a TopicP in the noun phrase? *University of Venice Working Papers in Linguistics* 6. 105-128.

Giusti, Giuliana. 2002. The functional structure of Noun Phrases: A bare phrase structure approach. In Guglielmo Cinque (ed.), *The Cartography of Syntactic Structures: Functional structure in DP and IP*, vol. 1, 54-90. Oxford: Oxford University Press.

Goldin-Meadow, Susan. 2012. Homesign: gesture to language In Roland Pfau, Markus Steinbach & Bencie Woll, *Sign language: An international handbook*, 601-625. Berlin: De Gruyter Mouton.

Greenberg, Joseph H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Human Language*, 73-113. Cambridge, MA: MIT Press.

Gries, Stefan T. 2013. *Statistics for linguistics with R. A practical introduction*. Berlin & Boston: De Gruyter Mouton.

Grimshaw, Jane. 1991. *Extended projection*. Manuscript. Waltham, MA: Brandeis University.

Guasti, Maria Teresa. 2007. *L'acquisizione del linguaggio. Un'introduzione*. Milano: Cortina.

Gumperz, John J. 1982. *Discourse Strategies*. Cambridge, UK: Cambridge University Press.

Haegeman, Liliane. 1994. *Introduction to government and binding theory.* Cambridge, MA: Blackwell Publishing.

Halle, Morris & Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In Kenneth Hale & S. Jay Keyser (eds.), *The view from building 20*, 111-176. Cambridge, MA: MIT Press.

Haspelmath, Martin. 1990. The grammaticalization of passive morphology. *Studies in*

Language 14(1). 25-71.

Hawkins, John A. 1983. *Word order universals*. New York, NY: Academic Press.

Hawkins, John A. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Jackendoff, Ray. 1977. *X-bar-syntax: A study of phrase structure*. Cambridge, MA: MIT Press.

Jaeger, T. Florian. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59. 434-446.

Johnston, Trevor. 2001. The lexical database of Auslan (Australian Sign Language). *Sign Language & Linguistics* 4 (1/2). 145-169.

Johnston, Trevor & Onno Crasborn. 2006. The use of ELAN software annotation software in the creation of sign language corpora. Presentation at *Electronic Metastructure for Endangered Languages Data (E-MELD)*. Lansing, MI. June 20-22.

Johnston, Trevor & Adam Schembri. 2006. Issues in the creation of a digital archive of a signed language. In Linda Barwick & Nicholas Thieberger (eds.), *Sustainable data from digital fieldwork*, 7-16. Sydney: University of Sydney Press.

Kayne, Richard. 1984. *Connectedness and binary branching*. Dordrecht: Foris Publishers.

Kayne, Richard. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.

Klima, Edward S. & Ursula Bellugi. 1979. *The Signs of Language*. Cambridge, MA: Harvard University Press.

Kroch, Anthony S. 2000. Syntactic change. In Mark Baltin & Chris Collins (eds.), *The handbook of contemporary syntactic theory*, 629-739. Malden, MA: Blackwell Publishing.

Labov, William. 1966. *The social stratification of English in New York City*. Washington, DC: Center for Applied Linguistics.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia, PA: University of Pennsylvania Press.

Labov, William. 1990. The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2. 205-254.

Labov, William. 2001. *Principles of linguistic change (social factors)*. Oxford: Blackwell Publishing.

Labov, William. 2008. *Quantitative reasoning in linguistics*. Manuscript, http://www.ling.upenn.edu/~wlabov/Papers?QRL.pdf. (20 August, 2014).

Lai, Yu-Ting Vivian. 2005. *Noun phrase in Taiwan Sign Language*. Taiwan: National Chung Cheng University, MA thesis.

Larson, Richard K. 1988. On the double object construction. *Linguistic Inquiry* 19(3). 335-391.

Littlewood, William. 1984. *Foreign and second language learning: language acquisition research and its implications for the classroom*. Cambridge, UK: Cambridge University Press.

Lucas, Ceil, Robert Bayley & Clayton Valli. 2001. *Sociolinguistic variation in American Sign Language*. Washington, DC: Gallaudet University Press.

MacLaughlin, Dawn. 1997. *The structure of Determiner Phrases: Evidence from American Sign Language*. Boston, MA: Boston University, doctoral dissertation.

Mantovan, Lara & Pietro Celo. 2008. Mettere i segni per iscritto. *L'Educazione dei Sordi*, vol. CIX, 3/4. 177-186.

Mantovan, Lara, Carlo Geraci & Anna Cardinaletti. 2014. Addressing the cardinals puzzle: new insights from non-manual markers in Italian sign language. *Language Resources and Evaluation Conference (LREC)*. 113-116.

Martin, Randi C. & Cristina Romani. 1994. Verbal working memory and sentence comprehension: A multiple-components view. *Neuropsychology* 8. 506–523.

McNally, Louise & Gemma Boleda. 2004. Relational adjectives as properties of kinds. In Olivier Bonami & Patricia Cabredo Hofherr (eds.), *Empirical issues in syntax and semantics*, vol. 5, 179-196. Paris: Colloqué de Syntax et Sémantique à Paris.

Meier, Richard P., Kearsy A. Cormier & David G. Quinto-Pozos (eds.). 2002. *Modality and structure in signed and spoken languages*. Cambridge, UK: Cambridge University

Press.

Meir, Irit, Wendy Sandler, Carol Padden & Mark Aronoff. 2010. Emerging sign languages. In Marc Marschark & Patricia E. Spencer (eds.), *Oxford handbook of Deaf studies, language, and education*, vol. 1, 267-280. Oxford: Oxford University Press.

Miller, George & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of mathematical psychology*, 419-491. New York, NY: John Wiley & Sons.

Mioni, Alberto M. 1983. Italiano tendenziale: osservazioni su alcuni aspetti della standardizzazione. In Paola Benincà, Manlio Cortelazzo, Aldo Luigi Prosdocimi, Laura Vanelli & Alberto Zamboni (eds.), *Scritti linguistici in onore di Giovan Battista Pellegrini,* vol. 1, 495-517. Pisa: Pacini.

Neidle, Carol & Joan Nash. 2012. The noun phrase. In Roland Pfau, Markus Steinbach & Bencie Woll (eds.), *Sign language: An international handbook*, 265-292. Berlin: De Gruyter Mouton.

Ohser, Erich. 2000. *Politische Karikaturen, Zeichnungen, Illustrationen und alle Bildgeschichten Vater und Sohn*. Konstanz: Sudverlag GmbH.

Padden, Carol. 1988. *Interaction of morphology and syntax in American Sign Language*. New York: Garland Publishing.

Pfau, Roland, Markus Steinbach & Bencie Woll (eds.). 2012. *Sign language: An international handbook*. Berlin: De Gruyter Mouton.

Pollock, Jean-Yves. 1989. Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry* 20(3). 365-424.

R Core Team. 2013. *R: A language and environment for statistical computing*. R foundation for statistical computing, Vienna. http://www.R-project.org/. (20 August, 2014).

Radutzky, Elena. 2001. *Dizionario bilingue elementare della lingua italiana dei segni*. Roma: Edizioni Kappa.

Rijkhoff, Jan. 2002. *The Noun Phrase*. Oxford: Oxford University Press.

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman (ed.),

Elements of grammar. Handbook in generative syntax, 281-337. Dordrecht: Kluwer.

Rutkowski, Paweł, Małgorzata Czajkowska-Kisil, Joanna Łacheta & Anna Kuder. 2014. The syntax of adjectival modification in Polish Sign Language (PJM). Presentation at *Olomouc Linguistics Colloquium (OLINCO)*. Univerzita Palackého v Olomouci, Olomouc. June 5.

Sankoff, Gillian & Suzanne Laberge. 1973. On the acquisition of native speakers by a language. *Kivung* 6. 32-47.

Schütze, Carson T. 2014. Judgment data. In Robert J. Podesva & Devyani Sharma (eds.), *Research methods in linguistics*, ch. 3. Cambridge, UK: Cambridge University Press.

Scott, Gary-John. 2002. Stacked Adjectival Modification and the Structure of Nominal Phrases. In Guglielmo Cinque (ed.), *The Cartography of Syntactic Structures: Functional structure in DP and IP*, vol. 1, 91–120. Oxford: Oxford University Press.

Shlonsky, Ur. 2004. The form of Semitic Noun Phrases. *Lingua* 114 (12). 1465-1526.

Sportiche, Dominique. 1988. A theory of floating quantifiers and its corollaries for constituent structure. *Linguistic Inquiry* 19. 425-449.

Sproat, Richard & Chilin Shih. 1991. The cross-linguistic distribution of adjective ordering restrictions. In Carol Georgopoulos & Roberta Ishihara (eds.), *Interdisciplinary approaches to language: Essays in honor of S. Y. Kuroda*, 565-593. Dordrecht: Kluwer.

Starks, Donna & Zita McRobbie-Utasi. 2001. Collecting sociolinguistic data: Some typical and some not so typical approaches. *New Zealand Journal of Sociology* 16. 79-92.

Szabolcsi, Anna. 1987. Functional categories in the noun phrase. In Istvan Kenesei (ed.), *Approaches to Hungarian: Theories and analyses*, vol. 2, 167–189. Szeged: JATE.

Szabolcsi, Anna. 1994. The noun phrase. In Ferenc Kiefer & Katalin E. Kiss (eds.), *The syntactic structure of Hungarian*, 179- 274. New York, NY: Academic Press.

Tagliamonte, Sali A. 2006. *Analysing sociolinguistic variation*, Cambridge, UK:

Cambridge University Press.

Tang, Gladys & Felix Sze. 2002. Nominal expressions in Hong Kong Sign Language: Does modality make a difference? In Richard P. Meier, Kearsy A. Cormier & David G. Quinto-Pozos (eds.), *Modality and structure in signed and spoken languages*, 296-321. Cambridge, UK: Cambridge University Press.

Van Herreweghe, Mieke & Myriam Vermeerbergen. 2012. Data collection. In Roland Pfau, Markus Steinbach & Bencie Woll (eds.), *Sign language. An international handbook,* 1023-1045. Berlin: de Gruyter Mouton.

Vázquez-Rojas, Violeta. 2011. Definite and indefinite numeral phrases in Shupamem. In Neil Myler & Jim Wood (eds.), *New York University Working Papers in Linguistics (NYUWPL)*, vol. 3. 231-244.

Vietti, Alessandro. 2005. Approcci quantitativi all'analisi della variazione linguistica: il caso di GOLDVARB. *Linguistica e Filologia* 20. 31-69. http://dspaceunibg.cilea.it/handle/10446/149. (20 August, 2014).

Volterra, Virginia (ed.). 1987. *La lingua italiana dei segni.* Bologna: Il Mulino.

Weinreich, Uriel, William Labov & Marvin Herzog. 1968. Empirical foundations for a theory of language change. In Winfred P. Lehmann & Yakov Malkiel (eds.), *Directions for historical linguistics*, 95-198. Austin, TX: University of Texas Press.

Yadroff, Michael & Loren Billings. 1998. The syntax of approximative inversion in Russian (and the general architecture of nominal expressions). In Željko Bošković, Steven Franks & William Snyder (eds.), *Formal approaches to Slavic linguistics: The Connecticut Meeting 1997*, 319–338. AnnArbor, MI: Michigan Slavic Publications.

Zeshan, Ulrike (ed.). 2006. *Interrogative and negative constructions in sign languages.* Nijmegen: Ishara Press.

Zeshan, Ulrike. 2008. Roots, leaves and branches. The typology of sign languages. In Ronice Müller de Quadros (ed.), *Sign languages: spinning and unraveling the past, present and future,* 671-695. Petrópolis: Editora Arara Azul.

Zeshan, Ulrike, Cesar Ernesto Escobedo Delgado, Hasan Dikyuva, Sibaji Panda & Connie de Vos. 2013. Cardinal numerals in rural sign languages: Approaching cross-

modal typology. *Linguistic Typology* 17. 357–396.

Zhang, Niina Ning. 2007. Universal 20 and Taiwan Sign Language. *Sign Language & Linguistics* 10. 55-81.

Università
Ca'Foscari
Venezia

DEPOSITO ELETTRONICO DELLA TESI DI DOTTORATO

DICHIARAZIONE SOSTITUTIVA DELL'ATTO DI NOTORIETÀ

(Art. 47 D.P.R. 445 del 28/12/2000 e relative modifiche)

Io sottoscritta Lara Mantovan

nata a Bassano del Grappa (prov. VI) il 09/11/1985

residente a Romano d'Ezzelino (prov. VI) in via Don Giovanni Minzoni n. 19

Matricola (se posseduta) 824989 Autrice della tesi di dottorato dal titolo:

Nominal modification in Italian Sign Language (LIS) ……...

……..…….

Dottorato di ricerca in Scienze del Linguaggio

Ciclo 27°

Anno di conseguimento del titolo a.a. 2014/2015

DICHIARO

di essere a conoscenza:

1) del fatto che in caso di dichiarazioni mendaci, oltre alle sanzioni previste dal codice penale e dalle Leggi speciali per l'ipotesi di falsità in atti ed uso di atti falsi, decado fin dall'inizio e senza necessità di nessuna formalità dai benefici conseguenti al provvedimento emanato sulla base di tali dichiarazioni;

2) dell'obbligo per l'Università di provvedere, per via telematica, al deposito di legge delle tesi di dottorato presso le Biblioteche Nazionali Centrali di Roma e di Firenze al fine di assicurarne la conservazione e la consultabilità da parte di terzi;

3) che l'Università si riserva i diritti di riproduzione per scopi didattici, con citazione della fonte;

4) del fatto che il testo integrale della tesi di dottorato di cui alla presente dichiarazione viene archiviato e reso consultabile via internet attraverso l'Archivio Istituzionale ad Accesso Aperto dell'Università Ca' Foscari, oltre che attraverso i cataloghi delle Biblioteche Nazionali Centrali di Roma e Firenze;

5) del fatto che, ai sensi e per gli effetti di cui al D.Lgs. n. 196/2003, i dati personali raccolti saranno trattati, anche con strumenti informatici, esclusivamente nell'ambito del procedimento per il quale la presentazione viene resa;

6) del fatto che la copia della tesi in formato elettronico depositato nell'Archivio Istituzionale ad Accesso Aperto è del tutto corrispondente alla tesi in formato cartaceo, controfirmata dal tutor, consegnata presso la segreteria didattica del dipartimento di riferimento del corso di dottorato ai fini del deposito presso l'Archivio di Ateneo, e che di conseguenza va esclusa qualsiasi responsabilità dell'Ateneo stesso per quanto riguarda eventuali errori, imprecisioni o omissioni nei contenuti della tesi;

7) del fatto che la copia consegnata in formato cartaceo, controfirmata dal tutor, depositata nell'Archivio di Ateneo, è l'unica alla quale farà riferimento l'Università per rilasciare, a richiesta, la dichiarazione di conformità di eventuali copie.

Data _____ Firma _____

AUTORIZZO

- l'Università a riprodurre ai fini dell'immissione in rete e a comunicare al pubblico tramite servizio online entro l'Archivio Istituzionale ad Accesso Aperto il testo integrale della tesi depositata;

- l'Università a consentire:
    - la riproduzione a fini personali e di ricerca, escludendo ogni utilizzo di carattere commerciale;
    - la citazione purché completa di tutti i dati bibliografici (nome e cognome dell'autore, titolo della tesi, relatore e correlatore, l'università, l'anno accademico e il numero delle pagine citate).

DICHIARO

1) che il contenuto e l'organizzazione della tesi è opera originale da me realizzata e non infrange in alcun modo il diritto d'autore né gli obblighi connessi alla salvaguardia di diritti morali od economici di altri autori o di altri aventi diritto, sia per testi, immagini, foto, tabelle, o altre parti di cui la tesi è composta, né compromette in alcun modo i diritti di terzi relativi alla sicurezza dei dati personali;

2) che la tesi di dottorato non è il risultato di attività rientranti nella normativa sulla proprietà industriale, non è stata prodotta nell'ambito di progetti finanziati da soggetti pubblici o privati con vincoli alla divulgazione dei risultati, non è oggetto di eventuali registrazione di tipo brevettuale o di tutela;

3) che pertanto l'Università è in ogni caso esente da responsabilità di qualsivoglia natura civile, amministrativa o penale e sarà tenuta indenne a qualsiasi richiesta o rivendicazione da parte di terzi.

A tal fine:

- dichiaro di aver autoarchiviato la copia integrale della tesi in formato elettronico nell'Archivio Istituzionale ad Accesso Aperto dell'Università Ca' Foscari;

- consegno la copia integrale della tesi in formato cartaceo presso la segreteria didattica del dipartimento di riferimento del corso di dottorato ai fini del deposito presso l'Archivio di Ateneo.

Data _____ Firma _____

La presente dichiarazione è sottoscritta dall'interessato in presenza del dipendente addetto, ovvero sottoscritta e inviata, unitamente a copia fotostatica non autenticata di un documento di identità del dichiarante, all'ufficio competente via fax, ovvero tramite un incaricato, oppure a mezzo posta

Firma del dipendente addetto ….....

Ai sensi dell'art. 13 del D.Lgs. n. 196/03 si informa che il titolare del trattamento dei dati forniti è l'Università Ca' Foscari – Venezia.
I dati sono acquisiti e trattati esclusivamente per l'espletamento delle finalità istituzionali d'Ateneo l'eventuale rifiuto di fornire i propri dati personali potrebbe comportare il mancato espletamento degli adempimenti necessari e delle procedure amministrative di gestione delle carriere studenti. Sono comunque riconosciuti i diritti di cui all'art. 7 D. Lgs. n. 196/03.

Estratto per riassunto della tesi di dottorato

L'estratto (max. 1000 battute) deve essere redatto sia in lingua italiana che in lingua inglese e nella lingua straniera eventualmente indicata dal Collegio dei docenti.

L'estratto va firmato e rilegato come ultimo foglio della tesi.

Studente: Lara Mantovan matricola: 824989

Dottorato: Scienze del Linguaggio

Ciclo: XXVII

Title of the thesis: "Nominal modification in Italian Sign Language (LIS)"

Abstract in English

This dissertation presents a syntactic analysis of nominal expressions in LIS combining three different theoretical frameworks: linguistic typology, generative linguistics, and sociolinguistics. Three empirical studies are presented: the first one investigates the syntactic distribution of nominal modifiers, the second one analyzes the duration of these modifiers, while the last study focuses on the syntax of cardinals in LIS. Results from the first two studies show that LIS nominal expressions display a considerable amount of syntactic variation which is influenced by both linguistic and social factors. The third study reveals a definiteness effect on the distribution of cardinals in LIS and analyzes the syntax of Measure Phrases by considering the lexical noun as the head of the entire construction. The combination of three different methodological approaches to the facts of language has proven to be extremely fruitful to account for language internal variation.

Titolo della tesi: “La modificazione nominale nella lingua dei segni italiana (LIS)”

Estratto in italiano

Questa tesi presenta un’analisi sintattica delle espressioni nominali in LIS attraverso la combinazione di tre diversi approcci teorici (tipologia linguistica, linguistica generativa e sociolinguistica). Nella tesi sono presentati tre studi empirici: il primo esamina la distribuzione sintattica dei modificatori nominali, il secondo analizza la durata di questi modificatori, il terzo si concentra sulla sintassi dei cardinali in LIS. I risultati dei primi due studi dimostrano che le espressioni nominali in LIS sono caratterizzate da una considerevole variazione linguistica che è influenzata da fattori sia linguistici che sociali. Il terzo studio rivela un effetto di definitezza sulla distribuzione dei cardinali e analizza la sintassi delle Measure Phrases considerando il nome lessicale come la testa dell’intera costruzione. La combinazione di tre diversi approcci metodologici ai dati linguistici si è rivelata proficua nella discussione della variazione intralinguistica.

Firma dello studente

_____