ma per seguir virtute e canoscenza

SISSA

Scuola Internazionale Superiore di Studi Avanzati
International School for Advanced Studies

Struggling for Structure

Cognitive origins of grammatical diversity and their implications for the Human Faculty of Language

Candidate:

Alan Langus

Supervisor:

Marina Nespor

Thesis submitted for the degree of Doctor of Philosophy in
Cognitive Neuroscience
Trieste, 2010

Jury

Laila Craighero

Human Physiology Section

Universita degli Studi di Ferrara, Ferrara Italy

Núria Sebastián-Gallés

Institució Catalana de Recerca i Estudis Avançats (ICREA) and Departament de Tecnologies de la Informació i les Comunicacions,

Universitat Pompeu Fabra, Barcelona, Spain.

Jacques Mehler

Cognitive Neuroscience Sector,

International School for Advanced Studies (SISSA/ISAS), Trieste, Italy.

Luigi Rizzi

Interdepartmental Center of Cognitive Studies on Language (CISCL),

Uinversita degli Studi di Siena, Siena, Italy.

Alessandro Treves

Cognitive Neuroscience Sector,

International School for Advanced Studies (SISSA/ISAS), Trieste, Italy.

Contents

List of Figures

Chapter 2

Chapter 3

Chapter 4

Acknowledgements

First and foremost, I would like to thank Marina Nespor. I could not have been luckier in finding a supervisor whose guidance has broadened my horizons, whose open-mindedness has allowed me to play with some pretty crazy ideas, whose belief in the project sent me to three different continents and whose support made these four years an experience of a life-time. Thank you!

I would also like to thank Jacques Mehler for taking me into his lab, for letting me carry out experiments, and for giving me all the advice and guidance.

This work would not have been possible without the Language Commando at SISSA – also known as the Language, Cognition and Development Laboratory. Special thanks for their time, assistance and invaluable advice go to Jean-Remy Hochmann, Ricardo A. Bion and Erika Marchetto. I am also grateful to Silvia Benavides, Luca Bonatti, Ansgar Endress, Francesca Gandolfo, Judit Gervain, David Maximiliano Gomez, Ágnes Kovács, Marcela Peña, Amanda Saksida, Mohinish Shukla, Ernő Téglás and Juan M. Toro. Last, but not least, I am in debt to Marijana Sjekloca for always taking the time to help me!

I am very grateful to Reiko Mazuka from the Laboratory of Language Development at RIKEN Brain Science Institute in Japan and Aslı Göksel from the Department of Western Languages and Literatures at Boğaziçi University in Istanbul for hosting me and for allowing me to use their facilities. I would also like to thank Alessandro Treves who helped me invaluably during the first months at SISSA.

Joana Costa do Carmo, Emiliana Papeo, Liuba Papeo and Chiara Nidasio – thank you for always being there for me!

Finally I would like to thank my mother for believing in me throughout the past ten years, through the ups and downs that brought me here!

Chapter 1

Introduction

There are between 5,000 and 8,000 distinct living languages spoken in the world today that are characterized by both exceptional diversity as well as significant similarities. Many researchers believe that at least part of this ability to communicate with language arises from a uniquely human Faculty of Language (c.f. Hauser, Chomsky, & Fitch, 2002; Pinker & Jackendoff, 2005). The traditional approach to the study of this uniquely human ability, is concerned with trying to understand how language is accommodated in the cognitive structure of the human mind, how it arises from the neural mechanisms that constitute the human brain and how it is linked to our genetic code. Most researchers have thus been looking at the similarities – especially the universal structural characteristics – among individual human languages (e.g. Greenberg, 1963; Chomsky, 1957; Jackendoff, 1997).

However, there are also significant differences among individual languages that can, by and large, be divided into two categories. On the one hand, some of the differences between individual languages have arisen due to the specific cultural context in which a given language has evolved through continuous use. For example, in English we say *dog* and in Italian *cane*. On the other hand, there exist also differences that appear too systematic to be simple coincidence of cultural evolution. For example, while the individual lexical entry for the concept 'dog' may vary across individual languages, all languages have the ability to assign a specific meaning to a sequence of sounds or signs that constitute a word.

The majority of theories of the Human Faculty of Language originate from the assumption of the existence of language universals shared by all individual human languages (Greenberg, 1963; Chomsky, 1957; Chomsky, 1995; Jackendoff, 1997 among others). This is especially true for many theories of grammar, that do not simply describe the specific grammars of individual languages, but attempt to make predictions about how linguistic knowledge is represented in the human mind (Chomsky, 1957). It is therefore reasonable to assume that the Language Faculty must also have the cognitive capacity to perform the computations of grammar to generate and interpret the structure of sentences (i.e. syntax). However, contrary to the specific lexical entries that vary randomly across languages, the differences in linguistic structure appear too systematic to emerge coincidentally through cultural evolution. For example, there are six logically possible ways of arranging words in a sentence according to their grammatical function of Subject, Object and Verb.[1] Were word order determined culturally, we would expect these six orders to be equally distributed among the world's languages. This is not the case: the majority of the world's languages rely on either the SOV or the SVO order (Dryer, 2005). Arguably, such systematic differences between linguistic structures are as indicative of the cognitive and biological structure of the Human Faculty of Language as the overall similarities between the worlds' languages (c.f. Kayne, 1994).

In the following, I will argue that by looking at the structural similarities between the world's languages, researchers have often failed to place equal value in the systematic differences observed among them (for a recent discussion see Evans & Levinson, 2009). The main aim of the thesis is to highlight the fact that a sustainable

[1] Humans can only utter one word at the time and, thus, words in the speech signal are arranged sequentially. If we combine the grammatical functions Subject, Object and Verb in all the possible combinations then we get six different word orders (OSV, OVS, SOV, SVO, VOS, VSO) – all of which have actually also been attested among the world's languages. (Dryer, 2005)

approach to understanding the Human Faculty of Language will not only pay attention to the structural similarities between the world's languages, but will also have to explain the systematic differences observed among them. The work presented below provides a concrete proposal of how the systematic structural differences among languages may emerge, what this means in terms of the nature and evolution of the Human Faculty of Language, and how it affects the way languages are acquired.

1.1 The structure of the Human Faculty of Language

It has been argued that the human faculty of language is modular and that it is possible to identify different cognitive systems responsible for specific linguistic tasks (Chomsky, 2000; Pinker, 1990; Fodor, 1983). The production and comprehension of language (either spoken or signed) require at least three task-specific cognitive systems: the conceptual system (semantics) that provides and interprets the meaning of linguistic utterances; the sensory-motor system (phonology and phonetics) that produces and perceives the actual sounds and signs of language; and the computational system of grammar (syntax) that links meaning with sounds (or signs) by generating the structure of sentences (Hauser, Chomsky, & Fitch, 2002; Pinker & Jackendoff, 2005).

For many linguists the primary difference between human language and the communication systems of other animals lies in the fact that in human language the interface between the sensory-motor and the conceptual system must necessarily be mediated by the computational system of grammar (Chomsky, 1957; Chomsky, 1995; Jackendoff, 1997). For example, in animal calls the sensory input (e.g. a holistic vocalization) is directly mapped to the meaning in the conceptual system (e.g. the presence of a snake) and there is no evidence that animals can use the combinatorial capacity of the computational system to create sounds with varied meaning (Hauser & Bever, 2008; Fitch & Hauser, 2004; Gentner, Fenn, Margoliash, & Nusbaum, 2006)[2].

[2] It is important to note that birds, rodents, and primates can compute some components of human grammatical competence. For instance, songbirds have been shown to be able to compute simple A^nB^n grammars (Gentner, Fenn, Margoliash, & Nusbaum, 2006) that were thought similar to recursion (Hauser, Chomsky, & Fitch, 2002; see however Hochmann,

In contrast, when humans communicate with language, they almost never abandon the syntax of their native language, and tend to use it robustly even when learning a new language (Odlin, 1989; Jansen, Lalleman, & Muysken, 1981). It is therefore that many of the approaches investigating the nature of the Language Faculty focus on the structure and processes of the computational system of grammar.

Despite the importance placed on the computational system of grammar, the approaches to examine it are marked by disagreement about the necessary or sufficient computations required to create the expressed languages of the world. Traditionally, for example in the 'principles and parameters' approach to grammar (Chomsky, 1981; 1986), the structural regularities related to signaling the grammatical relations within a sentence, i.e. those expressing the relations of 'who did what to whom', emerge from the computational system of grammar, i.e. syntax (Chomsky, 1957; Jackendoff, 1997; Pinker & Jackendoff, 2005). The Principles are linguistic universals that are common to all natural languages and are part of the child's native endowment. Parameters are options that allow for variation in linguistic structure and are set upon the child's exposure to linguistic input. One proposed principle is that phrase structure must consist of a head (e.g. a noun or a verb) and a complement (a phrase of specific types). However, the order of head and complement is not fixed: languages such as English have a "head-initial" structure (e.g. the verb phrase "catch fish") and languages such as Japanese have a "head-final" structure (e.g. "fish catch). Thus, the head-directionality parameter (initial/final) must be set through the child's exposure to linguistic input.

Representational approaches to grammar, i.e., the 'principles and parameters' theory, have been criticized for not being phylogentically and ontogenetically plausible (Chomsky, 1995). It was initially thought that the structural similarities and differences among world languages could adequately be explained by a handful of principles and parameters (Chomsky, 1981; 1986). However, subsequent research has complicated this simple view considerably (Chomsky, 1995). The necessity of additional parameters has two important consequences. On the one hand, the resulting complexity of syntax is difficult to explain in terms of natural selection (Hauser,

Azadpour, & Mehler, 2008 for a discussion about the appropriateness of the A^nB^n grammars). Importantly, although songbirds can combine different notes into a variety of songs, they don't integrate this combinatorial capacity with conceptual abilities to create sounds with varied meaning (Hauser & Bever, 2008).

Chomsky, & Fitch, 2002). On the other hand, while there are concrete proposals for setting the most important parameters, such as the head direction parameter, responsible for word order (e.g. Nespor, Shukla, Vijver, Avesani, Schraudolf, & Donati, 2008; Gervain, Nespor, Mazuka, Horie, & Mehler, 2008), there are many others for which no mechanisms have been found (Chomsky, 1995).

Faced with such shortcomings there are some researchers who believe that the structure of the computational system is much simpler and not all the structural diversity must be represented in the computational system of grammar (Chomsky, 1995). According to this view, the computational system of grammar is limited to a single syntactic (specifier-head-complement) structure (Kayne, 1994; 2004; Moro, 2000) and all the surface variation observed among the world's languages is derived from a handful of operations that map conceptual knowledge to sensory-motor programs (Chomsky, 1995; Hauser, Chomsky, & Fitch, 2002). These operations are thought to minimally include: computational devices such as hierarchies and dependencies among syntactic categories (e.g., the relationship between determiners such as "the" and "a" followed by nouns; the relationships between nouns and verbs); recursive and combinatorial operations (e.g. embedding phrases into phrases); and movement of parts of speech and phrases (e.g., to create a question, many languages move constructions such as "what" or "where" to the front of the sentence) (c.f. Chomsky, 1995). Thus, in comparison to the earlier views of the Human Faculty of Language that were largely representational (e.g. 'principles and parameters'), this approach emphasizes the derivational processes necessary for generating the structural diversity among world's languages. This effectively means that everything, except this restricted set of syntactic structures, is generated outside the computational system of grammar (Hauser, Chomsky, & Fitch, 2002).

The attempt to restrict the necessary computations and syntactic structures that have to be included in the computational system of grammar has been criticized for relegating many of the syntactic structures that have been the focus of the linguistic inquiry (e.g., subjacency, Wh-movement, the existence of garden-path sentences, morphology) to the periphery of the Language Faculty (c.f. Pinker & Jackendoff, 2005). In theory, the fact that linguistic structures emerge from outside the computational system of grammar need not mean that they are less important for human language than those structures and computations included in the core computations of syntax (Fitch, Hauser, & Chomsky, 2005). The real problem with the theory lies elsewhere: namely, if the computational system of grammar only has one underlying structure, how and where do all the alternative structural configurations emerge from. There is no empirical evidence that has unveiled the neuro-biological or cognitive basis of non-default grammatical configurations, neither have there been any concrete proposals for motivating either the ontogenetic or the phylogenetic emergence of the alternative grammatical configurations. In other words, there appears to be no reason for languages to differ structurally in the first place.

1.2 How does structural diversity emerge?

In order to understand why languages differ, the first part of this thesis focuses on the possibility that some of the structural regularities we observe among the world's languages may emerge from outside the computational system of grammar. The study investigates the cognitive bases of the two most common word orders in the world's languages: SOV (Subject–Object–Verb) and SVO (Dryer, 2005).

One the one hand, there is evidence that the computational system of grammar prefers SVO. For example, word order change is unidirectional from SOV to SVO (cf. Newmeyer, 2000), the SVO order emerges when children grammaticalize inconsistent linguistic input (Bickerton, 1981; Kouwenberg, 1994), SVO-languages appear to be syntactically most consistent (Steele, 1978); and theoretical arguments in syntax suggest SVO as the universal structure for computational system of grammar (Kayne, 1994). On the other hand, the reason for the prominence of SOV languages is not as clear. It is known, however, that deaf children born to hearing parents organize their spontaneous gestures (referred to as 'homesigns') in the OV order (Goldin-Meadow & Mylander, 1998); new sign languages that have emerged from homesign rely on the SOV order (Senghas, Coppola, Newport, & Supalla, 1997; Sandler, Meir, Padden, & Aronoff, 2005); and even normally hearing adults, who have to gesture instead of using their native language, produce gestures in the SOV order even if their native language is SVO (Goldin-Meadow, So, Ozyurek, & Mylander, 2008). The consistent neglect of native syntax in the improvised gesture systems of normally hearing adults suggests that the SOV order may emerge from outside the computational system of grammar.

The first part of the thesis consists of two gesture-production experiments and one gesture comprehension experiment (1,2 and 3) that show that SOV emerges as the preferred constituent configuration in participants whose native languages have orthogonal word orders (Italian: SVO; Turkish: SOV). This means that improvised communication does not rely on the computational system of grammar. The results of a fourth experiment, where participants comprehended strings of prosodically flat words in their native language, shows that the computational system of grammar prefers the orthogonal Verb–Object orders. The experiments show that linguistic structures can be generated outside the computational systems of grammar. This means that grammatical diversity may emerge without imposing complex data structures on the computational system of grammar: structural differences among languages may be the direct cause of a struggle among the individual cognitive systems trying to impose their preferred structures on human communication.

1.3 Further preferences of the computational system of grammar

While the first study showed that there are specific preferences for word order in the computational system of grammar, it is important to note that word order is not the only grammatical device that a language can utilize. The grammatical repertoire available to the Human Language Faculty has to include phrase structure, recursion, word order and morphological marking of case and agreement (Pinker & Jackendoff, 2005). Comparisons between word order and morphology are particularly interesting for investigating the preferences of the computational system of grammar because the two grammatical devices can, in theory, be equally effectively used to signal ‘who did what to whom’.

Despite the fact that word order and morphology may be used to accomplish exactly the same task, there are considerable differences between the two devices. For example, comparisons between different languages show that there are many more languages that rely on word order rather than morphology as a primary grammatical device (Dryer, 2005); languages that were thought to be non-configurational have been shown to have an underlying word order (Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009); word order and morphology are served by different neural subsystems (Newman, Supalla, Hauser, Newport, & Bavelier, 2010); and studies in language acquisition show that word order is acquired earlier than morphology (e.g. Hakuta, 1977; Slobin & Bever, 1982; Nagata, 1981). Taken together, these observations appear to indicate that there is a clear preference for word order over morphological marking.

The second part of this thesis consists of four cross-situational artificial grammar-learning experiments. In these experiments Italian and Japanese speaking participants were instructed to learn either a 'morphology rule' or a 'word order rule' by rapidly calculating cross-situational statistics for mapping the content of simple drawn vignettes to artificially synthesized nonsense sentences. While both linguistic groups readily learned the word order rule, they failed to perform above chance on the morphology rule. In fact, participants only learned morphology when they could rely on a fixed order. The results of the four experiments, where word order and morphological marking were rendered computationally comparable, show that word order is considerably easier to learn than morphology. Importantly, the results of the four experiments suggest that morphological marking as a grammatical device may emerge in language acquisition only after the language learners have acquired the basic word order. Because the experiments required participants to rapidly compute statistical relations on many different levels, these findings may mean that computationally the human mind is more adapted to processing word order than it is to process morphology.

1.4 Some implications for language acquisition

The idea that the computational system of grammar does not define all the grammatical diversity among the world languages has consequences also for language acquisition. Because the grammatical diversity is no longer defined by Principles and Parameters in the computational system of grammar (Chomsky, 1980), it has been implicitly assumed that languages cannot be acquired by parameter setting (Chomsky, 1995; Hauser, Chomsky, & Fitch, 2002). Instead, there is a strong trend to see language acquisition in terms of a combination of statistical computations (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), algebraic rule generalizations (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) and simple perceptual biases (c.f. Endress, Nespor, & Mehler, 2009). Recent studies go as far as to suggest that infants may use transitional probabilities between syllables (Saffran & Wilson, 2003) and identity relations between syllables (Kovács, & Endress, under review) to extract multi-level structural relations from continuous speech within the first year of life. Somewhat surprisingly, prosody has in this context received very little attention.

The third part of this thesis emphasizes the fact that the speech signal varies in duration, intensity and pitch (Lehiste 1970) and that the variation of these acoustic cues is systematically correlated to the hierarchical structure of syntax (Selkirk, 1984; Nespor & Vogel, 1986). Listeners rely on prosodic cues to segment continuous speech. However, they have also been found to group syllables according to prosodic cues: i.e. syllables that differ only in duration are grouped with the longest syllable in final position, and sequences of syllables that differ only in pitch are grouped with the higher-pitched syllable in initial position (e.g. Bion, Benavides, & Nespor, in press). This suggests that participants can use prosody also for finding relations between segmented units. By drawing a difference between the processes of segmentation and grouping, it is possible to see grouping as an effective mechanism for discovering hierarchical relations from continuous speech.

A series of three experiments show that participant use pitch declination to group sequences of six adjacent syllables ("sentences"), while simultaneously relying on final lengthening to further segment the input into trisyllabic units ("phrases"). Moreover, participants generalized "grammar-like" rules from the segmented units both on phrase- as well as on sentence-levels, a feat observed in previous studies on a single structural level only when the trisyllabic sequences were separated by silence (Peña, Bonatti, Nespor & Mehler, 2002). While there is no one-to-one mapping between syntax and prosody, the findings of this thesis suggest that language learners perceive prosodic cues (pitch and duration) to be organized hierarchically, use these cues to segment continuous speech, discover the hierarchical relations between the segmented units, and generalize structural regularities signaled by the prosodic constituency.

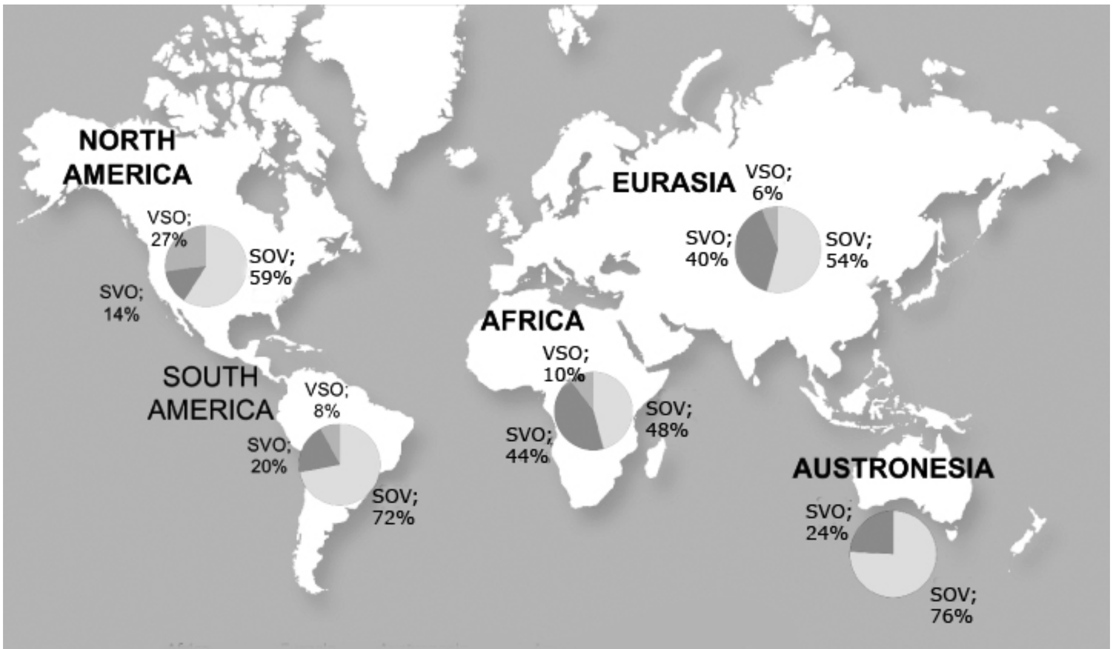
Chapter 2

How does structural diversity emerge?

2.1 Introduction

The first part of this thesis investigates the cognitive bases of the two most common word orders found in the languages of the world: Subject–Object–Verb (SOV) and Subject–Verb–Object (SVO). The way languages change over long periods of time, the analysis of the syntactic structures attested in the world's languages, the relative stability in word order, and how new languages, known as Creoles, emerge in situations of atypical language acquisition, suggest a syntactic preference for SVO. There is, however no parallel evidence that would account for the existence of at least as many SOV languages. It is proposed that it is possible to dissociate communication from grammar and hypothesize that the prominence of the SOV order among the world's languages lies in the cognitive mechanisms responsible for prelinguistic communication. The dominance of the two most common word orders in the world's languages would thus result from the struggle between the preferences of the computational system of grammar and the forces that govern prelinguistic communication.

This study takes a new look at word order variation, and argues that the structural diversity observed among the world's languages does not emerge solely from the computational system of grammar but rather from the ways in which the computational system of grammar interacts with the sensory-motor and conceptual systems. There are six logically possible ways of arranging words in sentences according to their basic grammatical functions of Subject, Object and Verb (OSV, OVS, SOV, SVO, VOS, VSO). Of these six, SVO and SOV characterize the basic orders of the great majority of the world's languages (76%: SOV% and SVO%) (for a detailed account of word order distribution among the world's languages see Figure 2.1) (Dryer, 1989; 2005). There is thus a clear preference for word orders where the Subject precedes the Object and where the Verb–Object constituent is preserved (as in both S–OV and S–VO) (Greenberg, 1963; Greenberg, 1978).



	Africa	Eurasia	Austronesia	N.America	S.America
SOV	**22**	**26**	**19**	**26**	**18**
SVO	21	19	6	6	5
VSO	5	3	0	12	2

Figure 2.1: Word order distribution among the world's languages. The table shows the count of different orders among the language families in the 5 large linguistic areas. The data adapted from Dryer (1989) is corrected for influences from language contact. The map shows the percentage of SOV and SVO among the languages in the 5 large linguistic areas.

There are, however, reasons to believe that SVO is the preferred structure for syntax. Studies in historical and comparative linguistics suggest that the computational system of grammar has one single preferred word order. Indeed, analyzes of how languages change over time, report that, when word order changes independently of language contact, it is unidirectional from SOV to SVO (cf. Newmeyer, 2000).[3] Thus, indirectly, historical and comparative linguistics suggest that when the word order of a language changes for language internal reasons, the computational system of grammar drives it towards the SVO order. Interestingly, SVO languages are more stable in word order than other languages, that usually have several additional alternative orderings of Subject, Verb and Object (Steele, 1978). This stability of SVO languages is an additional reason to consider SVO syntactically preferred.

Though there are alternative proposals (cf. Haider, 2000 based on German), convergent evidence in theoretical linguistics points to the universality of SVO as the basic word order for the computational system of grammar (Chomsky, 1995). It has been argued that there is a universal underlying structure, from which the surface syntactic forms of all languages are derived (2004; Kayne, 1994; Moro, 2000). In this basic structure, the heads of phrases universally precede their associated

[3] Word order change has extensively been studied in the Indo-European language family. Romance languages, such as French, have changed from SOV to SVO (Bauer, 1995), the same is true for Germanic languages such as English (Kiparsky, 1996), Swedish (Holmberg & Platzack, 1995) and Icelandic (Hróarsdóttir, 2000), as well as Slavic languages such as Russian (Leinonen, 1980). Also the Indo-European ancestor languages, such as Sanskrit (Staal, 1967) and Ancient Greek (Taylor, 1994) were SOV, suggesting that Proto-Indo-European had the SOV order. While the evidence for the unidirectional change from SOV to SVO is largely based on research on Indo-European languages, languages such as Finnish (Leinonen, 1980) and Austronesian languages such as Seediq (Aldridge, in press), that are not part of the Indo-European family, have undergone the same direction of change. In fact, if we look at the ancestor languages in Dryer (2005), it is clear that the SOV order was the dominant configuration among most of the sampled languages. Unfortunately, descriptions of historical change are only available for a restricted (though substantial) number of linguistic families. Of these, the only putative exception to the unidirectional change from SOV to SVO that occurs independently of language contact concerns Mandarin Chinese (Li, 1977). Li and Thomson (1974) have suggested that Mandarin has been undergoing a change from SVO to SOV through grammaticalization of serial-verb constructions. However, in contemporary Mandarin the SOV constructions are heavily marked and VO constructions vastly outnumber OV constructions, showing that SVO is the basic word order (Li, 1990). In fact, Sun and Givón (1985) have shown on the basis of written and spoken analyses that in contemporary Mandarin OV constructions only appear in about 10% of the cases. They further argue that there is no evidence either from their corpus or from the acquisition of Mandarin by native children that would suggest a drift toward the SOV order. Light (1979) comes largely to the same conclusion.

complements: for example, verbs precede their objects, prepositions precede nouns and main clauses precede subordinate clauses. In addition, specifiers – syntactic categories that specify the heads, as for instance 'some' in some apples – universally precede the head they are associated with. This specifier-head-complement configuration corresponds to the SVO order (Chomsky, 1995).

Interestingly, there are reasons to believe that the SOV order predominant in the world's languages is not particularly well suited for syntactic computations, whose task is to unambiguously map meaning to sound or signs (Hawkins, 1994). For example, verbs define the arguments they take, i.e. when one hears a verb like give, one is primed to expect two internal arguments pertaining to the object to be received, and to the recipient. However, in SOV languages the complements precede the heads – thus both direct and indirect object precede the verb. Given that the role of the arguments is defined by the verb, it is useful to use extra cues, i.e. morphological marking of case, to identify their semantic roles (Hawkins, 1994; Newmeyer, 2000). This suggests that the SOV order should be dispreferred by the computational system of grammar.

If we accept that the computational system of grammar has one single underlying order, it remains to be explained why grammatical diversity emerges in the first place – in particular, which cognitive mechanisms are responsible for the origin of the SOV word order, and why the SOV and SVO orders are equally prominent among the world languages. Hauser et al. (2002), argue that everything but the underlying SVO order is generated outside the computational system of grammar. However, neither theoretical proposals nor experimental evidence have clarified why and where in the language faculty the alternative configurations emerge.

Interestingly, a dichotomy between SVO and SOV has been found in two specific cases of atypical language acquisition. Creoles, new fully-fledged languages that arise in communities where children are exposed to a pidgin – a rudimentary jargon created by people who must communicate without sharing a native language (Bickerton, 1981) – use this jargon to develop a systematic SVO order in a single generation (Bickerton, 1984). In contrast, homesigners – deaf children not exposed to any sign language – create their own gestural vocabulary (Goldin-Meadow & Feldman, 1977) and use the Object–Verb order,[4] which parallels the SOV order, in

[4] It is generally agreed that it is more appropriate to describe the gesture systems of

their gestural expressions (Goldin-Meadow, 2005; Goldin-Meadow & Mylander, 1998).

While Creoles are syntactically fully-fledged languages (Bickerton, 1984; Muysken, 1988), the nature of the gesture systems of homesigning children is less clear. The SOV order that is dominant in homesign is also attested in the gestural utterances produced by normally hearing English (SVO), Chinese (SVO), Spanish (SVO) and Turkish (SOV) speaking adults, instructed to use only gestures to describe simple scenarios (Goldin-Meadow, So, Ozyürek, & Mylander 2008). The structural similarities between the gesture systems of homesigning children and the improvised gestures of normally hearing adults suggest a strong predisposition for the SOV order in simple improvised communication (Goldin-Meadow et al., 2008).

This study expands the hypothesis that SOV characterizes improvised communication[5] and suggests that the SOV order in gestures is prelinguistic in nature because it results from a direct interaction between the sensory-motor and the conceptual systems. Unlike in language where the mapping between signal and meaning has to necessarily be mediated by syntax, in improvised gestural communication the mapping between the signal (the gestures) and its meaning may be achieved without the intervening syntactic computations responsible for phrase structure. Several studies with adult speakers learning a new language show that they do not abandon their native grammar (Odlin, 1989). For example, immigrant workers learning Dutch – a language with SOV order in subordinate clauses and SVO order in main clauses – tend to use the SVO order when their native language is Moroccan Arabic (SVO), and the SOV order when their native language is Turkish (SOV) (Jansen, Lalleman, & Muysken, 1981). The fact that normally hearing English (SVO), Chinese (SVO) and Spanish (SVO) speaking adults in Goldin-Meadow et al. (2008) produced gesture strings in the SOV order and failed to transfer their native SVO

homesigners, as well as normally hearing adults asked to gesture, in terms of semantic roles (e.g. Actor, Patient, Action) rather than grammatical roles (e.g. Subject, Object, Verb). However, because the present paper directly compares word order in spoken language to the gesture order of normally hearing adults, and because in our experiments the semantic roles of words unanimously correspond to the same grammatical roles (e.g. the Actor is always the Subject, the Patient the Object, and the Action the Verb), for the sake of clarity we will use the terms of Subject, Object and Verb.

[5] The term 'improvised communication' is used throughout the thesis because in the experiments participants use pantomime-like gestures that they must create on the spot without any prior experience. Because the majority of the gestures participants used were imitations of the real world objects and actions, the resulting communication code is essentially an iconic one.

order to gestures, suggests that they bypassed their native linguistic structures. This may mean that it is possible to communicate simple events in a prelinguistic way, i.e. without relying on the computational system of grammar, a necessary ingredient of language.

The picture so far thus suggests that there is a general faculty of language that includes the sensory-motor system, the conceptual system and the computational system of grammar. The world languages emerge from the interaction of these three systems only when the computational system of grammar links meaning (the conceptual system) to sounds or signs (the sensory-motor system). However, language-like structures also emerge in improvised gestural communication that does not appear to rely on the computational system of grammar. Thus, these structures offer evidence that the different word orders observed in the world's languages are not uniquely defined by the computational system of grammar – were it so, we would expect the grammatical structures to exhibit much less variation than is attested among the world's languages.

In order to investigate whether the structural regularities in improvised gestures are grammatical in nature, the following experiments were carried out with normally hearing Italian and Turkish-speaking adult participants, whose native languages use different word orders, SVO for the former and SOV for the latter. Experiment 1 tested whether normally hearing Italian and Turkish-speaking adults introduce the structural regularities of their native grammars into their gesture strings. The intention was to replicate the results of Goldin-Meadow et al. (2008), though with a set of stimuli that could be systematically modified, in subsequent experiments, as to their complexity. Experiment 2 used more complex stimuli in order to investigate whether the structural regularities in improvised gestures rely on the computational system of grammar, that is, whether there is evidence for phrase structure. Experiment 3 investigated whether the preferences found in gesture production emerge also in gesture comprehension. Experiment 4 investigated the preferences of the computational system of grammar by testing the order preferences for prosodically flat sequences of words in participants' native language.

2.2 Experiment 1: Gestural descriptions of simple scenarios

Normally hearing adult speakers of English (SVO), Turkish (SOV), Spanish (SVO) and Chinese (SVO) asked to gesture instead of using their native language, have been found to order their gestures in the SOV order (Goldin-Meadow et al., 2008). In Experiment 1, tried to establish whether we find the same gesture regularities (i.e. the SOV order) with stimuli that can be systematically manipulated in complexity for subsequent experiments that can disentangle the cognitive origin of the SOV order. Thus, native speakers of Italian and Turkish were asked to describe simple scenarios depicted on drawn vignettes by using either only gestures or their native language. Italian (SVO) and Turkish (SOV) speaking adults were chosen because their native languages use orthogonal word orders. The results of Goldin-Meadow et al.'s (2008) predict that Italian and Turkish-speaking adults structure their gesture strings identically in the SOV order. Any deviance from the participants' native order would suggest that the structural regularities in gestural communication are independent of participants' native syntax.

2.2.1. Participants

Twenty-eight Italian native-speaking volunteers (15 females, 13 males, mean age 23.8, range 19– 27 years) recruited from the subject pool of the International School of Advanced Studies in Trieste (Italy) and 28 Turkish native-speaking volunteers (14 females, 14 males, mean age 21.4, range 19–24 years) recruited from the subject pool of the Boğaziçi University in Istanbul (Turkey). Participants reported no auditory or language related problems and did not know any sign language. Participants received a monetary compensation.

2.2.2. Stimuli

The stimuli of Experiment 1 consisted of 32 simple drawn vignettes that depicted someone doing something to someone or something else (e.g., a girl catches a fish) (for the full list of vignettes see Appendix A1). In all the vignettes, each of the three constituents unambiguously matched the category of the Subject, the Object or the Verb (e.g., the fish cannot catch the girl). In order to avoid possible frequency biases induced by different occurrences of individual constituents, in this and subsequent experiments, the depicted scenarios consisted of four different Subjects, Objects and Verbs that were distributed across the vignettes in a combinatorial manner. All constituents were thus equally frequent (N = 8) and participants saw them during the experiment in different combinations with other constituents an equal number of times. In order to avoid possible biases induced by certain constituents appearing either on the left or the right side of the vignettes, we created mirror images of each vignette and counterbalanced their appearance across participants.

2.2.3. Procedure

Participants were presented with the vignettes one by one in random order on a computer screen. After seeing each vignette, half of the participants in each linguistic group were instructed to describe it as clearly as possibly by using only gestures. Participants were asked not to speak. Participants were allowed to take as much time for describing each vignette as they thought it was necessary and to proceed to the following vignette when they thought they had accomplished the task. Participant's responses were videotaped and consequently coded for the order of individual gestures by two independent coders. The other half of the participants in each linguistic group was asked to describe the vignettes in their native language (Italian or Turkish). Their responses were audio recorded and coded for the order of words by two independent coders.

2.2.4 Results

Participants' responses were coded by two independent coders who had to determine the order of the gestures in participants' descriptions of the vignettes. Because we were interested in the order in which participants organized their gestures, rather than in how well and clearly they could gesture individual constituents, the coders could rely on the vignettes to determine the grammatical role of the gestures. This resulted in a confidence rating of 93% for coder 1 and 91% for coder 2 (the responses coded as ‘uncertain’ were eliminated from the analysis). Because participants sometimes made repeated attempts to gesture scenarios, the coders were asked to analyze the gesture-string that was produced last and ignore the failed attempts. Because participants sometimes described a scenario with several 2-gesture strings rather than with one 3-gesture string, the coders were asked to analyze the 2-gesture strings separately. The agreement of the two raters’ observations on coding gestures was measured with Cohen’s kappa coefficient, which resulted in a kappa value of 0.79 for Turkish (substantial agreement) and 0.83 for Italian (perfect agreement).

The gestures participants produced for describing the scenarios were always iconic, meaning that they figuratively imitated the form of the objects/persons or the movement of the limbs required to produce the actions. Because both objects and actions occurred several times in different vignettes, it was possible that participants could use the same gestures for the same objects and actions. The analysis of participants’ responses shows that they reused the gestures for both objects and actions by producing the same gestures as in previous occurrences of the same constituent on average of 78.44% of the cases (*SD* = 8.4).

While participants were instructed to describe vignettes with 3-gestures, when we look at the gesture strings of Italian as well as Turkish-speaking adults, we see that the gesture strings contained either two or three constituents. Participants thus sometimes omitted constituents and described a scenario with two 2-gesture strings. The gesture strings of Italian-speaking participants contained all three constituents on average in 58.6% (*SD* = 12.4) of the cases. For Turkish-speaking participants the three-constituent gesture strings made up on average 63.2% (*SD* = 10.4) of all the gesture strings.

In 2-gesture strings, Italian-speaking participants always gestured the Verb, and, additionally, gestured the Subject on average of 42.3% (*SD* = 10.3) and the Object on average of 57.7% of the cases. Similarly, Turkish-speaking participants always gestured the Verb, but gestured the Subject on average of 45.8% and the Object on average of 54.2% (*SD* = 9.5) of the cases. An ANOVA with two fixed factors (Constituent omission: Subject vs. Object omission) and (Participants’ native language: Turkish vs. Italian) showed a main effect for constituent omission (*F*(1, 26)

= 12.455, *P* = .032), but neither interaction with native language (*F*(1, 26) = 20.233, *P* = .211), nor a main effect of native language (*F*(1, 26) = 10.167, *P* = .097). This shows that participants omitted gestures Subjects more than for Object regardless of their native language.

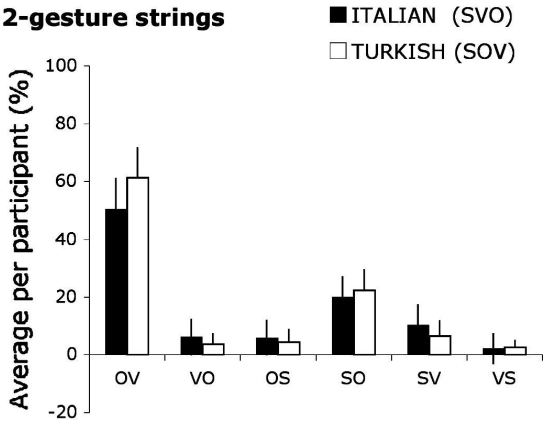


Figure 2.2 Italian and Turkish speakers' 2-gesture strings for describing simple scenarios: distribution of constituent orders for Subject, Object and Verb.

To see whether the 2-gesture strings were consistently organized within and across linguistic groups, a ANOVA with one within-subjects factor (gesture order: SV, VS, OV, VO, SO, and OS) and one between-subjects factor (participants' native language: Turkish vs. Italian) was carried out. For the distribution of the constituents in 2-gesture strings, see Figure 2.2 There was a main effect for gesture order (*F*(5, 26) = 83.586, *P* < .0001) but no interaction with participants' native language (*F*(1, 26) = 90.456, *P* = .867). Pair-wise Bonferroni-corrected comparisons show that Italian speakers were more likely to gesture Objects before, rather than after, Verbs (*P* < .0001); Subjects before Objects (*P* < .0001) and Subjects before Verbs (*P* = .032). The same tendency emerged for Turkish-speaking participants, who gestured Objects before Verbs (*P* < .0001); Subjects before Objects (*P* < .0001); and Subjects before Verbs (*P* = .023). Both Italian and Turkish-speaking participants were more likely to gesture Object–Verb than Subject–Verb (*P* < .0001) or Subject–Object (*P* < .0001) in their 2-gesture strings.

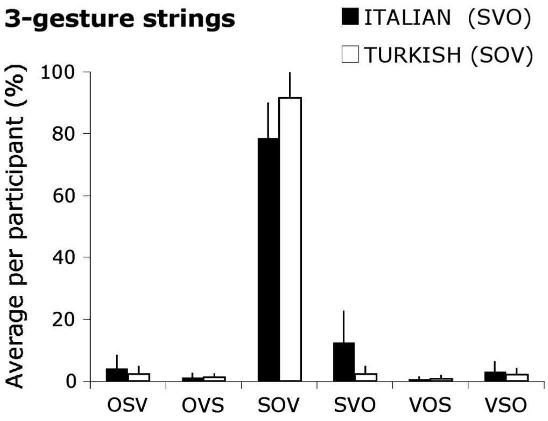


Figure 2.3 Italian and Turkish speakers' 3-gesture strings for describing simple scenarios: distribution of constituent orders for Subject, Object and Verb.

Among the 3-gesture strings, the most dominant order was Subject–Object–Verb (SOV) both for Italian (77.6%; *SD* = 8.9) as well as for Turkish (89.4%; *SD* = 10.6%) speaking adults (see Figure 2.3). In order to determine whether this ordering of constituents in 3-gesture strings was consistent between the two groups, an ANOVA with one within-subjects factor (gesture order: OSV, OVS, SOV, SVO, VOS and VSO) and one between-subjects factor (participants' native language: Turkish vs. Italian) was carried out. There was a main effect of gesture order ($F(5, 26) = 140.634$, $P < 0.0001$), but again no interaction with participants' native language ($F(1, 26) = 17.409$, $P < 0.543$), and no main effect of native language ($F(1, 26) = 33.232$, $P < 0.522$). Pair-wise Bonferroni-corrected comparisons show that Turkish speakers ordered their gesture strings predominantly in the SOV order ($P < .0001$). The SOV order was also the most dominant one for Italian participants ($P < .001$).

In order to determine whether participants were bypassing their native grammars, it was important to compare the order of constituents in participants' verbal descriptions with the constituent orders found in participants' gestural descriptions of the same vignettes. The participants of both linguistic groups, when asked to use their native language, always described the simple scenarios with sentences that contained the Subject, the Object and the Verb. Without exceptions, Italian speakers' spoken sentences were in the SVO and Turkish speakers' sentences in the SOV order, that is, for all participants, in the basic order of their native language.

To compare the 3-gesture strings to the verbal descriptions of the vignettes an ANOVA with one dependent variable (percentage of SOV) and two fixed factors (modality: speech vs. gestures; and participants' native language: Turkish vs. Italian) was performed. There was a main effect for modality (*F*(1, 52) = 128.834, *P* < .0001) and native language (*F*(1, 52) = 98.234, *P* < .0001) and an interaction between modality and native language (*F*(1, 52) = 90.233, *P* < .0001). Bonferroni-corrected post-hoc tests (*P* < .05) show that Italian participants 3-gesture strings had significantly more SOV order than their verbal descriptions (*P* < .0001). The differences between Turkish speaking participants' 3-gesture utterances and their verbal descriptions failed to reach significance (*P* < .309). Italian and Turkish-speaking participants verbal descriptions differed significantly (*P* < .0001), but the differences between their gestural utterances failed to reach significance (*P* = 0.655). This shows that at least Italian-speaking participants were bypassing their native grammar.

2.2.5 Discussion

These results show that, while in the speech test, Italian and Turkish participants used orthogonal word orders, both Italian and Turkish speakers produced gesture strings predominantly in the SOV order. Because Italian (SVO) and Turkish (SOV) have orthogonal word orders, the results of Experiment 1 results, like those of Goldin-Meadow et al. (2008), show that when asked to gesture, speakers of different languages introduce the same SOV order into their gesture strings. Because the SOV order is ungrammatical in SVO languages like Italian (Experiment 1) as well as in English, Chinese and Spanish (Goldin-Meadow et al., 2008), it has been suggested that SOV is a natural order – possibly semantic in origin – for describing simple events (Gershkoff-Stowe & Goldin-Meadow, 2002; Goldin-Meadow et al., 2008). Furthermore, it has been suggested that the fixed order of gestures may represent the seed of grammar, since also homesigning children introduce the Object–Verb order into their gesture strings (Goldin-Meadow, 2005; Goldin-Meadow & Mylander, 1998), and since new sign languages that emerged from homesigners in Nicaragua (Senghas, Coppola, Newport, & Supalla, 1997) and Israel (Sandler, Meir, Padden, & Aronoff, 2005) also appear to be organized in the SOV order. This interpretation has some plausibility, since the SOV order in gestures is indistinguishable from the canonical SOV order of simple clauses in Turkish.

However, a second interpretation of these results, as well as those of Goldin-Meadow et al.'s (2008), is also possible. When gesturing, participants must use the sensory-motor system for executing the physical gestures and rely on the conceptual knowledge stored in the conceptual system to convey the meaning of the vignettes with individual gestures. However, there is no reason to believe that the computational system of grammar is necessarily involved in producing these simple gesture strings. For example, on the basis of simple gesture-strings it is impossible to determine whether the gestural utterances have any internal language-like hierarchical organization of constituents such as specifiers, heads and complements. To decide between these two interpretations regarding the origin of the SOV order in gestures – whether or not it is grammatical in nature – a second experiment was carried out.

2.3 Experiment 2: Gestural descriptions of complex scenarios

One possible way to investigate whether the computational system of grammar is used when normally hearing adults are asked to gesture, is to increase the complexity of the scenarios that the participants are asked to describe. Thus, speakers of Italian and Turkish were asked to describe more complex scenarios depicted on drawn vignettes by using either their native language or only gestures. In natural language, the complex vignettes we used, would be described with complex sentences containing a main clause and an embedded clause (as in English [the man tells the child [that the girl catches a fish]]).

If the SOV order that emerged in the description of simple scenarios (Experiment 1) is grammatical for Turkish-speaking adults, it should also extend to more complex SOV like structures typical of Turkish (SOV). Participants should thus gesture the subordinate clauses in the same position as the Object of simple clauses, i.e. before the Verb of the main clauses, as in Turkish [Adam çocuga [kızın balık yakaladıgını] anlatır] (equivalent in English to [man child-to [girl fish catches] tells]). Furthermore, if gestural communication triggers SOV language-like constructions in the computational system of grammar, we would expect also Italian-speaking participants to produce complex gesture strings that follow the SOV language-like structures that are typical of Turkish, as well as other SOV languages.

2.3.1. Participants

Twenty-eight Italian native-speaking volunteers (14 females, 14 males, mean age 25.6, range 20–29 years) recruited from the subject pool of the International School of Advanced Studies in Trieste (Italy) and 28 Turkish native-speaking volunteers (16 females, 12 males, mean age 20.1, range 18–23 years) recruited from the subject pool of the Boğaziçi University in Istanbul (Turkey). Participants reported no auditory or language related problems, did not know any sign language and had not participated in Experiment 1. Participants received a monetary compensation.

2.3.2. Stimuli

The stimuli of Experiment 2 consisted of drawn vignettes more complex than those of Experiment 1. The 32 complex vignettes were created by randomly embedding the 32 simple drawings from Experiment 1 in a speech bubble in eight different scenarios (for an example of how the complex vignettes were created and a full list of the complex frames see Appendix A2).[6] In natural languages, vignettes like these would be described with complex sentences containing a main clause (e.g. the man tells the child) and an embedded clause (e.g. that the girl catches a fish), as in English the man tells the child that the girl catches a fish. In order to avoid possible biases induced by certain constituents appearing either on the left or the right side of the vignettes, we created mirror images of each vignette and counterbalanced their appearance across participants.

2.3.3. Procedure

The procedure of Experiment 2 was identical to that used in Experiment 1.

[6] Some of the verbs of the main clauses required a direct object and some did not. However, this is not relevant to this experiment.

Participants were presented with the vignettes one by one in random order on a computer screen. After seeing each vignette, half of the participants in each linguistic group were instructed to describe it as clearly as possible by using only gestures. Participants were asked not to speak. Participants were allowed to take as much time for describing each vignette as they thought it was necessary, and to proceed to the following vignette when they thought they had accomplished the task. Participant's responses were videotaped and consequently coded for the order of individual gestures by two independent coders. The other half of the participants in each linguistic group was asked to describe the vignettes in their native language (Italian or Turkish). Their responses were audio recorded and coded for the order of words by two independent coders.

2.3.4 Results

Participants' responses were coded by two independent coders who had to determine the order of the gestures in participants' descriptions of the vignettes. Because we were interested in the order in which participants organized their gestures, rather than in how well and clearly they could gesture individual constituents, the coders could rely on the vignettes to determine the grammatical role of the gestures. This resulted in a confidence rating of 98% for coder 1 and 96% for coder 2 (the responses coded as 'uncertain' were eliminated from the analysis). Because participants sometimes made repeated attempts to gesture scenarios, the coders were asked to analyze the gesture-string that was produced last and ignore the failed attempts. The agreement of the two raters' observations on coding word order was measured with Cohen's kappa coefficient, which resulted in a kappa value of 0.93 for Italian and 0.85 for Turkish (perfect agreement).

As in Experiment 1, the gestures participants produced for describing the complex scenarios in Experiment 2 were always iconic, that is, they figuratively imitated the form of the objects or the movement of the limbs required to produce the actions. Because the objects and actions occurred several times in different vignettes, it was possible that participants could use the same gestures for the same objects and actions. The analysis of participants' responses shows that they reused the gestures for both objects and actions on average of 82.32% of the cases (*SD* = 10.6).

When we look at the gestural descriptions of complex scenarios, it becomes evident that among the responses of the Italian and Turkish-speaking participants, there was not a single gesture string adhering to the syntactic structure typical of SOV languages. Italian speakers gestured the main clause before the subordinate clause in 87.5% (*SD* = 10.2) of the cases. The same order was also evident in Turkish speakers, who gestured the main clause before the subordinate clause in 96.7% (*SD* = 3.3) of the cases. An ANOVA with one within-subject factor (order of clauses: main clause before subordinate clause vs. subordinate clause before main clause) and one between-subjects factor (participants' native language) showed a main effect of order (*F*(5, 26) = 118.345, *P* < .0001), but no interaction with participants' native language (*F*(1, 26) = 34.324, *P* < .534), and no main effect of native language (*F*(1, 26) = 23.564, *P* < .690). This shows that participants gestured the main clauses before the subordinate clauses regardless of their linguistic background.

In order to determine whether participants were bypassing their native grammar, it was important to compare the order of clauses in participants' verbal descriptions to the order of clauses in participants' gestural descriptions of the same vignettes. To discover the most natural native syntactic constructions for describing the complex scenarios, the verbal descriptions of both Italian and Turkish-speaking adults were analyzed first. Without exceptions, in Italian speakers' sentences the main clause preceded the subordinate clause. In contrast, in Turkish speakers' sentences, the subordinate clause always preceded the verb of the main clause.

To compare the gestural descriptions to the verbal descriptions of the complex vignettes an ANOVA with one dependent variable (percentage of main clause followed by subordinate clause gesture strings) and two fixed factors (modality: speech vs. gestures; and participants' native language: Turkish vs. Italian) was performed. There was a main effect for both modality (*F*(1, 50) = 78.435, *P* < .0001) and native language (*F*(1, 50) = 67.564, *P* < .0001), and a significant interaction between modality and native language (*F*(1, 50) = 69.573, *P* < .0001). Bonferroni-corrected post-hoc tests (*P* < .05) show that Turkish-speaking participants' gestured scenarios had significantly more 'subordinate clause following main clause' than their verbal descriptions of complex scenarios (*P* < .0001). The differences between the same 'main clause following subordinate clause' constructions in Italian participants' gestured and verbal descriptions failed to reach significance (*P* < .204). Italian-speaking participants verbal descriptions had significantly more ‘main clause followed by subordinate clause’ constructions than Turkish speakers’ verbal descriptions (*P* < .0001), but the differences between Italian and Turkish gestural descriptions failed to reach significance (*P* = .096). This shows that at least Turkish-speaking participants were bypassing their native grammar.

2.3.5 Discussion

While in the speech task Italian speakers always described complex vignettes with sentences typical of SVO languages and Turkish speakers always described the same vignettes with sentences typical of SOV languages, when gesturing, neither Italian- nor Turkish-speaking adults produced even a single gesture-string that conformed to the structure of complex sentences typical of SOV languages, like Turkish. Turkish-speaking participants failure to gesture the subordinate clause before the verb of the main clause, that was common among the Turkish-speaking participants when using their native language, demonstrates that gestural communication does not follow the grammar of Turkish. Thus Experiments 1 and 2 taken together show that, when gesturing, both Italian and Turkish-speaking adults bypassed their native linguistic structures.

In the computational system of grammar, the majority of SOV languages are syntactically left branching, thus the subordinate clauses usually precede the verb of the main clause, and the majority of SVO languages are syntactically right-branching, thus subordinate clauses usually follow the main clauses (Chomsky, 1957). This is clearly not the case in the results of Experiment 2, where for both Italian and Turkish speakers the main clauses were gestured before the subordinate clauses – a construction typical of SVO but not of SOV languages. This shows that the SOV order in improvised gesturing does not generalize to more complex SOV language-like constructions: it thus does not instantiate the typical linguistic hierarchical organization of constituents. Our results thus indicate that participants were not using the computational system of grammar and that improvised gesture communication is the product of a direct link between the conceptual and the sensory-motor systems.

2.4 Experiment 3: Gesture comprehension

If gesturing does not utilize the computational system of grammar and relies instead on a direct link between the conceptual and the sensory-motor systems, the preference for the SOV order should not only prevail in gesture production, but also be observable in gesture comprehension. Experiment 3 therefore investigated the gesture order preferences in comprehension by using the same simple scenarios that participants described in Experiment 1.

2.4.1 Participants

Thirty-six Italian native-speaking volunteers (18 females, 18 males, mean age 21.2, range 18–28 years) recruited from the subject pool of the International School of Advanced Studies in Trieste (Italy) and 36 Turkish native-speaking volunteers (20 females, 16 males, ages 19–22) recruited from the subject pool of the Boğaziçi University in Istanbul (Turkey). Participants reported no auditory or language related problems, did not know any sign language and had not participated in Experiments 1 and 2. Participants received a monetary compensation.

2.4.2 Stimuli

The stimuli of Experiment 3 consisted of the 32 simple vignettes used in Experiment 1 (see Appendix A1), and 32 video clips where a person described each of these vignettes by using only gestures. Like in Experiments 1 and 2, each vignette was counterbalanced with its mirror image across participants. In order to determine whether there is a preference for a specific constituent order in gesture comprehension, the video clips were constructed digitally in all the possible six orders of Subject, Object and Verb (SOV, SVO, OSV, OVS, VSO and VOS). To avoid possible biases for gesture order introduced by the gesturer, she was asked to produce individual gestures for each of the four Subjects, Objects and Verbs that were depicted on the simple scenarios. The individual gestures were then digitally edited so that they all were equal in length (2000 ms) and then combined into all the logically possible six orders of Subject, Object and Verb (6000 ms). Following this procedure, the video-clips describing the same vignette thus only differed in the order of the constituents (for an example see Video 1 that can be found online at http://dx.doi.org/10.1016/j.cogpsych.2010.01.004).

2.4.3 Procedure

Participants were seated in front of a computer screen. They were first told that they would see video clips of someone describing simple situations with gestures. They were then instructed to choose as quickly as possible, immediately after each gesture clip, between two drawn vignettes (used in Experiment 1), the one that depicted the content of the gesture clip they just saw (dual forced choice task).

In the dual forced choice task one of the vignettes corresponded to the gesture clip and the other one did not by semi-randomly deviating in either one of the three constituents (of Subject, Object or Verb) (for an example see Appendix A3). As the vignettes were created according to a combinatorial design, the distracting vignette of one trial was the correct vignette of another trial. Each participant saw each of the scenarios gestured in the video clip once in each of the six logically possible orders (192 trials). Each participant saw each of the vignettes six times as the correct target and six times as the distracter. Participants had 1500 ms to choose before the next trial began. Reaction Times (RTs) were measured from the onset of the dual forced choice task (i.e. from the moment when the two vignettes appeared on the screen).

To determine whether there is a gesture order preference, all participant saw each of the 32 different scenarios once in each of the six logically possible orders of Subject, Object and Verb. The experiment thus had 192 trials and was divided into six experimental blocks: in each block participants saw each of the 32 scenarios only once in semi-randomly determined gesture order. Between experimental blocks participants could take a break for as long as they wished.

2.4.4 Results

Participants’ responses show that they were not having difficulties with the task, as Italian-speaking participants only failed to give an answer on average in 7.6% (*SD* = 2.3) and Turkish-speaking participants on average in 6.3% (*SD* = 1.5) of the trials. Similarly, the percentage of correct answers was on average 91.1% (*SD* = 8.9) of all the answered trials for Italian- and 94.3% (*SD* = 5.7) of the answered trials for Turkish-speaking adults.

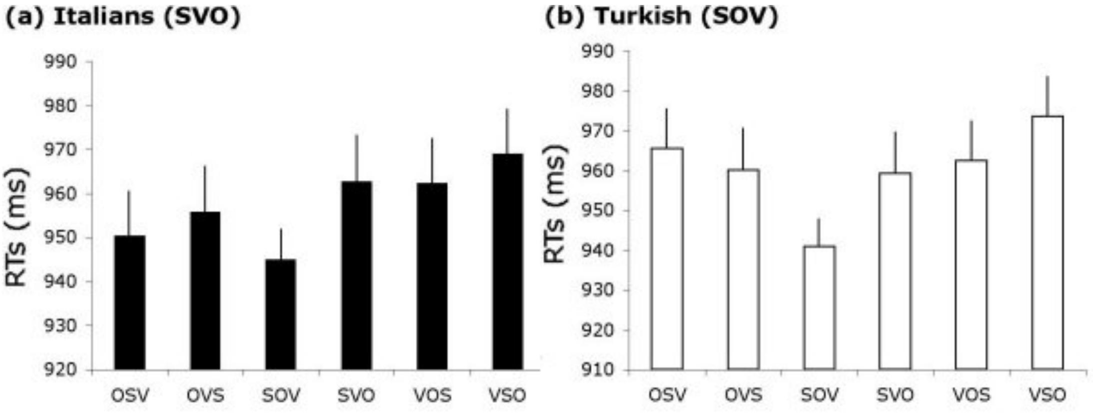


Figure 2.4 Participants’ average Reaction Times in gesture comprehension: (a) Italian-speaking participants and (b) Turkish-speaking participants.

In order to investigate order preferences in gestures, it was first necessary to compare participants’ performance on individual constituent orders. The ANOVA between all constituent orders in Turkish speakers’ responses shows that gesture order influenced their RTs (*F*(5, 31) = 10.23, *P* < .01). Post-hoc tests show that the SOV order elicited fastest RTs for Turkish-speaking adults (Bonferroni-corrected *P* < .04) (see Figure 2.4). Also the ANOVA between all constituent orders in Italian speakers’ responses shows that constituent order influenced their RTs (*F*(5, 30) = 12.7, *P* < .01). Post-hoc tests show that the SOV order elicited the fastest RTs also for Italian-speaking participants (Bonferroni-corrected *P* = .045) (see Figure 2.4). To see whether we find these differences also when taking items, rather than subjects, as random variables, ANOVAs with the vignettes as random variables for Turkish (*F*(5, 30) = 30.48, *P* = .037) and Italian (*F*(5, 30) = 25.21, *P* = .048) participants were carried out. Post-hoc tests show that the SOV order elicited the fastest RTs for both groups in the item based analysis as well (Bonferroni corrected Turkish: *P* = .033;

Italian: $P = .041$). Italian speakers' shorter RTs with SOV than with their native SVO order with gestures suggests that the same preference – non-grammatical in nature – we observe in the production of improvised gestures, also prevails in comprehension.

We observe consistent preferences between Italian and Turkish speaking participants' performance also when we look at all the six logically possible orders. For Italians Object–Verb orders (OSV, OVS, SOV) elicited on average significantly shorter RTs than Verb–Object orders (SVO, VOS, VSO) (2-tailed *t*-test between Object–Verb and Verb–Object orders: $t(35) = 2.969$, $P < 0.01$). Exactly the same preference for Object–Verb orders over Verb–Object orders emerged also in Turkish-speaking participants' RTs (2-tailed *t*-test between Object–Verb and Verb–Object orders: $t(35) = 3.696$, $P < 0.01$) (see Figure 2.6). Because Italian is a Verb–Object order language, the preference for Object–Verb orders in Italian-speaking participants' RTs must be independent of participants' native language.

2.4.5 Discussion

Italians' faster reaction times with SOV than with their native SVO order suggests that the same preference – non-grammatical in nature – we observe in gesture production (Experiment 1), prevails also in gesture comprehension. While participants neglecting their native syntactic structures in gesture production and comprehension is clearly caused by the fact that they had to either produce or interpret gestures instead of sentences in their native language, it is unclear why the SOV order should prevail as the preferred configuration.

One possibility is that we are observing a simple modality effect due to the use of gestures rather than speech, with this particular order of constituents emerging as a by-product of gesturing and with no consequence for the word order distribution in the languages of the world. While the non-native SOV order clearly emerges because participants cannot use their native language to describe the vignettes, there are reasons to believe that this order is not caused by gesturing *per se*. For example, Gershkoff-Stowe and Goldin-Meadow (2002) showed that the same order that prevails in improvised gestures (Agent-Patient-Act) emerges also when participants have to stack together transparencies depicting individual constituents. In this task, where each transparency contained one constituent, participants consistently picked first the transparency with the Agent, followed by the transparency containing the Patient and finally the transparency containing the Act. Furthermore, if the SOV order were particularly well suited for the manual modality, we would expect the sign languages of the world to be predominantly in the SOV order. While detailed word order distributions for sign languages have not been carried out, sign languages do show word order variation just like spoken languages (Klima & Bellugi, 1979), suggesting that the order of constituents is not determined by the manual modality.

Alternatively, it is possible that the SOV order prevails in gestures because gesturing relies on the direct interaction between the sensory-motor and the conceptual system. While there appears to be no particular reason why SOV would be good for the sensory-motor system, SOV might be preferred by the conceptual system. It has been argued that semantic relations (e.g. verbs) require the presence of the entities (e.g. nouns) they link (Gentner & Boroditsky, 2009). The two word orders that satisfy this requirement are SOV and OSV. The latter is, however, extremely rare among world languages because the Object precedes the Subject (Greenberg, 1963; Greenberg, 1978). The SOV order thus satisfies the condition that the entities (the Subject and the Object) precede the relations (the Verb) in the most optimal way.

According to this view, participants bypass their native linguistic structures and prefer a non-grammatical gesture order because they are not using the computational system of grammar. This view becomes plausible when considering how resistant adult speakers of a language are in abandoning the linguistic structures of their native language when using a foreign language (Muysken, 1988, cf. Odlin, 1989). Evidence in favor of this view comes also from American Sign Language, which has undergone a change from SOV to SVO (Kegl, 2008). In order to provide experimental evidence in favor of either one of the two alternatives, namely whether the SOV order in gestures is a simple modality effect or whether it reflects a preference of a cognitive system other than the computational system of grammar, a fourth Experiment was carried out.

2.5 Experiment 4: Speech comprehension

Experiment 4 investigated whether also the computational system of grammar has word order preferences, and whether these preferences differ from the constituent order preferences found for communication in the absence of the computational system of grammar in Experiments 1, 2 and 3. Evidence from Creole languages suggests that when children have a pidgin's vocabulary at their disposal during language acquisition, they grammaticalize the input by engaging the computational system of grammar (Bickerton, 1984). It may therefore be that when normally hearing adult speakers of a language hear word strings in their native language, they may also make use of the computational system of grammar to organize them. Experiment 4 therefore tested the comprehension of artificially synthesized and prosodically flat word strings in participants' native language. If the computational system of grammar is involved in the comprehension of these word strings, we would expect participants to perform fastest on their native word orders: Italian speakers with the SVO and Turkish speakers with the SOV orders.

While there are many reasons, both theoretical and based on language change, that suggest a preference for the SVO order in the computational system of grammar, there is no direct evidence for this preference (Newmeyer, 2000). The reason for the lack of clear evidence may lay in the fact that it is difficult to determine whether a non-native word order is computationally better for speakers of a language that has an alternative canonical word order: participants are simply always better on their native order.

A direct comparison between Italian and Turkish speakers is thus not possible. It may, however, be possible to determine more general word order preferences even in adult speakers of a language. For example, it has been proposed that there is a clear preference for orders where the Subject is in first (SOV and SVO) rather than in medial (OSV and VSO) or final (OVS and VOS) position (Greenberg, 1963; Greenberg, 1978). It has also been argued that world's languages can be classified roughly into Object–Verb (OVS, OSV, SOV) and Verb–Object (SVO, VOS, VSO) languages, because the languages in each of these two groups of act syntactically alike in many ways (Lehmann, 1973, 1978; Vennemann, 1974, 1976). On the basis of this reasoning we would expect, that when exposed to artificially synthesized prosodically flat strings of words in their native language, both Italian and Turkish-speaking participants would on average be faster on orders where the Subject is in the initial position and where the Verb precedes the Object. Because Turkish (SOV) is an Object–Verb language, Turkish-speaking adults performing better on Verb–Object orders and thus overcoming the native Object–Verb constituent, would be especially strong evidence for order preferences in the computational system of grammar.

2.5.1 Participants

Thirty-six Italian native-speaking volunteers (19 females, 17 males, mean age 25.4, range 23–29 years) recruited from the subject pool of the International School of Advanced Studies in Trieste (Italy) and 36 Turkish native-speaking volunteers (22 females, 14 males, ages 20–24) recruited from the subject pool of the Boğaziçi University in Istanbul (Turkey). Participants reported no auditory or language related problems, did not know any sign language and had not participated in Experiments 1, 2 and 3. Participants received a monetary compensation.

2.5.2 Stimuli

The stimuli of Experiment 4 consisted of the 32 simple vignettes used in Experiment 1 (see Appendix A1), and artificially synthesized audio clips describing each of these vignettes with a sentence consisting of a prosodically flat sequence of three words in both Italian and Turkish (see Appendix A4). Like in Experiments 1–3, each vignette was counterbalanced with its mirror image across participants. To determine whether there is a preference for a specific constituent order in speech comprehension, the audio clips were constructed in all the possible six orders of Subject, Object and Verb (SOV, SVO, OSV, OVS, VSO and VOS). To avoid a bias for certain orders through prosodic and phonological cues, the Italian (exemplified in Audio 1 that can be found online at http://dx.doi.org/10.1016/j.cogpsych.2010.01.004) and Turkish (exemplified in Audio 2 that can be found at http://dx.doi.org/10.1016/j.cogpsych.2010.01.004) words were synthesized by using MBROLA (Dutoit et al., 1996; 1997) and PRAAT

(Boersma, 2001). Phoneme files were constructed for each sentence with a phoneme length of 80 ms, pauses between the words of 80 ms and a constant pitch of 200 Hz. To obtain different word orders of the same sentence the order of the words in the phoneme files were changed before synthesizing the sentences. For Italian the It4 voice and for Turkish the Tr1 voice were used. The artificially synthesized sentences were prosodically flat and the sentences describing the same vignette only differed in the order of words. Four native speakers of Italian and four native speakers of Turkish verified that all the audio clips could be clearly understood.

Importantly, Italian is a language that uses Verb-agreement and the verbs of each sentence are marked for the Subject of the sentence. In general, speech perception studies have demonstrated that in Italian there is a clear preference for Subject initial position and the SVO order. Verb agreement, however, also plays a role in parsing (Bates, Devescovi, & D'Amico, 1999). The synthesized sequences of words therefore preserved the inflectional markings in Italian, even though, in the present task they could not be used to disambiguate between the two nouns, given they were both singular. Turkish is additionally a case-marking language where Objects are marked for case. Despite varying the order of words, the auditory sentences used in Experiment 4 preserved verb-agreement in Italian and both verb-agreement and case-markings in Turkish (see Appendix A4). This was not only important because Object initial sentences without case marking are ungrammatical in Turkish (Erguvanli, 1984), but also because the marking of case in Turkish could liberate listeners from relying on the linear order of words. In fact, MacWhinney, Osmán-Sági, and Slobin (1991) found that when Turkish participants had to act out sentences with a Verb and two Nouns, of which the Object was clearly case-marked, they had the same accuracy with all word orders and could interpret all the sequence on the basis of case-marking alone. Based on these findings, any possible preference for one of the logically six possible word orders among Turkish-speaking participants in this reaction-times experiment would be especially strong evidence for one of the orders being more natural for the computational system of grammar.

2.5.3 Procedure

The procedure of Experiment 4 was identical to the procedure of Experiment 3. Participants were seated in front of a computer screen. The participants were given instructions where they were told that they would hear a sentence in their native language (Italian or Turkish). Immediately after each audio clip, participants were instructed to choose as quickly as possible between two drawn vignettes (used in Experiment 1) the one that depicted the content of the sentence they just heard (dual forced choice task).

In the dual forced choice task, one of the vignettes corresponded to the audio clip and the other one did not by randomly deviating in either one of the three constituents (of Subject, Object or Verb) (for an example see Appendix A3). As the vignettes were created according to a combinatorial design, the distracting vignette of one trial was the correct vignette of another trial. Each participant listened to each of the scenarios once in each of the six logically possible orders (192 trials). Each participant saw each of the vignettes six times as the correct target and six times as the distracter. Participants had 1500 ms to make a choice before the next trial would begin. Reaction Times (RTs) were measured from the onset of the dual forced choice task (from the moment when the two vignettes appeared on the screen).

To determine whether there is a word order preference, each participant saw each of the 32 different scenarios once in each of the six logically possible orders of Subject, Object and Verb. The experiment thus had 192 trials and was divided into six experimental blocks: in each block participants saw each of the 32 scenarios only once in semi-randomly determined gesture order. Between experimental blocks participants could take a break for as long as they wished.

2.5.4 Results

Like in gesture comprehension, participants' responses to the artificial synthesized prosodically flat strings of words in their native language show that they were not having difficulties with the task, as Italian-speaking participants only failed to give an answer on average in 5.3% (*SD* = 1.6), and Turkish-speaking participants on average in 5.9% (*SD* = 1.2) of the trials. Similarly, the percentage of correct answers was on average 95.3% (*SD* = 4.7) of all the answered trials by Italian- and 91.2% (*SD* = 8.2) of the answered trials by Turkish-speaking adults.

In order to determine whether the computational system of grammar is involved, it was first necessary to compare participants' performance on individual word orders. The ANOVA between all word orders in Turkish speakers' responses shows that word order influenced their RTs ($F(5, 31) = 6.8$, $P < 0.01$). Post-hoc tests show that the native SOV order elicited fastest RTs for Turkish-speaking adults (Bonferroni-corrected $P < 0.01$) (see Figure 2.5). Also the ANOVA between all constituent orders in Italian speakers' responses shows that word order influenced their RTs ($F(5, 31) = 7.436$, $P < 0.01$). Post-hoc tests show that the native SVO order elicited the fastest RTs for Italian-speaking participants (Bonferroni-corrected $P < 0.02$) (see Figure 2.5). To see whether we find these differences also when taking items, rather than subjects, as random variables, ANOVAs with the vignettes as random variables for Turkish ($F(5, 30) = 21.22$, $P = .029$) and Italian ($F(5, 30) = 40.56$, $P = .031$) participants were also carried out. Posthoc tests show that the SOV order elicited the fastest RTs for Turkish (Bonferroni-corrected $P = .048$) and SVO for Italian (Bonferroni-corrected $P = .028$) participants in the item based analysis as well. To comprehend artificially synthesized prosodically flat strings of words in their native language, both groups were using the computational system of grammar because they performed fastest on their native word orders: Italians on SVO and Turkish on SOV.

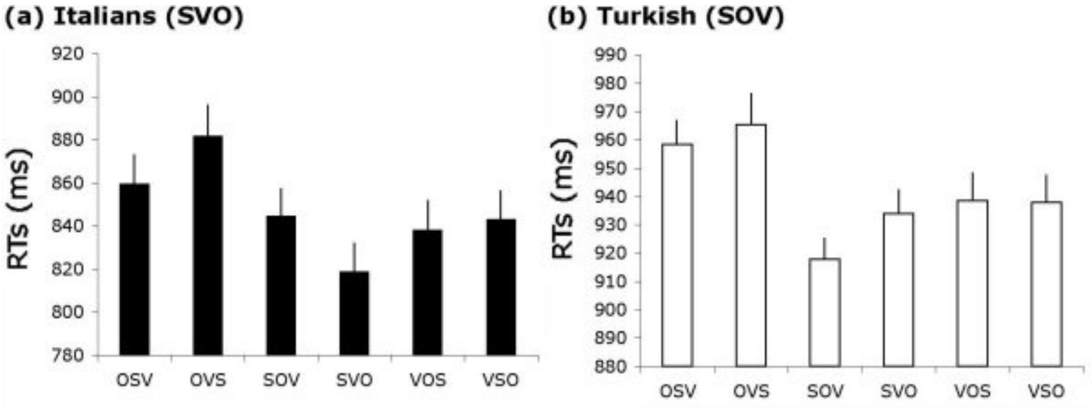


Figure 2.5 Participants' average Reaction Times in the comprehension of artificially synthesized strings of words in their native language.

When we look at all the six logically possible orders, however, we observe consistent preferences across Italian and Turkish speaking participants' performance. For Italians, word orders where the Subject is in initial position (SVO and SOV) elicited significantly shorter RTs than orders where the Subject is in either second (OSV, VSO: 2-tailed *t*-test: *t*(35) = 2.391, *P* < .01) or third position (OVS and VOS: 2-tailed *t*-test: t(35) = 2.201, *P* < .01). Importantly the comparison between Subject in second and third positions failed to reach significance (2-tailed *t*-test: *t*(35) = 9.392, *P* = .39). The same tendency to prefer Subject initial word orders was evident also for Turkish-speaking adults (2-tailed *t*-test between S-initial and S-second position: *t*(35) = 2.670, *P* < 0.01; 2-tailed *t*-test between S-initial and S-third position: *t*(35) = 1.976, *P* < 0.01; 2-tailed *t*-test between S-second and S-third position: *t*(35) = 4.392, *P* = .231). There is thus a clear preference for word orders where the Subject is in the initial position (as in SVO and SOV), but no significant difference whether it occurs in second or third position.

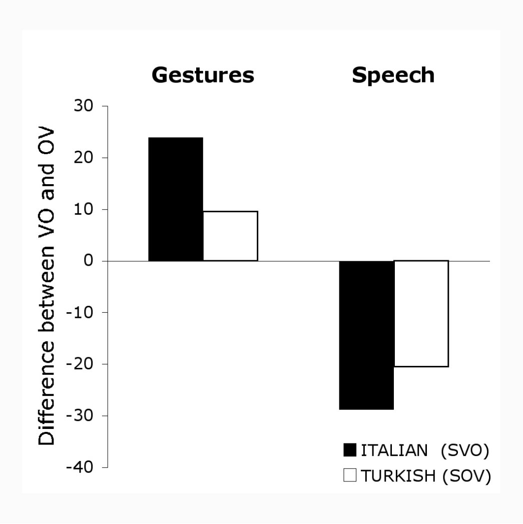


Figure 2.6 Participants' average Reaction Times to Object–Verb and Verb–Object orders in the comprehension of gestures and artificially synthesized strings of words in their native language.

Because both Italian and Turkish are Subject initial languages, it was also important to look at the positions of Objects and Verbs. For Italians, Verb–Object orders (SVO, VOS, VSO) elicited on average significantly shorter RTs than Object–Verb orders (OSV, OVS, SOV) (2-tailed *t*-test between Verb–Object and Object–Verb orders: *t*(35) = 2.591, *P* < 0.01). Exactly the same preference for Verb–Object orders over Object–Verb orders emerged also in Turkish-speaking participants' RTs (2-tailed *t*-test between Object–Verb and Verb–Object orders: *t*(35) = 4.202, *P* < 0.01) (see Figure 2.6).

One possible explanation for the VO preference could be that among the VO orders there are more cases where the Subject precedes the Object (SVO, VSO) than there are among the OV orders (SOV). However, there is a significant preference for the Subject only in the initial position, but no significant difference between Subject in second and third positions (see above). This allows us to compare orders where the Subject is in a non-initial position – either Subject-second (OSV; VSO) or Subject-third (OVS; VOS) – because in these orders the position of the Subject does not matter. Comparing OV (OSV; OVS) to VO (VSO; VOS) orders, we still find a significant preference for VO orders for both Turkish (2-tailed *t*-test between OV (OSV; OVS) and VO (VOS; VSO) orders: $t(35) = 8.736$, $P < 0.001$) and Italian participants (2-tailed *t*-test between OV (OSV; OVS) and VO (VOS; VSO) orders: $t(35) = 9.143$, $P < 0.001$). Because Turkish is an Object–Verb order language, the preference for Verb–Object orders in Turkish-speaking participants' RTs must be independent of participants' native language.

2.5.5 Discussion

While expectedly, in speech comprehension, both groups of subjects were faster on their native order, when we consider all six orders together, both Italian and Turkish-speaking adults show on average shorter RTs with word orders where the Subject is in the initial position (SVO and SOV) as opposed to when it is in the medial or final position, and with word orders where the Verb precedes the Object as opposed to where the Verb follows the Object. While the preference for the Subject in the initial position is clearly interesting, it must be noted that both Italian (SVO) and Turkish (SOV) are Subject initial languages. The overall preference for Subject initial orders may thus be due to the fact that participants were simply fastest on their native order.

This cannot be the case for the preference for Verb–Object orders over Object–Verb orders, because both Italian and Turkish-speaking adults were on average faster with word orders where the Object follows the Verb. Because Turkish (SOV) is an OV language and because the shortest RTs emerged for the native SOV order, participants' general preference for the VO orders (SVO, VOS, VSO) is especially compelling – on average Turkish-speaking adults prefer word orders that violate their native language's Object–Verb directionality. Because Italian and Turkish-speaking participants prefer the same Verb–Object orders, our findings show that the computational system of grammar does have specific preferences for arranging words in sentences. These preferences are independent of participants' native language.

Importantly, when comparing the findings of speech comprehension (Experiment 4) to the findings of gesture comprehension (Experiment 3), we see that improvised gesture and speech have complementary word order preferences: when perceiving sequences of unknown gestures, both Italian and Turkish-speaking adults prefer Object–Verb orders, when perceiving sequences of known words both Italian and Turkish-speaking adults prefer Verb–Object orders. This is the first experimental evidence showing that the computational system of grammar privileges the Verb–Object orders, it also enforces the idea that the SOV order in gestures arises from the direct interaction between the sensory-motor and conceptual system.

Importantly, the complementary order preferences in gesture (OV) and speech (VO) comprehension parallel the word order distribution among the world's languages where the SOV and SVO orders are distributed almost equally (Dryer, 2005). This suggests that the SOV order in improvised gestures is not simply a modality effect, but could very well emerge for the same reason the SOV order prevails among world languages as one of the dominant configurations. Because improvised gesturing bypasses participants' native grammar both in production as well as in comprehension, and fails to show the internal language-like organization of constituents, it is likely that it emerges as the preferred constituent configuration in the direct interaction between the conceptual and sensory-motor systems.

2.6 General Discussion

The present study proposes that the prominence of the SOV and the SVO orders among the world's languages originates from different cognitive systems: SOV is the preferred constituent order in the direct interaction between the sensory-motor and the conceptual system; the SVO order is preferred by the computational system of grammar.

The results of Experiment 1 show that when participants are asked to gesture, they prefer the SOV order in the production of simple clauses, independently of whether their native language is SOV or SVO (see also Gershkoff-Stowe & Goldin-Meadow, 2002; Goldin-Meadow et al., 2008). These results indicate that gesture production is independent of the participants' native grammar. In order to decide whether the SOV strings produced by our participants have the structural properties that characterize the hierarchical constituent structure of SOV languages, or are just a flat sequence of individual gestures, Experiment 2 tested the production of more complex sentences that require a main and a subordinate clause. If grammar were responsible for the SOV order observed in Experiment 1, then subordinate clauses should occupy the position immediately before the verb, as in SOV languages, and in our Turkish-speaking participants' verbal descriptions of the same complex vignettes. The results of Experiment 2 show that when gesturing complex scenarios, neither Italian nor Turkish participants respect the structure of SOV languages. Gesture production thus does not appear to be governed by the computational system of grammar.

Experiments 1 and 2 taken together show that SOV is the preferred order in gesture production for the description of simple scenarios, and that it is not language-like in nature, since the SOV structure breaks down as soon as participants have to describe more complex scenarios. These findings confirm the hypothesis that simple improvised communication is the result of a direct interaction between the sensory-motor and the conceptual systems. Human language, in contrast, must necessarily also make use of the computational system of grammar (Chomsky & Lasnik, 1977).

Experiment 3 tested whether the SOV order preferred in gesture production emerges also in gesture comprehension. The results show that Turkish as well as Italian speaking adults were fastest in choosing the correct vignette after seeing the gestured videos in the SOV order, even though in Italian the SOV order is ungrammatical. Furthermore, on average, both linguistic groups showed a preference for orders where the Object precedes the Verb (OSV, OVS, SOV) over orders where the Object follows the Verb (SVO, VOS, VSO). Because Italian is a Verb–Object language, this preference for orders where the linear order of Verb and Object is reversed, is especially strong evidence for the SOV preference in simple gestural communication.

Why do improvised communication (SOV) and language (SVO) prefer different word orders? While the SOV order in improvised gestural communication parallels the Object–Verb orders found in homesigning children (Goldin-Meadow, 2005), the SVO order proposed for language can be found in children who grammaticalize the pidgin input into Creole languages (Bickerton, 1981; Bickerton, 1984). The difference between these two atypical language acquisition situations – the former having to create a vocabulary and the latter already having the pidgin lexicon – suggests that lexical input may be sufficient to trigger the computational system of grammar. While the experiments on gesture production and comprehension mimic the situation of homesigners, Experiment 4 aimed at creating a task that parallels the situation of children exposed to a pidgin.

Experiment 4 therefore tested the comprehension of artificially synthesized prosodically flat word strings in participants' native language – thus guaranteeing an existing lexicon – and varied the order of the words within the strings. While Italian (SVO) and Turkish (SOV) speaking participants were fastest in choosing the correct vignette after hearing strings in which the words appeared in the order of their respective native language, when comparing participants performance on all the six logically possible word orders, we found that both linguistic groups prefer word orders where the Object follows the Verb (SVO, VOS, VSO) over orders where the Object precedes the Verb (OSV, OVS, SOV). Because Turkish is an Object–Verb language, the findings of Experiment 4 provide strong evidence for the Verb–Object order preference in the computational system of grammar. The asymmetry in the results of the organization of unknown gestures and of known words, strengthens the hypothesis that improvised gesture production, as well as comprehension, is not mediated by the computational system of grammar.

Taken together, Experiments 3 and 4 provide the first cross-linguistic evidence for word order preferences in comprehension. Italian-speaking participants bypassing their native linguistic structures in comprehending improvised gestures, demonstrates that a direct link between the sensory-motor and the conceptual systems that prevails in gesture production, is discernable also in gesture comprehension. The fact that participants chose the correct vignettes faster after seeing gestured videos in the Object–Verb than in Verb–Object orders, shows that this link – unmediated by the computational system of grammar – prefers word orders where the Objects precede the Verbs. In comprehending artificially synthesized words in their native languages, participants were fastest in choosing the correct vignette after hearing sequences of words in their native word orders, showing that the computational system of grammar is involved in processing the word sequences. However, both groups eliciting shorter reaction times on Verb–Object orders, confirms that also the computational system of grammar has a word order preference that is independent of participants' native language and orthogonal to the order preference we found for the direct interaction between the sensory-motor and the conceptual system.

When considering the differences between participants who could rely on their native language as opposed to participants who were faced with the production (Experiments 1 and 2) or comprehension (Experiments 3) of gestures, the findings show a crucial difference. On the one hand, with gestures, participants did not rely on their native syntactic structures nor could they utilize any lexical knowledge, since they had to improvise the gestures in the production experiments and interpret unknown gestures in the comprehension experiment. With artificially synthesized words, instead, participants could at least rely on the lexicon of their native language. Similarly, in homesign, children have to invent their gestures de novo, and when doing so, they introduce the Object–Verb order into their gesture strings (Goldin-Meadow & Mylander, 1983). In contrast, when children are exposed to the unstructured mix of pidgin words, whose meaning they learn from the input, they grammaticalize the pidgin and introduce the SVO order. It is therefore possible, that the prominence of the SOV and SVO orders in atypical language acquisition as well as in the experiments presented above is not due to the fact that in one case participants dealt with gestures and in another with their native language, but to the fact that in one case they did, and in the other they did not, have a lexicon at their disposal.

Proposals concerning the preferences for certain linguistic structures over others in the computational system of grammar have been highly controversial. For instance, it has been argued that recursive structures are easier to understand and process in SVO languages like English and Italian, characterized by rightward embedding, than they are in SOV languages like Japanese and Turkish, characterized by centre embedding (Frazier & Rayner, 1988). It has, however, been shown that Japanese-speakers can very well disambiguate multiple centre-embedded clauses (Mazuka et al., 1989). Thus the preference for one order over the other in the computational system of grammar does not emerge from the inability of the system to process certain syntactic structures.

It has instead been proposed that the preferences for some structural regularities – such as the SVO order and right-branching syntactic structures in general – may arise from the optimality with which they are processed in the computational system of grammar (Hauser et al., 2002). For example, Hawkins (1994) has noted that left-branching languages are likely to violate the branching direction with syntactically heavy embedded clauses, which are often postposed to the right. Because this construction – where subordinate clauses follow the main clauses – is typical of SVO languages, it has been assumed that there is a performance advantage for the SVO order. However, because different languages are not directly comparable, this hypothesis has proven difficult to confirm. The results of Experiment 4 are the first experimental evidence of cross-linguistic preferences for one relative order of verb and object over the other, and show that these are even more fine tuned than previously thought: participants show a preference for Verb–Object orders even with simple artificially synthesized three-word strings in their native language, independently on the language's word order.

The reasons why SOV should be so widespread among the world's languages, as well as in simple improvised communication are less clear. Hawkins (1994) argues that the SOV order is not particularly good for the computational system of grammar because it is possible for adjacent nouns to assume different functions (e.g. a girl can be either the actor or the patient). Thus in SOV languages, it is has often proven useful to additionally overtly mark the grammatical function of nouns with morphological endings (Hawkins, 1994). The findings of Experiment 1–4 show that the preference for the SOV order is motivated outside the computational system of grammar, and must thus originate from either the sensory-motor or the conceptual system. While there appears to be no reason why SOV should be good for the sensory-motor system, Gentner and Boroditsky (2009) have argued that relational terms – such as verbs – require the presence of the entities they link – such as nouns; suggesting that the SOV order may originate from the requirements imposed by the semantic relations in the conceptual system of grammar. It is however also possible that the SOV order results from the different conceptual accessibility of nouns and verbs. For instance, Bickerton (1992) has argued that while nouns have concrete counterparts in the environment, the correspondence of a verb to an action is considerably more vague and therefore more abstract. This may mean that the concepts the nouns represent are more accessible than concepts pertaining to verbs. In fact, Bock and Warren (1985) have shown that the syntactic organization of words in sentences is influenced by the different conceptual accessibility of nouns and verbs. In either case, it appears that the SOV order may be best suited for the conceptual system of grammar.

One question that needs to be answered is how the computational system of grammar emerges in human communication. Goldin-Meadow and Mylander (1998) have observed that when homesigning children gesture sequences of one Noun and one Verb, the Noun they gesture is more likely to represent a patient than an agent, a tendency typical of ergative languages. Based on the idea that the production probability of specific constituents is evidence for syntactic structure, Goldin-Meadow (1982) has argued that also homesigners use the computational system of grammar. However, in situations where homesigns have been grammaticalized into a language, as has happened in the school for the deaf in Nicaragua (Senghas et al., 1997) and a Bedouin village in Israel (Sandler et al., 2005), there is no evidence of consistent agent omission. Thus the production probability of specific constituents is not sufficient evidence in favor of syntax. Since both languages rely on the SOV order (Senghas et al., 1997; Kegl, 2008; Sandler et al., 2005), this suggests that analyzing the production probabilities of specific constituents in homesigning children's gesture strings may be misleading. Thus, until there is evidence that also homesigners organize constituents hierarchically rather than simply beading them together sequentially, it may be concluded that – just as normally hearing adults asked to gesture – also homesigners do not use the computational system of grammar.

Comparing the cases of isolated homesigning children (Goldin-Meadow, 2005; Goldin-Meadow & Mylander, 1998) to the situations where a group of homesigners was brought together – as has happened in the school for deaf in Nicaragua (Shenghas et al., 1997) and the Bedouin village in Israel (Sandler et al., 2005) – it appears that children exposed to no linguistic input during the window of opportunity in which language can be acquired (Lenneberg, 1967), do not use the computational system of grammar. Consistent SOV order emerges – as happened in both Nicaragua and Israel – only when the gestures of homesigners are grammaticalized by a new generation exposed to them (Kegl, 2008). It is therefore likely that the processes that grammaticalize the SOV order of improvised gestures and trigger the computational system of grammar are similar to those explored experimentally by Hudson and Newport (2005) who have shown that children over-regularize the input they receive. While the presence of a group may thus considerably affect the process, Singleton and Newport (2004) have shown that the over-regularization can also occur when a single child receives inconsistent linguistic input from his homesigning parents. According to such a scenario the SOV order that emerges in improvised gestural communication – that relies on the direct link between the sensory-motor and the conceptual system – only grammaticalizes when a child's acquires the gesture system from the input received as its native language.

There are a number of questions that remain to be answered. For instance, why does an existing vocabulary engage the computational system of grammar in Experiment 4 and in the case of Creole languages, whereas when adults have to create the vocabulary while they produce the gesture expressions, it does not? Similarly, the differences between single homesigning children as well as individual adult participants asked to gesture on the one hand, and a group of homesigners brought together in Nicaragua and Israel on the other hand, suggests that also the factor of the group may play a role in how communication systems grammaticalize and engage the computational system of grammar. Further work is necessary to flesh out the necessary and sufficient conditions for the emergence of language in the individual.

In terms of the nature of the human language faculty our results suggest that the structural diversity observed in the world's languages does not emerge from the computational system of grammar alone. The computational system of grammar responsible for generating and interpreting unambiguous structures prefers Verb–Object orders (Experiment 4), and is possibly limited to the SVO order (Chomsky, 1995), thus to right-branching structures (Kayne, 1994). This means that all the alternative grammatical configurations must originate from elsewhere in the language faculty. We have shown that one of the alternatives – the SOV order that is at least as widespread as the SVO order in the world's languages – emerges from the direct interaction between the sensory-motor and the conceptual system. It is therefore likely, that also other word orders, may originate outside the computational system of grammar. Thus, the structural diversity observed among the world's languages may be the result of a struggle between the individual cognitive systems and their interactions trying to impose their preferred structure on human language.

The way one cognitive system imposes its structure on another is frequent in other cognitive domains. For example, observers obligatorily see illusory contours, such as Kanizsa triangles, even when perceiving them impairs their performance of a certain task (Davis & Driver, 1998). Recently, Endress and Hauser (2010) showed that a cognitive system (syntax) can impose its preferences on another cognitive system (motor-system) even in the Human Faculty of Language. For example, simple repetition based grammars (Marcus, Vijayan, Rao, & Vishton, 1999) are so readily learned by newborns (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008) and even rats (Murphy, Mondragon, & Murphy, 2008) that they are considered to constitute a perceptual primitive in speech perception (Endress, Nespor, & Mehler, 2009). However, when the repetitions were based on syntactic categories such as nouns and verbs, listeners fail to detect grammatically impossible rules (Endress & Hauser, 2010). When participants listened to three-word sequences that either started or ended with two words from the same syntactic category (e.g., AAB noun–noun–verb and verb–verb–noun or ABB noun-verb-verb and verb-noun-noun), participants learned the repetition patters when these were consistent with syntactically possible structures (AAB: Noun-Noun-Verb and Adjective-Adjective-Noun, ABB: Verb-Noun-Noun and Noun-Adjective-Adjective) but not when they were syntactically impossible (AAB: VVN and AAV; ABB: NVV and VAA). Importantly, participants identified the categories and learned repetition patterns over non-syntactic categories (e.g., animal–animal–clothes), but they failed to learn the repetition pattern over syntactic categories, even when explicitly instructed to look for it. This shows that when human adults hear a sequence of nouns and verbs, their syntactic system enforces an interpretation and, as a result, listeners fail to perceive the simpler pattern of repetitions (Endress & Hauser, 2010) and suggests that the individual preferences of the cognitive systems may play a role in defining the surface characteristics of world's languages.

On the basis of the experimental results concerning the cognitive systems responsible for improvised gestural communication, it is now also possible to advance a hypothesis about the evolution of the human faculty of language. Communication that can satisfy the simple needs of interpersonal interaction is possible in the absence of the computational system of grammar. By relying on the direct interaction of the sensory-motor and the conceptual system, communication might have emerged as a non-linguistic interaction with its own structural regularities. We thus suggest that human language rests on more primitive cognitive systems still available to humans. Proto-capacities have been shown to co-exist with more modern and fine-tuned cognitive capacities in other cognitive domains as well. For example, it has been suggested that number representation derives from magnitude estimation, a cognitive capacity that is also separately present in modern humans (Feigenson, Dehaene, & Spelke, 2004). The results of experiments 1 to 4 suggest that also our linguistic abilities coexist with, and possibly derive from, a more primitive form of communication that relies on the direct mapping between the conceptual and the sensory-motor system.

These findings also indicate that human language is not a perfect product of engineering, but rather, that evolution has tinkered a patchwork solution (Jacob, 1977) from different, partially conflicting, cognitive systems. Simple communication that relies on the direct interaction between the sensory-motor and the conceptual system prefers the SOV order. If the computational system of grammar had evolved gradually, to enhance the structural coherence and the computational complexity of human communication, we would expect it to have adapted to the structural preferences of the simpler – already existing – form of communication. In such a case, the computational system of grammar should also prefer the SOV order. However, as was shown in Experiment 4, the computational system of grammar prefers Verb–Object orders: orders that are orthogonal to the Object–Verb orders found for simple improvised communication. This suggests that in a particular period in the history of language, the computational system of grammar must have emerged through a process of ''recycling" a pre-existing and evolutionarily older cognitive capacity (Hauser et al., 2002). This process of ''recycling" has recently been proposed for a different cognitive domain: mental arithmetic. In an imaging study, Knops, Thirion, Hubbard, Michel, and Dehaene (2009) showed that participants recycle brain areas used for spatial attention – an evolutionarily older cognitive ability – when engaging in mental arithmetic – a newer cognitive ability for which evolution has not yet dedicated specific brain mechanisms. It is therefore possible that also for human language, the computational system of grammar could have been recruited from pre-existing computational capacities that were already used to process information in a manner that in the language faculty translate to the SVO order.

Chapter 3

Further preferences of the computational system of grammar

3.1 Introduction

Human language uses a number of different grammatical devices for mapping meaning to sound. Minimally these include phrase structure, recursion, word order and morphological marking (Pinker & Jackendoff, 2005). While all known languages are thought to have phrase structure and recursion in their grammatical repertoire (Pinker & Jackendoff, 2005; for a suggested exception see Everett, 2005; and for its a refutation see Nevins, Pesetsky, & Rodrigues, 2009), they differ in their use of word order and morphological marking to represent the function of words. For example, while English relies mainly on word order (Greenberg, 1963), and Mohawk uses mainly morphological marking (Baker, 2001), Japanese has a rich morphology but utilizes also word order (Kuno, 1973; Miyagawa, 1996).

In theory, word order and morphology can be used to accomplish exactly the same task: to signal who did what to whom. To achieve this, word order exploits the fact that the physical realization of words in the speech signal is sequential, and assigns the grammatical functions of Subject, Object and Verb in a consistent order across sentences. In contrast, morphological marking relies on the decomposability of the speech signal into segments and defines the function of words primarily with suffixed (e.g. Japanese), but in some languages also with prefixed (e.g. Tukang Besi) or, rarely, with infixed (e.g. Tagalog) morphemes. In principle, because the morphology of case and agreement can define the function of each word in a sentence, it can liberate human language from linear order.

In practice, however, comparisons between different languages show that, while the great majority of the world's languages relies on word order (with or without morphology), the non-configurational languages that use only morphological marking are very infrequent (Dryer, 2005). Recent findings (Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009) demonstrate that even a language classified as non-configurational, like Basque, has a basic word order that facilitates language processing, suggesting that morphological marking must be accompanied by word order. This landslide preference for order over morphology is taken by some linguists as evidence that word order is in some way a more fundamental or optimal grammatical device than morphological marking of case and agreement (Kayne, 1994; Chomsky, 1995; Hauser, Chomsky, & Fitch, 2002; see however Pinker & Jackendoff 2005 for an alternative view).

Artificial grammar learning studies show that (artificial) languages with and without morphology show different degrees of learnability. For instance, Braine (1966) showed that 9-10-year-old children readily learn the relative position of non-frequent variable tokens with respect to constant marker elements. Green (1979) showed that when morphological markers were consistently ordered with respect to content words and phrases, participants found the grammars easier to learn than when the marker elements were co-occurring with content words and phrases randomly. Furthermore, Morgan (1987) showed that both bound and free morphemes (suffixes and function words, respectively) facilitated the discovery of phrase structure in artificial grammar learning when they were consistently marked with respect to content-words and phrase boundaries. However, the artificial grammars in these studies were semantically empty, thus morphological markers did not represent semantic relations between words. Therefore, morphology, in these experiments, enabled participants break into the continuous speech signal, but was not really acquired to represent the function of words.

There is experimental evidence that shows that morphology may help young infants to break into continuous speech without being acquired as a grammatical device. For example, there are measurable differences in the relative frequency of close class items (e.g. determiners, case morphology and verb agreement) and open class items (i.e. nouns and verbs) in continuous speech (Kucera & Francis, 1967; Cutler & Carter, 1987; Cutler, 1993; Gervain, Nespor, Mazuka, Horie & Mehler, 2008). Gervain et al. (2008) exploited these frequency differences and familiarized Japanese and Italian infants with a sequence of syllables that alternated in their frequency, mimicking the relative frequency of content-words vs. function words with a frequency of one to nine. Following the familiarization phase, infants were tested with a looking-time procedure for their memory of frequent-infrequent or infrequent-frequent syllable pairs. The results demonstrate that infants preferred the relative order of frequent and infrequent words that mirrored the word orders of their native languages: Japanese infants preferred infrequent-frequent and Italian infants frequent-infrequent syllable pairs. This suggests that infants know this basic aspect of the word order of their mother tongue and become sensitive to distributional cues already during the first year of life.

Following the findings of Gervain et al. (2008), Hochmann, Endress and Mehler (2010) familiarized 17-month-old Italian infants with the same stream of syllables that alternated in their frequency, and consequently engaged the infants in a word-object pairing task. The results show that 17-month-old Italian infants treat infrequent words as content words (Hochmann, Endress & Mehler, 2010), suggesting that distributional cues (e.g. frequency) differentiating content words from function words, can help young infants in word learning. There is, however, to date, no evidence that infants in an experimental setting can associate frequent words to grammatical functions of any kind.

There are, instead, differences observed in the acquisition of morphology and word order in older children. Using comprehension tests in which children are asked to act out sentences where morphology defines the grammatical function of words, Slobin and Bever (1982) tested Croatian children between ages 2;0 and 4;8 and Hakuta (1977) Japanese children between ages 2;3 and 6;2. Croatian and Japanese use both word order and morphological marking, and can thus be used to investigate the interaction of order and morphology in language acquisition. The findings of both studies show that morphological marking is acquired later than word order. Importantly, children could interpret morphological markings only when the words were in the canonical word order of their language of exposure and failed to correctly interpret morphology when the words in the sentences were in a non-canonical order. This suggests that morphological marking is learnt as a grammatical tool only after children have learned the basic word order of their mother tongue. This may mean that knowing the basic word order is a prerequisite for learning morphological marking.

The only experimental evidence that has shown learning of morphology as an independent grammatical device comes from Nagata (1981; 1983; 1984). Nagata (1981) taught Japanese adult participants simple artificial grammars by showing them images paired with sentences that described them in a nonsense language. The semantic relations in these scenarios were represented in the sentences by word order and/or morphology. The results show that participants' performance was best when they could rely on both order and morphology. Additionally participants' could also learn the semantic relations with either one of these cues alone, with no significant differences in performance between the word order and the morphology conditions. In two subsequent studies, Nagata varied the transparency between the semantic relations and the grammatical devices. Nagata (1983) incorporated among the sentences that relied on either word order or morphology, sentences that had no consistent structure, and did thus not transparently signal how the semantic relations in the images mapped to the structure of the sentences. Nagata (1984) varied systematically the number of words in the sentences, thus pairing images that should be described with 5-word sentences with auditory stimuli that sometimes had five words and sometimes only had three words. The results of Nagata (1983; 1984) demonstrate that significant differences between word order and morphological markings emerged only when the number of words was reduced: in this case word order was learnt significantly better than morphology. This suggests that the differences in acquiring word order and morphology emerge only when the input is at least partially inconsistent. Given the difficulty with which infants learn morphology during language acquisition (c.f. Hakuta, 1977; Slobin & Bever, 1982) – the ease with which participants learned the morphology with consistent input in Nagata's experimental studies is surprising.

However, the finding that participants could master grammatical regularities with cross-situational learning is interesting in terms of recent findings of rapid vocabulary learning on the basis of statistical distributions in a cross-situational learning-paradigm. In cross-situational learning, words are not mapped to a potential referent within, but across, multiple encounters and learning trials, i.e. the potential meanings of words are disambiguated across different occasions of use (Akhtar & Montague, 1999; Klibanoff &Waxman, 2000; Yu & Smith, 2007; Smith & Yu, 2008). For instance, Yu and Smith (2007) showed participants slides containing simultaneously pictures of two, three or four uncommon objects that were auditorily accompanied with pseudowords in a random order: since it was impossible for participants to directly map an object to a word, they had to pay attention to which words occurred with which objects across several different trials. The results show that participants keep track of which objects occurred with which pseudowords, and could consequently also learn the randomly ordered word-referent pairings after a remarkably sparse exposure (e.g. from six repetitions of each word-referent pair only). These findings – that have been replicated with 12- and 14-month infants – show that cross-situational statistics is a potentially powerful tool in the acquisition of the lexicon. This suggests that if words can be mapped to objects through cross-situational statistics, it may also be possible to map semantic relations in the environment to grammatical devices in speech by paying attention to cross-situational statistics.

In order to compare word order and morphological marking cross-linguistically, it was first necessary to overcome the problem of cross-linguistic variation in linguistic structure. For example, there are six logically possible ways to order words in a sentence according to the grammatical categories of Subject, Object and Verb. In addition, morphological marking varies in terms of its position (prefixed, infixed or suffixed) and its complexity (Clark, 1998). Considering the marking of case alone, a language may assign as few as four (German) or as many as fourteen (Estonian) different cases. In order to overcome this disparity between natural languages, the experiment used a cross-situational learning paradigm in which artificial grammars signalled the functional role of words by using either word order or morphology. The stimuli of the experiments consisted of computer generated nonsense speech, in which the surface-complexity of the 'sentences' was identical across conditions. Because grammatical relations can only be acquired with an existing vocabulary (Moeser &

Bergman, 1973) and deduced from the semantic relations in the environment (Moeser & Bregman, 1972), drawn vignettes accompanied the auditory stimuli of the experiments. These vignettes depicted real world situations and were created in a combinatorial manner to avoid cross-situational differences.

3.2 Experiment 1A: Cross-situational learning of order and morphology

Experiment 1 contrasted adult Italian and Japanese speakers performance in learning word order and morphology in a cross-situational familiarization experiment. Native speakers of Italian and Japanese were chosen because Italian uses word order to signal the function of words and Japanese relies primarily on suffixed morphology (Miyagawa, 1996). Experiment 1 attempted to clarify whether adults can acquire morphological marking and word order equally well as predicted by the studies of Nagata (1981), or whether morphology is learned significantly worse than word order as predicted by the studies on Japanese (Hakuta, 1977) and Croatian (Slobin & Bever, 1982) children. The design of the experiment allowed us to test whether cross-situational statistics is a powerful enough tool to acquire grammatical devices such as word order and morphology.

3.2.1 Participants

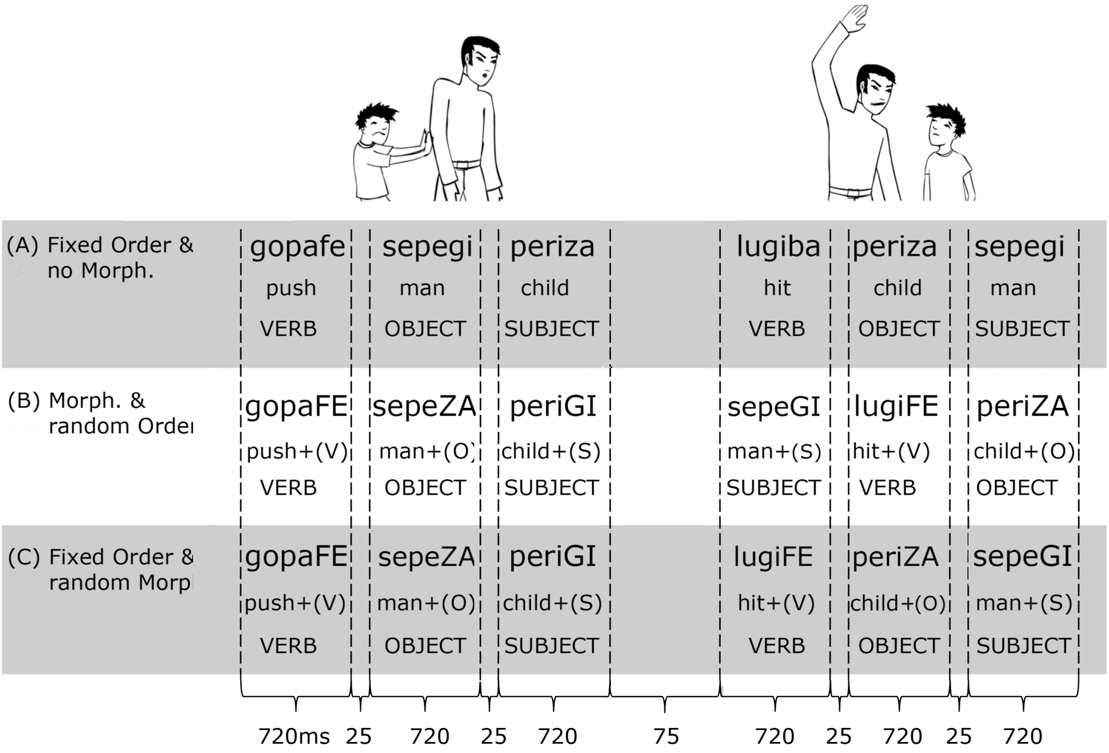
Thirty-nine adult native speakers of Japanese (20 females, mean age 20.7, range 18-23 years) from the subject pools of Tokyo Gakugei University (Tokyo, Japan) and RIKEN Brain Science Institute (Saitama, Japan), and 39 adult native speakers of Italian (18 females, mean age 22.6, range 18-26 years) from the subject pool of SISSA – International School of Advanced Studies (Trieste, Italy). Participants reported no auditory, vision or language related problems and received a monetary compensation. Participants' count does not include 2 Italian and 3 Japanese participants who failed to complete the task.

3.2.2 Stimuli

A cross-situational learning paradigm (e.g. Smith & Yu, 2007) was used to teach participants the grammatical regularities. The stimuli consisted of nonsense auditory sentences in an imaginary computer generated language that 'described' simple drawn vignettes. The black and white vignettes – of the kind 'someone is doing something to someone else' – contained three persons (*a boy, a man, a woman*) and three actions (*hit*, *tell*, *push*). The persons and actions were allocated to the vignettes according to a full combinatorial design. This design resulted in 18 different vignettes (for a full list of vignettes see Appendix B).

To keep the surface complexity of the sentences identical, each sentence had three words (two nonsense nouns and one nonsense verb), and each word three CV (consonant-vowel) syllables (see Figure 3.1). To guarantee identical surface complexity, the auditory stimuli were synthesized by using MBROLA (Dutoit et al., 1996; 1997) and PRAAT (Boersma, 2001). The phonemes were 120ms long and the pitch was constant at 240Hz. To leave the impression of continuous speech but avoid a segmentation task, 25ms long subliminal pauses were inserted between words (cf. Peña et al. 2002), and 75ms pauses between sentences. The words were meaningless in both Japanese and Italian.

The sentences differed in the instantiated regularities. Three sets of sentences were designed to describe the 18 vignettes. Each set of these sentences used consistently one of three regularities: fixed order; morphology with random order; fixed order with random morphology (see Fig. 2.1).



(A) Fixed Order & no Morph.	gopafe push VERB	sepegi man OBJECT	periza child SUBJECT	lugiba hit VERB	periza child OBJECT	sepegi man SUBJECT
(B) Morph. & random Orde	gopaFE push+(V) VERB	sepeZA man+(O) OBJECT	periGI child+(S) SUBJECT	sepeGI man+(S) SUBJECT	lugiFE hit+(V) VERB	periZA child+(O) OBJECT
(C) Fixed Order & random Morp	gopaFE push+(V) VERB	sepeZA man+(O) OBJECT	periGI child+(S) SUBJECT	lugiFE hit+(V) VERB	periZA child+(O) OBJECT	sepeGI man+(S) SUBJECT

Figure 3.1 Example of the vignettes and the accompanying auditory stimuli in Experiment 1.

The three sets of sentences had the following properties:

(A) In the 'fixed word order' sentences (n=18), each word described one person or action depicted in the vignettes. The order of the grammatical categories (of Subject, Object and Verb) was kept constant in all sentences. To ensure that participants were not influenced by the word order of their native language, the sentences used the Verb-Object-Subject (VOS) – a non-canonical order in both Italian and Japanese, and dispreferred cross-linguistically (Dryer, 2005)[7].

(B) In the 'morphological marking' sentences (n=18), each word contained a word-

[7] Italian (SVO) is a considerably more rigid word order language than Japanese (SOV). In fact, it is common in Japanese to have sentences in which words violate the canonical SOV order. However, in Japanese there is one restriction to alternative word order configurations: sentences are Verb final and, if non-Verb-final sentences occur, they are prosodically marked (Tsujimura, 1999). The stimuli in this study were prosodically flat and thus, the VOS order we used, would not be allowed in Japanese. In contrast, morphological markers in Japanese are suffixed to the word-stems in the form of a single consonant-vowel syllable. Because the morphology in this study was also in the form of a single consonant-vowel syllable suffixed on the word-stems, Japanese speakers should, if anything, have an advantage in learning morphological markings over word order.

stem and a suffix. The two-syllables long word-stems ‘described’ the three persons and three actions depicted in the vignettes. Three different one-syllable long suffixes determined the grammatical category and function of each word (Subject, Object or Verb). The suffixed syllables were identical to the final syllables of the words in (A). To guarantee that participants relied only on morphology to determine the grammatical category of the word-stems, the order of words was randomly varied across sentences.

(C) The ‘fixed word order and random morphology’ sentences (n=18) used the word-stems and suffixes of (B). Crucially the word-stems were always in the non-canonical VOS order, as in (A), but the suffixed syllables varied randomly across sentences. The random morphology with fixed order condition was intended to determine whether the suffixed syllables in condition (B) were processed as morphological regularities.

3.2.3 Procedure

Participants were assigned to one of three conditions (fixed order; morphology with random order; or fixed order with random morphology). Participants were told that in the first part of the experiment they would see simple drawings accompanied by computer-generated sentences in an imaginary language (Familiarization phase). Participants were told to pay attention because in the second part of the experiment they would be asked to discriminate correct from incorrect sentences (Test phase).

The familiarization phase consisted of half of the vignettes and the corresponding nonsense sentences (n=9), repeated 11 times each in random order (the familiarization phase was approximately 3 minutes long). In the test phase, participants heard 36 test sentences, half of which correctly described the accompanying vignette (the 9 nonsense sentences withheld during the familiarization phase) and half of which did not. In order to guarantee that participants relied on the grammatical regularities to identify the correct sentences, and not on a change in vocabulary, the incorrect sentences used the same non-sense words (or word-stems) as the correct sentences, but: in (A) – the ‘fixed order’ condition – the correct sentences were in the VOS order and the incorrect sentences in the VSO order; in (B) – the ‘morphology with random order’ condition – the suffix Subjects and Objects in the correct sentence became the suffix of Objects and Subjects (respectively) in the incorrect sentence; in (C) – the ‘fixed order with random morphology’ condition – the correct sentences were in the VOS order and the incorrect sentences in the VSO order as in (A). Participants had to give a YES/NO answer to: “Does this sentence describe this image?”

To determine whether participants learned the meaning of the nonsense words, following the test phase, participants were presented with a list containing all the nonsense words and asked to write down their meanings in their native language.

3.2.4 Results

Figure 3.2 presents Japanese- and Italian-speaking participants’ performance in learning the three grammatical regularities. A two-way analysis of variance showed a main effect for the type of regularity (*F*(2,1) = 63.71, *P* < .001), and for participants’ native language (*F*(2,1) = 9.15, *P* < .010), as well as a significant interaction between type of regularity and native language (*F*(2,1) = 3.29, *P* < .05). Bonferroni-corrected post-hoc comparisons show that Japanese chose the correct and rejected the incorrect sentences significantly better in the ‘fixed order’ than either in the ‘morphology with random order’ (M_{DIF} = 22.69, *Std.error* = 3.56) or in the ‘fixed order with random morphology’ (M_{DIF} = 8.53, *Std.error* = 3.56) conditions. The same pattern emerged also for Italian speakers who performed significantly better in the ‘fixed order’ than in either the ‘morphology with random order’ (M_{DIF} = 35.03, *Std.error* = 3.81) or the ‘fixed order with random morphology’ (M_{DIF} = 10.25, *Std.error* = 3.81) conditions. Furthermore, both groups also performed significantly better in the ‘fixed order with random morphology’ than in the ‘morphology with random order’ conditions (Japanese: $\mathrm{M}_{\mathrm{DIF}}$ = 14.15, *Std.error* = 3.56; Italians: $\mathrm{M}_{\mathrm{DIF}}$ = 24.78, *Std.error* = 3.81). In fact, neither Italian (*t*(24)=1.213, *P* = .237) nor Japanese (*t*(24) = .839, *P* = .410) speakers’ performance differed significantly from chance in the ‘morphology with random order’ condition. Comparisons between the linguistic groups indicate that Italians were better than Japanese at learning the ‘fixed order’ (M_{DIF} = 11.13, *Std.error* = 4.67) and also significantly better than Japanese at learning the 'fixed order with random morphology' (M_{DIF} = 9.42, *Std.error* = 3.27) condition.

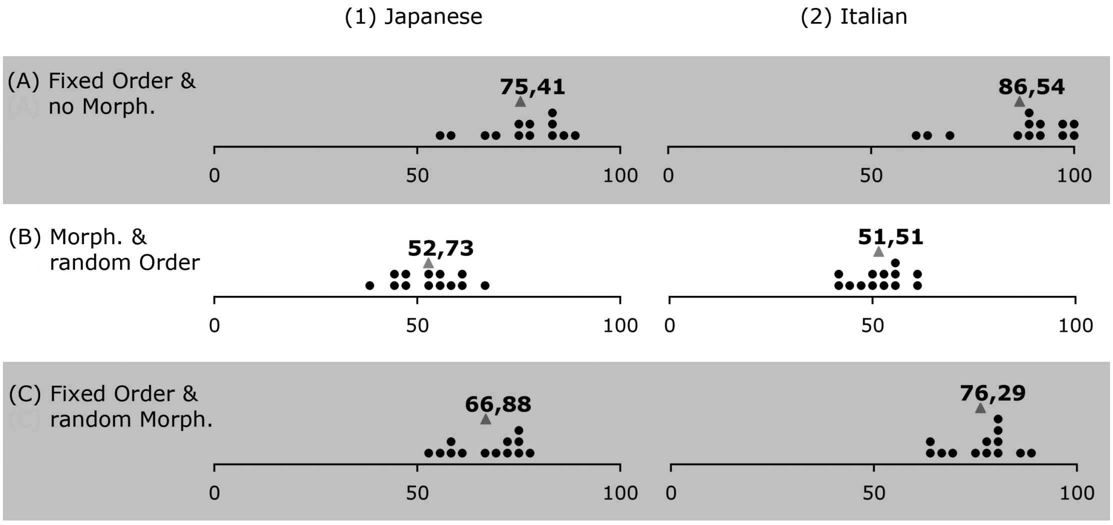


Figure 3.2 Japanese and Italian participants' performance in learning grammatical regularities in Experiment 1A.

Participants' responses in the final vocabulary test show that they could not explicitly state the meaning of the six nonsense words: Italian speakers gave the correct meaning on average to 1.3 and Japanese speakers to 1.6 of the total of 6 words. Participants' performance did not differ significantly from chance either in the 'fixed word order' (*t*-test against chance Italian: *t*(24) = 2.124, *P* = .112; Japanese *t*(24) = 2.124, *P* = .237), in the 'morphology with random order' (*t*-test against chance Italian: *t*(24) = 2.124, *P* = .101; Japanese *t*(24) = 2.124, *P* = .097); or in the 'fixed order with random morphology' condition (*t*-test against chance Italian: *t*(24) = 2.124, *P* = .081; Japanese *t*(24) = 2.124, *P* = .267).

3.2.5 Discussion

The results of Experiment 1A show that both Italian and Japanese speaking adults learned a nonnative word order (Figure 3.2A), but failed to learn morphological marking (Figure 3.2B). Because in Japanese the grammatical categories of words are determined primarily by morphology (Kuno, 1973; Miyagawa, 1996), one would expect Japanese participants to learn also the morphological regularity. Instead, participants' native language does not seem to have determined their strategy for identifying the function of words. In fact, the importance of word order for both Japanese and Italian speakers is also evident when comparing participants' performance with random word order (B) to participants' performance with random morphology (C). While rendering word order random resulted in chance performance when learning the morphological regularity (Figure 3.2B), random morphology only reduced the learning of the word order regularity (Figure 3.2C). This suggests that both linguistic groups targeted word order rather than morphological marking.

Participants might have failed to learn the morphology regularity because they did not notice the morphological markings; alternatively, they might have noticed the suffixed morphology, but failed to link the morphemes with their grammatical functions. The fact that participants performed significantly worse on the same fixed word order (Figure 3.2A) when it additionally had randomly varying morphology (Figure 3.2C), shows that morphological marking did affect participants' performance. Participants' performance in the third condition being significantly above chance (Figure 3.2C) indicates that they extracted the word-stems to learn the word order regularity. This suggests that participants noticed the morphological markers also in the second condition (B) but failed to link them with the grammatical function they represented.

While participants' native language did not determine which of the two grammatical devices they learned, subtle differences between the linguistic groups emerged. Presumably because Italian uses a rigid word order, Italian speakers performed better in learning the word order regularities (Figure 3.2A and 2.2C) than did Japanese-speaking participants, in whose native language word order shows considerably greater variability than in Italian. It is known that one's native language influences the acquisition of nonnative linguistic phonology (e.g. Goudbeek, Cutler & Smits, 2008; Braun, Lemhöfer & Cutler, 2008). That Japanese speakers perform worse on word order compared to Italian speakers, demonstrates that also the knowledge of native morpho-syntactic devices can constrain the acquisition and use of nonnative ones.

Interestingly, the poor performance on the vocabulary test, suggests that it is not necessary for participants to be explicitly aware of the meaning of the nonsense words. While participants had to match the auditory stimuli to the semantic constituents and their relations in the vignettes (e.g. the nouns and the verbs), the results suggest that the rapid acquisition of grammatical regularities such as word order may proceed independently from the explicit acquisition of vocabulary. This is in accordance with the findings of Yu and Smith (2007) in whose cross-situational learning experiment, participants were also reporting that they had learned nothing but could still perform significantly better than chance in choosing the correct word-reference pairs in a four alternative choice task.

3.3 Experiment 1B: Does doubling the familiarization help?

Experiment 1B tests whether a longer familiarization might aid participants in learning morphology. The results of Experiment 1A suggest that participants noticed the morphological markings at the end of the words in condition C, where morphology was randomly varying. This may mean that in condition B (consistent morphology with random word order), participants did not have enough exposure to map the morphology to the grammatical function of words that the suffixed syllables represented. To test whether this is the case, Experiment 1B used a twice as long familiarization phase and hypothesized that if the problem consisted in the brevity of exposure, then we ought to see some improvement in participants' performance in learning morphology in Experiment 1B.

3.3.1 Participants

Twenty-six adult native speakers of Italian (13 females, mean age 22.4, range 18-26 years) from the subject pool of University Milano-Bicocca (Milan, Italy). Participants reported no auditory, vision or language related problems and received course credit for participation.

3.3.2 Stimuli

In order to see whether longer familiarization can induce the learning of morphology, the stimuli of Experiment 1B were identical to the stimuli of Experiment 1A, condition A (fixed word order without morphology), and condition B (morphological marking with random order). The structure of the sentences, the words that made up the sentences and the way the stimuli were synthesized, was identical to Experiment 1A.

3.3.3 Procedure

The procedure of Experiment 1B was identical to the procedure of Experiment 1A with the only exception that in the familiarization phase the sentences were presented twice as many times, resulting in 22 repetitions of the 9 sentences repeated (the familiarization phase was approximately 6 minutes long).

3.3.4 Results

Figure 3.3 presents participants' performance in learning word order and morphology with double long familiarization. Participants who were familiarized with sentences in the fixed VOS order (condition A) chose the correct and rejected the incorrect sentences significantly above chance ($t(22) = .438$, $P < .0001$). However, the performance of participants who were familiarized with sentences that had consistent morphology and random word order, did not differ significantly from chance ($t(22) = .943$, $P = .321$). Participants' performance in the fixed VOS order condition (condition A) was significantly better than participants' performance in the consistent morphology condition (condition B) ($t(22) = .1002$, $P < .001$). When comparing participants' performance in condition A in Experiment 1A and 1B, there was a significant increase in choosing the correct and rejecting the incorrect sentences with double the familiarization ($t(22) = .812$, $P = .039$); when comparing the performance of participants in condition B. In both Experiment 1A and 1B, there differences failed to reach significance ($t(22) = .234$, $P = .821$).

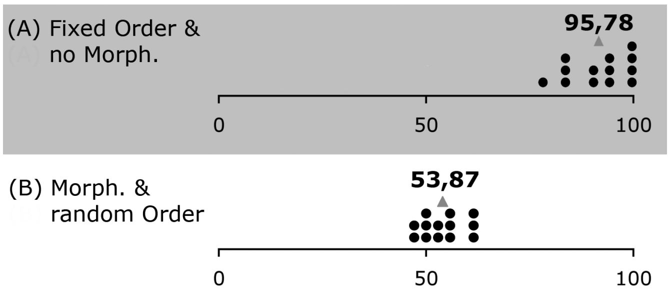


Figure 3.3 Participants' performance in learning grammatical regularities in Experiment 1B.

3.3.5 Discussion

The results of Experiment 1B show that doubling the exposure during the familiarization phase improved participants' performance on word order but not on morphology. The failure to learn morphology in Experiment 1A was thus probably not caused by the fact that participants' did not have enough exposure to the sentences where morphology consistently signalled the function of words. Were it so, we would have expected to see some improvement in learning morphology in Experiment 1B. Instead, the results suggest that morphology, as an independent grammatical device that, by itself signals the function of words, is very difficult to acquire. Thus, Experiment 2 explores an alternative way in which morphology may be acquired.

3.4 Experiment 2: Learning morphology through word order

Following participants' failure to learn morphology in Experiment 1, Experiment 2 explored the possibility that morphology can only be acquired through word order. The findings of Hakuta (1977) on Japanese children and the findings of Slobin and Bever (1982) with Croatian children suggest that children learn morphology only when it is consistent with the canonical word order of their language of exposure.

This is evident in Japanese and Croatian children's systematic failure to interpret morphology when the order of words differs from the canonical word order of their language. To explore the possibility that morphology can only be learned through word order, Experiment 2 used the same cross-situational learning-paradigm as Experiment 1. Italian-speaking participants were familiarized with sentences that had the same VOS order and, additionally, also had suffixed morphology that consistently signaled the function of words. Given consistent word order, we would expect Italian participants to also learn the morphological markings.

3.4.1 Participants

Twenty-six adult native speakers of Italian (12 females, mean age 20.6, range 18-23 years) from the subject pool of University Milano-Bicocca (Milan, Italy). Participants reported no auditory, vision or language related problems and received course credit for participation.

3.4.2 Stimuli

The stimuli of Experiment 2 were created in an identical manner to the stimuli used in Experiment 1 (see above). The difference between the two experiments was that in the familiarization phase of Experiment 2, participants listened to sentences that had words consistently ordered in the VOS order (fixed order) and morphological markings consistently suffixed to two syllable long word-stems (good morphology). Both order and morphology were consistently signaling the grammatical function of words (i.e. whether a word was a Subject, an Object or a Verb).

3.4.3 Procedure

The procedure of Experiment 2 was similar to the procedure of Experiment 1 (see above). The difference was that participants were randomly assigned to one of two conditions. In both conditions participants were familiarized with the same stimuli as described above. The difference between the two conditions was in the Test Phase. In the 'fixed order with inconsistent morphology' condition (condition A) participants had to choose between: correct sentences that had the VOS order (as the sentences of the familiarization phase) and morphology that consistently signalled the function of words (e.g. Subject, Object or Verb); and incorrect sentences that were in the VOS order (as the sentences during the familiarization phase) but had inconsistent morphology that used the suffixed syllables of the familiarization phase but had them randomly suffixed to word stems so that they no longer represented the function of words. In the 'random order with consistent morphology' condition (condition B) participants had to choose between: correct sentences that were in VOS order (as the sentences during the familiarization phase) and morphology that consistently signalled the function of words; and incorrect sentences were in random order and had consistent morphology. Participants had to give a YES/NO answer to: "Does this sentence describe this image?"

3.4.4 Results

Figure 3.4 presents Italian-speaking participants' performance in learning morphology with fixed order. Participants in the 'fixed order with inconsistent morphology' (condition A) chose correct sentences and rejected the incorrect sentences on average 57.3% of the cases (*t* against chance (22) = 4.168, *P* = .002). Participants in the 'random order with consistent morphology' (condition B) chose correct sentences and rejected the incorrect sentences on average 67.7% of the cases (*t* against chance (22) = 2.495, *P* = .030). Participants chose significantly more correct sentences and rejected incorrect sentences in the 'random order with consistent morphology' (condition B) than in the 'fixed order with random morphology' (condition A) (*t*(22) = 1.246, *P* = .023).

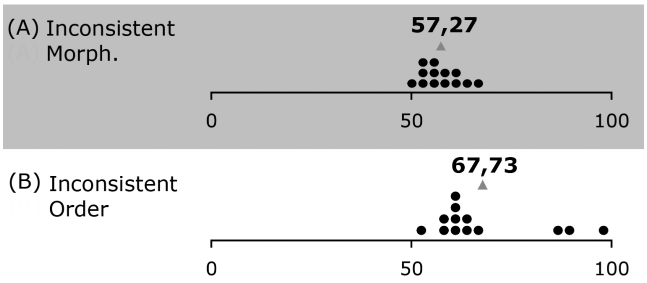


Figure 3.4 Participants' performance in learning grammatical regularities in Experiment 2.

3.4.5 Discussion

In Experiment 2 participants were familiarized with fixed order that had additionally consistent morphology that signalled the function of words, and then queried on test-sentences, half of which always had the correct VOS order and consistent morphology. The only difference between the two experimental conditions was that either the other half of the test-sentences had wrong word order (condition A), or inconsistent morphology (condition B). Participants must have learned the non-native VOS word order because they accepted the correct sentences and rejected the sentences that had inconsistent word order (condition B) significantly above chance (Figure 3.4B). The findings suggest that participants did learn also some morphology because they accepted the correct sentences and rejected the sentences with inconsistent morphology (condition A) significantly above chance (Figure 3.4A). It must however be mentioned that their performance in condition (A) was only 57.3%.

The fact that participants performed significantly better in condition (B) where the incorrect sentences violated the VOS order, than in condition (A) where the incorrect sentences violated the morphology, suggests that word order is better learned than morphology. This result is interesting in relation to the results of Experiment 1, where participants failed to learn morphology. Because during the familiarization, in Experiment 2, participants had both fixed order and consistent morphology available to them, and because word order was learned better than morphology, it appears that fixed word order is a prerequisite for learning morphology as a grammatical device that signals the function of words. Morphology being considerably easier to learn on the basis of word order, may explain why the majority of world’s languages rely on word order – rather than on morphology – as a primary grammatical device.

3.5 General Discussion

In Experiment 1 Italian (SVO) and Japanese (SOV) speaking adults were familiarized with sentences that were identical in their surface complexity but differed in the manner in which they signaled the grammatical function of words. The results show that participants performed better in learning the artificial grammars when the function of words was signaled by fixed word order (VOS) than by morphological markings. In fact, random morphology appeared to reduce participants’ performance in learning fixed word order. These results suggest that participants noticed the randomly varying morphology, but were incapable of assigning the grammatical function of words to the morphological markings.

These findings are in contradiction with previous artificial grammar learning experiments where Japanese participants’ succeeded in learning morphology (c.f. Nagata, 1981; 1983; 1984). Participants’ failure to learn morphology in Experiment 1, as opposed to Japanese speakers’ success in Nagata’s experiments may lay in the fact that, in the studies by Nagata, participants could study the relation between the semantic relations and the grammatical devices on the experimental slides for as long as they thought necessary. In the experiments reported above, the speed of the auditory stimuli defined the length of the familiarization trials and therefore also the exposure to the input: participants could thus not at will manipulate the duration of the exposure. In other words, these experiments presupposed that the extraction of grammatical relations would be considerably more automatic and less mediated by introspection.

This suggests that word order as a grammatical device is simpler to learn through cross-situational statistics than is morphological marking. If we consider the findings of Yu and Smith (2007), that showed that cross-situational statistics is a powerful-tool for extracting word-referent pairs from the environment, even when there is no structure in the speech signal, it becomes evident that in order to learn word order participants could first have extracted the meaning of words during the familiarization phase without relying on any structure and consequently noticed that words always occurred in a certain order. However, for learning the morphological marking, participants had to additionally discover the suffixed syllables – which occurred systematically with certain words and not with others – and only then, could they map the morphology to the semantic relations in the vignettes. This suggests that cross-situational statistics is a suitable tool for extracting word-referent pairs from the environment and acquire word order, but may not be as efficient for acquiring morphology as an independent grammatical device.

In fact, Experiment 2 explored the possibility that morphology might be acquired instead through the basic word order of a language. The hypothesis relied on the findings with Japanese (Hakuta, 1977) and Croatian (Slobin & Bever, 1982) children, who were shown to understand morphology, only if the words in the sentences were in the canonical word order. By familiarizing Italian participants with sentences that had fix word order and consistent morphology, the results of Experiment 2 showed that participants performed significantly better than chance on both word order and morphology. Because participants learned word order significantly better than morphology also in Experiment 2, it is likely that morphology may be acquired through the canonical word order of the language. The findings of Experiment 2 gain strength when we consider that even non-configurational languages, such as Basque, rely on word order (Erdocia et al., 2009), suggesting that in the vast majority of the world's languages word order can be exploited to learn morphology.

Chapter 4

Can prosody be used to discover hierarchical structure in speech?

4.1 Introduction

In order to understand language, it is necessary to process its hierarchical structure. Sentences often contain more than one phrase, phrases more than one word, words more than one morpheme. Adult speakers apply generative rules at each level of this hierarchy, producing from a finite number of morphemes, and words, an infinite number of phrases and sentences. However, it is not clear how language learners manage to keep track of these different levels of linguistic processing when interpreting spoken language. This study investigates the possibility that at least part of the human ability to hierarchically organize words into phrases and phrases into sentences can be acquired from prosody, that is, by tuning into the acoustic properties of the speech signal.

The prosody of speech is characterized by changes in duration, intensity and pitch (Lehiste 1970; Cutler, Dahan, & van Donselaar, 1997). Speakers can intentionally manipulate these acoustic cues in order to convey information about their emotional states (e.g. irony or sarcasm), to define the type of statement they are making (e.g. a question or a statement), to highlight certain elements over others (e.g. by contrasting them), or even to define the meaning of words (e.g. vowel length is phonemic in Estonian and can be used to differentiate lexical entries, e.g. *ma* 'I', *maa* 'land'; pitch is similarly phonemic in tonal languages like Chinese). Additionally, prosody contains information about the syntactic structure of a language. The variation of the specific acoustic properties such as duration, intensity and pitch are, in fact, systematically related to the hierarchical structure of syntax (Selkirk, 1984; Nespor & Vogel, 1986; Nespor, Shukla, van de Vijver, Avesani, Schraudolf, & Donati, 2008). Thus at least some aspects of syntactic information are deducible from the prosodic contour.

4.1.1 The Prosodic hierarchy

Just like syntax, phrasal prosody is structured hierarchically from the prosodic word to the utterance (e.g. Selkirk, 1984; Nespor & Vogel, 1986; Hayes, 1989). The different levels of the prosodic hierarchy are organized so that lower levels are exhaustively contained in higher ones (e.g. Selkirk, 1984). This is best exemplified by considering the two prosodic constituents most relevant for the present paper: the Phonological Phrase and the Intonational Phrase. The Phonological Phrase extends from the left edge of a phrase to the right edge of its head in head-complement languages; and from the left edge of a head to the left edge of its phrase in complement-head languages (Nespor and Vogel 1986)[8]. The constituent that immediately dominates the Phonological Phrase is the Intonational Phrase – a more variable constituent as to its domain – that is coextensive with intonation contours,

[8] The head of a phrase is the word that determines the syntactic type of the phrase of which it is a member, with the other elements modifying the head. In the sentence *that dog has chased many kids*, *chased* is the head of the verb phrase *has chased the kids*. This sentence has two Phonological Phrases: the first beginning at the beginning of the sentence and ending after *dog*; and the second beginning with *has* and extending to the end of the sentence [[that dog]$_{PP1}$[has chased]$_{PP2}$ [many kids]$_{PP3}$].

thus accounting for natural break points in speech (Pierrehumbert & Hirschberg, 1990). While the number of Phonological Phrases contained in an Intonational Phrase may vary, Phonological Phrases never straddle Intonational Phrase boundaries – Phonological Phrases are exhaustively contained in Intonational Phrases.

Despite the similarities between syntactic and prosodic structures, there is no one-to-one correspondence between the two (c.f. Steedman, 1990; Hirst, 1993; Inkelas & Zec, 1990; Shattuck-Hufnagel & Turk, 1996; see Cutler, Dahan, & van Donselaar, 1997 for an overview). The prosodic hierarchy is flatter than the syntactic hierarchy (Selkirk, 1984; Nespor & Vogel, 1986), in that there are fewer different levels in prosody than there are in syntax (Cutler, Dahan, & van Donselaar, 1997), and prosodic structure is not recursive (Selkirk, 1984; Nespor & Vogel, 1986). Prosody thus systematically fails to cue certain syntactic-constituent boundaries. In addition, prosody also creates intonational boundaries that do not coincide with the edges of syntactic constituents. For example, the bracketed string in [The very eminent professor of the London School of Economics (was most avidly reading) the latest wonderful book by Derek Bickerton on the origin of language] can form a single Intonational Phrase, but it does not constitute a syntactic constituent (Selkirk, 1984; see also Steedman, 1990).

However, since in speech production, prosodic structure is mapped from syntactic structure automatically (Nespor & Vogel, 1986), many prosodic cues do signal syntactic boundaries. The boundaries of major prosodic units are associated with acoustic cues like final lengthening and pitch reset or decline (c.f. Klatt, 1976; Cooper & Paccia-Cooper, 1980; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992; Beckman & Pierrehumbert, 1986). These cues are organized so that different levels of the prosodic hierarchy use at least partially different prosodic cues. For example, among the strongest cues for Phonological Phrase boundaries is final lengthening, and for Intonational Phrase-boundaries the declining pitch contour at the right edge and by pitch resetting at the left edge (Price et al., 1991; Wightman et al., 1992). Importantly, the largest variations in pitch and duration, typical of boundaries of prosodic constituents, most often coincide with edges of syntactic constituents (Vaissiere, 1974, 1975; O'Shaughnessy, 1979; Cooper & Sorensen, 1981).

While speakers automatically map syntactic structure onto prosodic contours, there is also evidence that listeners are sensitive to the prosodic cues that signal syntactic constituents. It has in fact been found that listeners locate major syntactic boundaries in the speech stream by relying on prosody alone (Collier & ‘t Hart, 1975; de Rooij, 1975, 1976; Collier, de Pijper, & Sanderman, 1993). Final lengthening in Phonological Phrases (Lehiste, 1973; Lehiste, Olive, & Streeter, 1976; Scott, 1982; Nooteboom, Brokx, & de Rooij, 1978) and the declining pitch contour that characterizes Intonational Phrases (Cooper & Sorensen, 1977; Streeter, 1978; Beach, 1991; Wightman et al., 1992) are the most reliable cues for segmenting continuous speech into syntactic constituents (Fernald & McRoberts, 1995; Cutler, Dahan, & van Donselaar, 1997). For example, adults use final lengthening to segment artificial speech streams (Bagou, Fougeron, & Frauenfelder, 2002) and can use both Phonological Phrase and Prosodic Word boundaries to constrain lexical access (Christophe et al., 2004; Millotte et al., 2008; Marslen-Wilson & Tyler, 1980). While prosodic cues thus appear to signal syntactic constituency in fluent speech, there is, to our knowledge, no evidence that participants can keep track of distinct prosodic cues from different levels of the prosodic hierarchy. Thus, the first question the experiments in this study address is whether listeners view prosody as hierarchically structured, and assign the different cues (e.g. duration and pitch) to specific levels of the prosodic hierarchy.

4.1.2 The two roles of prosody

The possibility that language learners see prosodic cues as hierarchically structured poses the question of how they make use of them. Experimental evidence from language acquisition suggests that prosody has two primary roles in breaking into the continuous speech stream. On the one hand, infants use prosodic cues to segment speech (c.f. Jusczyk, 1998). Infants can discriminate pitch change by 1–2 months of age (Kuhl & Miller, 1982; Morse, 1972). By 4.5 months, infants prefer passages with artificial pauses inserted at clause boundaries rather than other places in the sentence (Jusczyk, Hohne, & Mandel, 1995; Hirsh-Pasek et al., 1987; Kemler Nelson et al., 1995; Morgan, Swingley, & Miritai, 1993). At 6 months, infants are able to use prosodic information consistent with clausal units (Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000) and also demonstrate some sensitivity for prosodic information consistent with phrasal units (Soderstrom et al., 2003). At 9 months, infants show a preference for passages with pauses coincident with phrase boundaries over passages where the pauses are inserted elsewhere in the sentence (Jusczyk et al., 1992). By 13 months of age, infants can use Phonological Phrase boundaries to constrain lexical access (Gout, Christophe, & Morgan, 2004). In sum, the sensitivity to cues carried by prosody appears to emerge within the first year of life.

On the other hand, there is also some evidence that language learners may be able to use the variation in pitch and duration for grouping speech sequences into prosodic constituents, and thus likely also into syntactic constituents. According to the Iambic-Trochaic law (ITL) (Hayes, 1995), elements that alternate in duration are grouped iambically (weak-strong i.e. short-long) and elements that alternate in intensity are grouped trochaically (strong-weak i.e. high-low) (Hay and Diehl, 2007). Nespor et al. (2008) argue that the ITL could also cue word order because Phonological Phrase prominence is signaled mainly with pitch and intensity in complement-head languages, where it is in initial position, and mainly with duration in head-complement languages, where it is in final position. Bion et al. (in press) showed that 7-month-old Italian infants habituated to syllables alternating in pitch, preferred to listen to pairs of prosodically flat syllable pairs that had high pitch on the first syllable during the familiarization phase. The trochaic preference has also been found with English infants (Jusczyk, Cutler, & Redanz, 1993; Thiessen & Saffran, 2003) and an iambic preference with 7-month-old English bilinguals with Japanese, Hindi, Punjabi, Korean or Farsi as the other language (Gervain & Werker, 2008). While age and linguistic environment appear to play a crucial role in the development of these grouping preferences (c.f. Yoshida et al., in press), infants ability to use pitch and duration cues for discovering constituents from continuous speech – just like segmenting continuous speech – seems to emerge during the first year of life.

Thus on the one hand, prosody signals breaks in the speech stream, providing to the listener the edges of individual constituents. For example, phrasal prosodic constituents can be exhaustively parsed into a sequence of non-overlapping words (e.g., Selkirk, 1984; Nespor & Vogel, 1986; Selkirk, 1996; Shattuck-Hufnagel & Turk, 1996). Since phrasal prosodic constituent boundaries are also word boundaries, they can be used for discovering words (Shukla, Nespor, & Mehler, 2007; Millotte, Frauenfelder, & Christophe, 2007). Let's call this process 'segmenting' the speech stream. On the other hand, because prosody uses at least partially different cues to signal breaks at different levels of the prosodic hierarchy, it also provides information about how the segmented units relate to each other. For example, the declining pitch contour does not only signal where an Intonational Phrase begins and ends, but it also groups together the Phonological Phrases that it contains. Lets call this process ‘grouping’ (for a similar distinction in syntactic processing, see Cutler, Dahan, & van Donselaar, 1997). In theory, thus, prosody can play a crucial role in language acquisition both for finding words to build a lexicon and for discovering at least part of the syntactic structure according to which words are arranged into sentences.

However, because the majority of studies on prosody have used single prosodic cues (i.e. either pitch or duration), have not manipulated the individual prosodic cues with respect to their position in the prosodic hierarchy (i.e. words and phrases), and relied on constituents at a single structural level (i.e. either words or phrases), grouping in these studies often assimilates to segmentation. This has as a consequence that prosody is primarily seen as a tool for finding constituents in the speech stream and is often neglected as a viable cue for bootstrapping into the hierarchical syntactic structure. There is, to our knowledge, no evidence that participants can use prosody for understanding the hierarchical structural relations between different prosodic constituents, and thus also possibly have a cue to the structural relations between morphosyntactic constituents, i.e. words, phrases and sentences in the speech stream. Therefore, the second question the experiments in this study address is whether we can highlight the difference between segmentation and grouping by considering the interaction of different levels of the prosodic hierarchy: i.e. whether participants are capable of using prosodic cues from different levels of the prosodic hierarchy to segment continuous speech and additionally are also capable of using these cues for grouping the segmented speech units hierarchically.

4.1.3 Mechanisms for extracting the hierarchical structure from the speech stream

There is some evidence that infants approach the speech stream as if expecting it to be hierarchically structured. Adult participants (Saffran, Newport, & Aslin, 1996) as well as 8-month-old infants (Saffran, Aslin, & Newport, 1996) are able to discover nonsense words from a continuous artificial speech by calculating Transitional

Probabilities (TPs)[9] between adjacent syllables. Following this finding, Saffran and Wilson (2003) investigated whether 12-month-old infants can engage in two statistical learning tasks to discover simple multi-level structure in the speech stream. Infants in that study were familiarized with a recurring sequence of 10 words that conformed to a finite state grammar for two minutes. In this sequence the TPs were always 1.0 within words and .25 at word boundaries.[10] In the test-phase infants listening times to grammatical and ungrammatical syllable sequences showed that they could: (1) use TPs to discover the words and (2) consequently also discover the relations between the segmented words.

While the findings of Saffran and Wilson (2003) show that statistical computations can be used for discovering syntactic-like structures from simple artificial speech, statistical computations alone appear to be insufficient for both segmentation as well as grouping in the acquisition of a real language. For example, even though TPs appear to signal the boundaries of multi-syllabic words, statistical computations fail to segment monosyllabic words. Yang (2004) argued that because monosyllabic words have no word-internal TPs, they are invisible to statistical computations that compare TPs between adjacent syllables and assign segmental breaks between syllables where the TPs drop. Additionally, the size of the lexicon and the possible combinations in which all the words can be arranged, suggests that while the differences between within-word TPs and TPs at word boundaries may differ sufficiently to discover possible word candidates, the differences between TPs at different word boundaries are bound to be considerably smaller than the differences

[9] Corpus studies have shown that there are measurable statistical regularities between sounds that occur within words and those that occur across word boundaries (Harris, 1955; Hayes & Clark, 1970; for a discussion of statistical cues to word boundaries see Brent & Cartwright, 1996). Within a language, the transitional probability from one sound to the next will generally be higher when the two sounds follow one another within a word and lower when they occur at word boundaries. For example, given the sound sequence *pretty#baby*, the transitional probability from *pret* to *ty* is greater than the transitional probability from *ty* to *ba* (Hayes & Clark, 1970). The transitional probability of a sound pair is: YX = (frequency of XY) / (frequency of X) (Saffran, Aslin, & Newport, 1996).

[10] Saffran and Wilson (2003) used a familiarization stream that contained 10 bisyllabic words. The words were allocated to five families (A: dato, kuga; B: pidu, gobi; C: buto, tiga; D: badu, tubi; and E: dipa, tako) and formed 16 sentences. The sentences conformed to a final-state grammar (sentence: A→B→C→D→E). During familiarization every second sentence was preceded by the syllable /la/. Because every word could be followed by either one of the two words from the following family and some words shared the final-syllable (e.g. daTO buTO), the resulting stream had a TP of 1.0 within words and 0.25 TPs at word boundaries.

between within-word TPs and TPs at word boundaries (c.f. Saffran, Newport, & Aslin, 1996). Thus while statistical computations may play a role in discovering possible word-candidates in continuous speech, they are bound to be extremely inefficient in discovering the structural grouping principles between words.

In fact, there is experimental evidence that suggests that transitional probabilities are not always effective cues for finding possible word-candidates. Shukla, Nespor and Mehler (2007) investigated the relative strength of transitional probabilities in tackling the continuous speech stream by looking at the interaction between statistical computations and prosody in speech segmentation. In that study, adult participants were familiarized for eight minutes with an artificial speech-stream that contained statistically defined words (word-internal TP=1.0) that occurred either within 10-syllable-long Intonational Phrases (defined by pitch decline) or straggled Intonational Phrase boundaries. In the test-phase participants were asked to discriminate between words that occurred in the familiarization phase and words that did not. The results show that participants recognized the words only when they occurred within the Intonational Phrase boundaries, but not when they straggled them. This suggests that prosodic cues (i.e. the declining pitch contour) are used as filters to suppress possible statistically well-formed word-like sequences that occur across Intonational Phrase boundaries. However, it is unknown whether prosody constrains statistical computations only at the Intonational Phrase level or also at lower-levels of the prosodic hierarchy.

Statistical computations are not the only processes that researchers have used to demonstrate structural learning in young infants. It is known that infants as young as 7-months can learn simple structural regularities of the kind ABA or ABB (e.g. “gatiga” and “nalili”) from as brief exposures as two minutes (Marcus et al., 1999) and the neonate brain appears to distinguish structural sequences such as ABB (e.g. ‘‘mubaba,’’) from structureless sequences such as ABC (e.g. “mubage”) already during the first days of life (Gervain, Macagno, Cogoi, Peña, & Mehler, 2008). Kovács & Endress (under review) thus investigated with a modified head-turn preference procedure (see Gervain, Nespor, Mazuka, Horie, & Mehler, 2008) whether 7-month-old infants can learn hierarchically embedded structures that are based on identity relations on two different levels. They habituated infants with a stream of syllables that contained words formed by syllable repetitions (“abb” or “aba”, where each letter corresponds to a syllable), and sentences that were formed by repetition of the words ("AAB" or "ABB", where each letter corresponds to a word). In the test-phase, infants' looking times showed that they were able to discriminate novel syllable sequences adhering to the repetition rules from illegal syllable sequences both at the word and at the sentence level. This is one more piece of evidence that suggests that infants do approach the speech signal as if expecting it to be organized on multiple-structural levels.

However, because the majority of studies investigating grammar-like rule learning have used segmented artificial streams, Peña, Bonatti, Nespor, and Mehler (2002) investigated whether a continuous speech stream also allows the extraction of structural generalizations. They familiarized participants with a syllable stream composed of a concatenation of trisyllabic nonsense words. In each word, the first syllable predicted the last syllable with certainty, whereas the middle syllables varied[11]. To identify words and rules, participants could thus not rely on adjacent TPs, but had instead to compute TPs between nonadjacent syllables. The results demonstrate that participants could compute distant TPs, but only for segmenting the speech stream and not for generalizing the dependency between the first and the last syllable of words. After a 10 minute long familiarization participants preferred words that occurred during familiarization over part-words that occurred during familiarization but violated the word-boundaries signaled by TPs. However, they did not prefer novel-rule words that had a novel middle syllable over part-words that actually occurred during the familiarization but violated the word-boundaries signaled by TPs. Participants failed to generalize the long-distance dependencies even after a 30 minute long familiarization. Only when words were separated by subliminal pauses (25ms), could participants generalize the dependency between the first and the last syllable by choosing rule-words that had a novel middle syllable over part-words that occurred during the familiarization but violated the word-boundaries (see

[11] In the long-distance dependencies the first syllable of each tri-syllabic word predicted the last syllable of the word with certainty and the middle syllable varied (e.g. PURAKI, PULIKI, PUFOKI, BERAGA, BELIGA, BEFOGA, TARADU, TALIDU, TAFODU; PU predicts KI with certainty, BE predicts GA with certainty, and TA predicts DU with certainty). Participants were familiarized either with a continuous stream of these words (…PURAKIBELIGATAFODUTALIDUBERAGA …) or with a stream where 25ms long subliminal pauses were inserted between the words (…PURAKI-25ms-BELIGA-25ms-TAFODU-25ms-TALIDU-25ms-BERAGA …). Following the familiarization participants were tested: (1) on words that occurred during the familiarization vs. part-words that occurred during the familiarization but violated the word-boundaries signaled by TPs (e.g. PURAKI vs. RAKIBE); and (2) novel rule-words with the correct long-distance dependency but a novel middle syllable vs. part-words (e.g. PUbeKI vs RAKIBE).

however Perruchet et al., 2004 for criticism; and Bonatti, Peña, Nespor, & Mehler, 2006 for a reply). On the basis of these findings Peña et al. (2002) argued that structural generalizations can be only drawn from a segmented speech stream and that the subliminal pauses that facilitated rule-generalization mimicked the cues provided by prosodic constituency (i.e. Nespor & Vogel, 1986; Selkirk, 1984; cf. Bonatti, Peña, Nespor, & Mehler, 2006 for a thorough discussion). The necessity of segmental cues in the form of pauses for long-distance regularity learning is also found in 12-month-old infants (Marchetto & Bonatti, under review; Marchetto & Bonatti, in prep).

However, in natural languages words are not separated by subliminal pauses (Perruchet et al., 2004). In fact, pauses have been found to be unreliable cues for segmentation in natural speech (for a discussion about pauses as segmentation cues cf. Fernand & McRoberts, 1995). It remains unknown whether real prosodic cues facilitate grammar-like rule learning, and whether they do so at ever level of the prosodic hierarchy. Thus the third question the experiments in this paper address is whether rule-generalization is facilitated also with more realistic prosodic cues (duration and pitch) that correspond to actual cues present in the speech stream. Importantly, because in natural language the structural regularities are organized hierarchically, this study asks whether listeners can generalize hierarchical structural regularities by using cues from different levels of the prosodic hierarchy.

4.1.4 Can prosody be used for discovering hierarchical structure in continuous speech?

Prosody signals syntactic constituency in fluent speech – a characteristic that could facilitate the acquisition of hierarchical structures from the speech stream. However, several crucial questions remain open: (1) Do listeners view prosody as hierarchically organized and assign the different cues (e.g. duration and pitch) to specific levels of the prosodic hierarchy? (2) Can listeners use hierarchically structured prosody to both segment the speech stream and group the segmented units hierarchically? And (3) what is the role of prosody in drawing generalizations from continuous speech?

This study investigates these questions in 3 artificial grammar experiments where participants were first familiarized with an artificially synthesized speech stream that contained prosodic cues that signal constituents at different levels of the prosodic hierarchy, and then tested for learning grammar-like regularities on a dual forced-choice task. The experiments relied on the two prosodic cues that are most reliable for syntactic processing (as discussed above): duration and pitch. These cues were implemented onto two distinct levels of the prosodic hierarchy: duration as Phonological Phrase final lengthening, and a declining pitch contour spanning the Intonational Phrases. The prosody was artificially synthesized over an imaginary language composed of phrases and sentences that contained long-distance dependencies where the first syllable predicted with certainty the last syllable of any given constituent (c.f. Peña et al., 2002) (in the remainder of the article, if not specified otherwise, "phrase" and "sentence" refer to the two structural levels at which we instantiated the long-distance dependency rules). Furthermore, the structure of the familiarization streams was created so that we could investigate both the interaction of prosody and statistical computations and the role of prosody in the extraction of generalization from continuous speech.

In order to see whether participants can keep track of prosodic cues from different levels of the prosodic hierarchy, in Experiment 1 participants were familiarized with both prosodic cues (pitch and duration) simultaneously and half of the participants' were queried for Phonological Phrase level rule-learning and the other half for Intonational Phrase level rule-learning. In order to investigate whether participants relied indeed on both prosodic cues, in Experiment 2 one group of participants was familiarized with a speech stream that contained only final-lengthening as a prosodic cue to phrases and another group with a speech stream that contained only pitch declination as a cue to sentences. In order to investigate whether the findings of Experiments 1 and 2 could be explained by specific properties of the speech streams, participants in Experiment 3 were habituated with a prosodically flat stream.

4.2 Experiment 1: Rule learning with hierarchical prosodic cues

Experiment 1 investigated three questions: (1) Can listeners keep track of prosodic cues from different levels of the prosodic hierarchy? (2) Do listeners perceive prosody as organized hierarchically? and (3) Can listeners use hierarchically structured prosody to acquire hierarchically organized rule-like regularities? Thus, a familiarization paradigm was used to test whether listeners rely on duration to group syllables into phrases, while simultaneously relying on pitch declination to group phrases into sentences, and consequently generalize structural regularities on both levels. The familiarization stream consisted of phrases that followed long-distance dependency rules structurally identical to those used in Peña et al. (2002). However, in order to instantiate rules also at a higher level, let's call it the sentence level, the experiment did not change the order of phrases (contrary to Peña et al., 2002) but paired each two subsequent phrases into a sentence (resulting in 1.0 TPs between phrases rather than 0.5). In the familiarization stream contained two prosodic cues from two distinct levels of the prosodic hierarchy: (A) final lengthening that mimicked Phonological Phrases and was instantiated over the final vowel of each phrase; and (B) pitch declination that mimicked Intonational Phrases and was instantiated over sentences. Participants were expected to be able to use both final lengthening and pitch declination to extract the rules on the phrase- as well as on the sentence-levels.

4.2.1 Participants

Twenty-eight native speakers of Italian (13 females, mean age 20.1, range 19-25 years) from the subject pool of SISSA – International School of Advanced Studies (Trieste, Italy). Participants reported no auditory, vision or language related problems. They received a monetary compensation.

4.2.2 Procedure

The experiment followed a between-subjects design. Participants were randomly assigned to one of two conditions: (1) phrase level rule learning with prosody; or (2) sentence-level rule learning with prosody. In the first part of the experiment participants listened to an imaginary computer-generated language (familiarization phase). Participants were instructed to pay attention because in the second part of the experiment (test phase) they would be asked in a dual forced choice task to discriminate: (a) novel rule-phrases from part-phrases (conditions 1); and (b) novel rule-sentences from part-sentences (conditions 2).

4.2.3 Materials

Participants in both conditions listened to the same familiarization stream. The syllabic structure of the familiarization stream was created so that it contained trisyllabic phrases that formed two-phrase sentences. In order to increase surface variation in the familiarization stream, the phrases and the sentences followed long-distance dependency rules (i.e. the first syllable of each constituent predicted the last syllable with certainty) (c.f. Peña et al., 2002). Figure 4.1 (D, E and F) shows the structure of the familiarization stream. We used three long-distance dependency rules (A_x_C) on the phrase level. In all these rules the first syllable (A_1) always predicted the third syllable (C_1) with a probability of 1.0. The middle syllable (x) varied between three different syllables that were the same for all three rules. Two consecutive phrases formed a sentence. Because the phrases were repeated in the same order throughout the familiarization stream, there were exactly three long-distance dependency rules on the sentence level. In all these rules the first syllable of the first phrase (A_1) always predicted the final syllable of the second phrase (C_2) with a probability of 1.0 and the last syllable of the first phrase (C_1) always predicted the first syllable of the second phrase (A_2) with a probability of 1.0. In the familiarization stream each of the three long-distance dependency rules that formed the phrases (A_x_C) was repeated 60 times (a total of 180 phrase repetitions) and each of the three long-distance dependency rules that formed the sentences (A_1... C_2) was repeated 30 times (a total of 90 sentence repetitions).

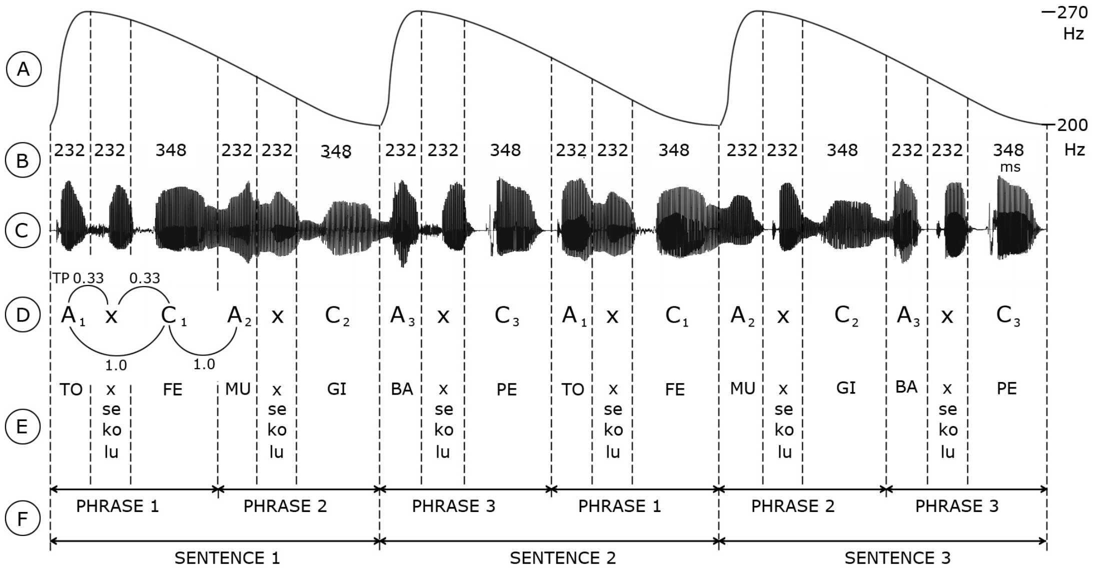


Figure 4.1 Prosodic and syllabic structure of the habitation streams: (A) declining pitch contour; (B) syllable length; (C) the speech signal; (D) the long-distance dependency rules; (E) the actual syllables used; and (F) how syllables formed phrases and sentences. The streams were structurally identical in all four conditions. In the no-prosody conditions (not depicted) the pitch was constant at200 Hz and the syllable length was 232 ms.

The prosodic structure of the familiarization stream was manipulated to see whether prosody can guide the discovery of rules on multiple levels. Prosodic cues were included for Phonological Phrases (final lengthening) and for Intonational Phrases (declining pitch contour) (see Figure 4.1: A, B, C). The final lengthening was instantiated by increasing the duration of the final vowel of each phrase by 50%, resulting in a phoneme length of 232 ms[12]. All the other phonemes were 116ms long. The declining pitch contour started from a baseline of 200Hz with a rapid initial ascent that peaked at 270Hz on the centre of the vowel of the first syllable of the first phrase of the sentence and then declined to 200Hz at the centre of the vowel of the

[12] Final lengthening was instantiated over the final vowel of each phrase (and not over the whole final syllable that consistent of a consonant and a vowel) because pilot experiments showed that participants did not notice lengthening when it was instantiated over the whole syllable. This is in line with the finding that in English that consonants tend to be longer in word-initial position than in word-final position (Oller, 1973; Klatt, 1974; Umeda, 1977). Thus because words are exhaustively contained in Phonological Phrases (e.g. Selkirk, 1984; Nespor & Vogel, 1986), also the consonants at the end of Phonological Phrases must be shorter than at the beginning of Phonological Phrases. This may mean that participants failed to perceive final lengthening over the final syllable because lengthening the consonants was in conflict with the expectation of finding longer consonants at the beginning of the phrases.

last syllable of the second phrase of the sentence. In between these points, pitch was interpolated and then smoothed quadratically (4 semitones). These parameters fall within the range used in previous studies (c.f. Bion, Benavides, & Nespor, in press), and are within the limits of pitch and syllable durations in natural speech (c.f. Shukla, Nespor, & Mehler, 2007). The familiarization stream was 2 min. 26 sec. long. In order to prevent participants from finding the phrases and sentences simply by noticing the first or the final phrase of the familiarization stream, the initial and final 10 sec of the file were ramped up and down in amplitude to remove onset and offset cues.

To test whether participants had acquired the rules at multiple levels, the test phase consisted of 36 trials of a dual forced choice task between two prosodically flat syllable sequences. In the phrase-level rule learning (conditions 1) these sequences were nine novel rule-phrases that had the same A_x_C long-distance dependency but a middle syllable (x) that had not occurred in this position before, and nine part-phrases that were present in the familiarization phase but violated the prosodically signaled phrase boundaries (for a full list of rule-phrases and part-phrases see Appendix B). If participants choose rule-phrases that conform to the long-distance dependency but have a novel surface form over syllable sequences they actually hear during the habituation, they must have generalized the regularities. In the sentence-level rule learning (conditions 2), the test sequences were nine novel rule-sentences that had the same long-distance dependency (A_1 … C_2) as the familiarization sentence, but had a middle syllable (x) that had not occurred in this position before, and nine part-sentences that paired two rule-phrases that occurred during the familiarization but that did not form a sentence in the familiarization phase (for a full list of rule- and part-sentences see Appendix C). If participants choose rule-sentences (that conform to the long-distance dependency but have novel surface forms) over syllable sequences (that they actually hear during the familiarization but that did not form a sentence), they must have on the one hand learned the order of the phrases and on the other have also generalized the long-distance dependencies. All the phonemes in the test items were 116 ms long (test phrases were 696 ms and the test sentences were 1392 ms long). We used prosodically flat test items, in order not to bias the choice of the participants. Thus, the phrases and sentences heard during test are acoustically different from those heard during familiarization.

All the stimuli of this experiment as well as of the following experiments were synthesized with PRAAT (Boersma, 2001) and MBROLA (Dutoit et al., 1996; Dutoit 1997) by using the French female diphone database (fr2). The French diphone database was used since pilot studies showed that artificial speech synthesized using this database resulted in speech that was perceived by Italian adults better than with other similar databases, including the Italian diphone database (notice that the diphone database does not encode sentential prosody).

4.2.4 Results

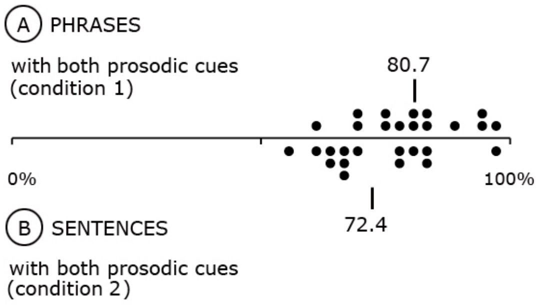


Figure 4.2 Participants' responses in rule learning with hierarchical organized prosodic cues (Experiment 1): (A) the average percent of correctly chosen novel rule-phrases over part-phrases on the word level rule learning with prosody (condition 1); (B) the average percent of correctly chosen novel rule-sentences over part-sentences on the sentence level rule learning with prosody (condition 2).

Figure 4.2A presents the percent of correctly chosen novel rule-phrases. Participants in condition 1 chose novel rule-phrases over part-phrases on average 80.7% of the cases (*t*-test against chance with equal variance not assumed: *t*(13) = 11.670, *P* < .001). Figure 4.2B presents the percent of correctly chosen novel rule-sentences. Participants in condition 2 chose novel rule-sentences over part-sentences on average 72.4% of the cases (*t*-test against chance with equal variance not assumed: *t*(13) = 7.474, *P* < .001). Participants in phrase-level rule learning (condition 1) chose correct novel rule-phrases significantly more than participants in sentence-level rule learning (condition 2) chose novel rule-sentences (*t*-test: *t*(26) = 1.923, *P* = .046).

4.2.5 Discussion

The results suggest that participants used the prosodic cues to learn rules for both phrases and sentences. On the one hand, participants chose significantly more rule-phrases with novel surface structure (novel middle syllables that they had not heard in this position before) than part-phrases they had actually heard during the familiarization but that violated the prosodic boundaries (condition 1). Were they not relying on final lengthening that signalled the phrase-boundary, we would expect them to have preferred part-phrases that actually occurred in the familiarization stream. On the other hand, participants chose significantly more rule-sentences that had a novel surface structure (novel middle syllables that they had not heard in these positions before) over part-sentences that contained phrases they had actually heard during the familiarization phase, but that did not conform to the rule-sentence (condition 2).

Importantly, participants in both conditions were familiarized with the same stream that contained both prosodic cues and they were not informed beforehand whether they would be queried for phrases or sentences. In order to perform above chance on both phrase-level and sentence-level rule learning, listeners must thus have been able to keep track of prosodic cues from different levels of the prosodic hierarchy.

However, significant differences emerged also between phrase-level (condition 1) and sentence-level (condition 2) rule learning. There are two possible explanations for this. It is possible that participants segmented and grouped syllables together according to their respective prosodic cues online. Alternatively, the differences between sentences and phrases may have emerged because there were double as many instances of phrases as there were sentences. The relative difficulty of finding the sentence-level rules could also be increased by the fact that final lengthening is sometimes seen as a stronger prosodic cue than a declining pitch contour. This view is supported by evidence that final lengthening appears to be a more consistent cue to segmentation than the declining pitch contour (de Rooji, 1976; Streeter, 1978; Beach, 1991; for a discussion see Fernald & McRoberts, 1996).

However, there is another possibility that could explain participants' poorer performance on sentences than on phrases. In the familiarization phase the TPs between phrases were always 1.0 (that is, the order of phrases did not change). In the test-phase, participants in condition 2 had to choose between novel rule-sentences that conformed to this order and part-sentences that violated the order of phrases (the phrases that formed part-sentences occurred in the familiarization stream, but the part-sentences themselves did not). Given that participants' found the phrases by relying on final lengthening, it might have been enough to remember the order of all the phrases and not process the pitch-declination at all. Participants thus might have performed worse on sentences because they did not perceive pitch declination: they had to find the phrases first and then the order between all the three phrases. In order to distinguish between these two alternatives, according to which participants were either using both prosodic cues, or only final lengthening, a second Experiment was carried out.

4.3 Experiment 2 Rule learning with individual prosodic cues

Experiment 2 attempted to determine whether participants were indeed using both prosodic cues (final lengthening and pitch declination) to extract rules from the familiarization streams. Experiment 2 kept the syllable structure of the familiarization streams used in Experiment 1, but familiarized participants with either only final lengthening as a cue to phrases or with only pitch declination as a cue to sentences. Participants were expected to learn phrase-level rules only when they had final lengthening as a cue to phrases and to learn sentence-level rules only when they had pitch declination as a cue to sentences: (1) participants who were familiarized with prosodic cues to phrases (final lengthening) were to choose novel rule-phrases over part-phrases that actually occurred in the familiarization stream; (2) participants who were familiarized with prosodic cues to phrases (final lengthening) not to choose novel rule-sentences over part-sentences; (3) participants who were familiarized with prosodic cues to sentences (pitch declination) to choose novel rule-sentences over part-sentences; and (4) participants who were familiarized with prosodic cues to sentences (pitch declination) not to choose novel rule-phrases over part-phrases. Importantly, if participants in Experiment 1 relied only on final lengthening (and not on pitch declination), we would expect them to fail on the sentence-level rule learning because the only prosodic cue they had available was pitch declination.

4.3.1 Participants

Twenty-eight native speakers of Italian (14 females, mean age 21.4, range 20-26 years) from the subject pool of SISSA – International School of Advanced Studies (Trieste, Italy). Participants reported no auditory, vision or language related problems. They received a monetary compensation.

4.3.2 Procedure

The procedure of Experiment 2 was identical to that of Experiment 1, except that the familiarization stream varied either only in duration or only in pitch (instead of varying in both pitch and duration). Therefore, this experiment comprises four conditions - 2 training conditions (stream varying in pitch or stream varying in duration) and 2 test conditions (investigating listeners segmentation of phrases or listeners segmentation of sentences): (1) rule learning with prosodic cues for only phrases (familiarization contained only final lengthening) with test on part-phrases against novel rule-phrases; (2) rule learning with prosodic cues for only phrases (familiarization contained only final lengthening) and test on part-sentences vs. novel rule-sentences; (3) rule learning with prosodic cues for only sentences (familiarization contained only pitch declination) and test on part-sentences vs. novel rule-sentences; and (4) rule learning with prosodic cues for only sentences (familiarization contained only pitch declination) and test on part-phrases vs. novel rule-phrases.

4.3.3 Materials

The syllabic structure of the familiarization stream was identical for all conditions and to the one used in Experiment 1 (see the Materials section of Experiment 1). The crucial difference in respect to Experiment 1 was that the individual prosodic cues were separated into two familiarization streams. Participants in conditions 1 and 2 listened to a familiarization stream that contained prosodic cues for only phrases (final lengthening). Final lengthening was identical to that used in Experiment 1. The resulting familiarization stream was 2 min. and 26 sec. long. Participants in conditions 3 and 4 listened to a familiarization stream that contained prosodic cues for sentences only (declining pitch contour). Pitch declination was identical to that used in Experiment 1. The resulting familiarization stream was 2 min. 5 sec. long (the familiarization stream was shorter for conditions 3 and 4 because there was no final lengthening, but it contained the same number of instances of phrases and sentences as the familiarization stream for conditions 1 and 2). The test phase was identical to that of Experiment 1. The synthesis of the stimuli was identical to that in Experiment 1.

4.3.4. Results

Figure 4.3A presents the percent of correctly chosen novel rule-phrases following the familiarization with final lengthening only. Participants in condition 1 chose novel rule-phrases over part-phrases on average 82.32 % of the cases (*t*-test against chance with equal variance not assumed: $t(13) = 7.221$, $P < .001$). Figure 4.3B presents the percent of correctly chosen novel rule-sentences following the familiarization with final lengthening only. Participants in condition 2 did not significantly choose novel rule-sentences over part-sentences (*t*-test against chance with equal variance not assumed: $t(13) = 10.348$, $P = .66$). Figure 4.3C presents the percent of correctly chosen novel rule-sentences following the familiarization with pitch declination only. Participants in condition 3 chose novel rule-sentences over part-sentences on average 74.80 % of the cases (*t*-test against chance with equal variance not assumed: $t(13) = 10.644$, $P < .001$). Figure 4.3D presents the percent of correctly chosen novel rule-phrases over part-phrases following the familiarization with pitch declination only. Participants in condition 4 did not significantly chose novel rule-phrases over part-phrases (*t*-test against chance with equal variance not assumed: $t(13) = 7.342$, $P = .67$). Participants in phrase-level rule learning (condition 1) chose correct novel rule-phrases significantly more than participants in sentence-level rule learning (condition 3) chose rule-sentences (*t*-test: *t*(26) = 4.237, *P* = .045).

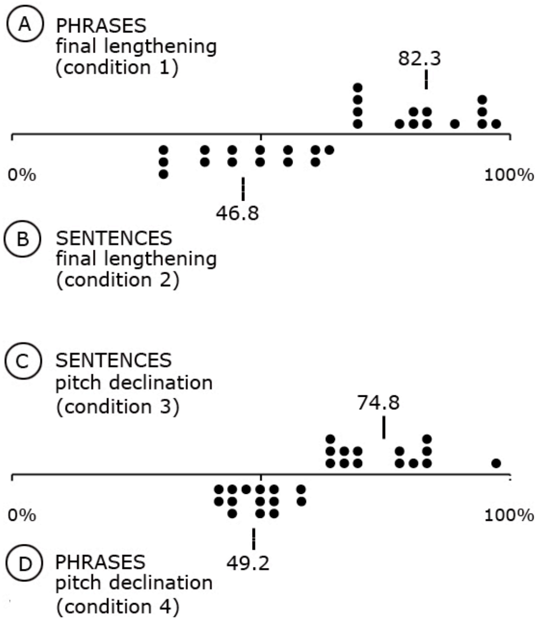


Figure 4.3: Participants' responses in rule learning with individual prosodic cues (Experiment 2): (A) the average percent of correctly chosen novel rule-phrases over part-phrases at the phrase-level rule learning with final lengthening as a cue (condition 1); (B) the average percent of correctly chosen rule-sentences over part-sentences on the sentence level rule learning with final lengthening as a cue (condition 2); (C) the average percent of correctly chosen novel rule-sentences over part-sentences at the sentence-level rule learning with pitch declination as a cue (condition 3); (D) the average percent of correctly chosen novel rule-phrases over part-phrases at the phrase-level rule learning with pitch declination as a cue (condition 4).

No significant differences emerged in rule learning between Experiment 1 and 2. Participants who were familiarized with both prosodic cues (final lengthening and pitch declination) and tested on phrase-level rule learning (Experiment 1 condition 1) did not perform significantly better than participants who were familiarized only with final lengthening and tested on phrase-level rule learning (Experiment 2 condition 1) (*t*-test: *t*(26) = 3.020, *P* = .231). Also participants who were familiarized with both prosodic cues and tested on sentence-level rule learning (Experiment 1 condition 2) did not perform significantly better than participants who were familiarized only with pitch declination and tested on sentence-level rule learning (Experiment 2 condition 3) (*t*-test: *t*(26) = 4.121, *P* = .134).

4.3.5 Discussion

The results show that participants in Experiment 2, just like participants in Experiment 1, used prosodic cues to learn rules for both phrases and sentences. On the one hand, participants familiarized with an artificial speech stream that contained prosodic cues for phrases (condition 1), chose significantly more rule-phrases with novel surface structure (novel middle syllables that they had not heard in this position before) than part-phrases they had actually heard during the familiarization but that violated the prosodic boundaries. Were they not relying on final lengthening that signalled the phrase-boundary, we would expect them to have preferred part-phrases that actually occurred in the familiarization stream. On the other hand, participants familiarized with an artificial speech stream that contained prosodic cues for sentences (condition 3), chose significantly more rule-sentences that had a novel surface structure (novel middle syllables that they had not heard in these positions before) over part-sentences that they had actually heard during the familiarization but that violated the prosodic boundaries. Were they not relying on pitch declination that signalled the beginning and the end of sentences, we would expect them to have preferred part-sentences that had actually occurred in the familiarization stream.

Importantly, we can also rule out the possibility that participants in Experiment 1 were relying only on final lengthening to learn besides the phrase-level rules also the sentence-level rules (and were not relying on pitch declination at all). Participants, who were familiarized with only final lengthening as a cue to phrases, did not choose significantly more novel rule-sentences over part-sentences (condition 2). This suggests that while final lengthening alone enabled participants to generalize the phrase-level rules, it was not sufficient to learn the rules for sentences. Additionally, participants, who were familiarized with only pitch declination as a cue to sentences, did not choose significantly more novel rule-phrases over part-phrases (condition 4). This suggests that while pitch declination alone enabled participants to generalize the sentence-level rules, it was not sufficient to learn the rules at the phrasal-level. Because the syllabic structure of the familiarization streams in Experiment 1 and 2 were identical, participants must have relied on the individual cues: on final lengthening for phrase-level rules and on pitch declination for sentence-level rules.

The results of Experiment 2 also suggest that in Experiment 1 participants were not choosing significantly less novel rule-sentences than novel rule-phrases because they only relied on final lengthening. The findings of Experiment 1 left open the possibility that participants were performing better on the phrasal level than on the sentence level simply because they only used final lengthening to find phrases and then, because the order of the phrases did not vary, discovered the sentences. If this were the case, we would have expected participants who were familiarized only with final lengthening to perform like participants in Experiment 1. In fact, participants in Experiment 2 did choose more novel rule-phrases than novel rule-sentences (conditions 1 and 3). However, participants who were familiarized with final lengthening (Experiment 2 condition 2) did not choose novel rule-sentences significantly over part-sentences, suggesting that final lengthening alone was insufficient to discover rules on both phrasal- as well as sentence-level. The differences between phrase-level and sentence-level rule learning in Experiment 1 and 2 must thus have emerged either because final lengthening is a stronger cue to phrase boundaries than pitch declination is to sentence boundaries, or because in the familiarization phase there were twice as many instances of phrases as there were sentences.

Interestingly, the comparisons between the findings of Experiment 1 and Experiment 2 also suggest that the strength of a prosodic boundary is not the sum of the strength of individual prosodic cues. In the familiarization phase of Experiment 1 the end each sentence was marked by both final lengthening and pitch decline. In contrast, in Experiment 2 participants familiarized with one cue only, in that sentences were only marked by pitch declination. However, there were no significant differences between participants' performance in learning sentence-level rules in Experiment 1 (condition 2) and in Experiment 2 (condition 3). The results thus suggest that pitch declination and final lengthening are not additive in signaling prosodic boundaries. Instead, participants seem to have assigned the individual cues to specific levels in the prosodic hierarchy and used them separately to discover rules at the phrase-level and sentence-level.

While the findings of Experiment 2 showed that participants did indeed rely on both prosodic cues (final lengthening and pitch declination) to learn the rules at the phrase-level and at the sentence-level, both Experiment 1 and Experiment 2 relied on the assumption that if participants did not learn the rules they would be performing at chance. However, this need not be the case as the familiarization streams contained also statistical regularities between syllables that could be used for segmentation (Saffran, Aslin and Newport, 1996; Saffran, Newport, & Aslin, 1996; Aslin, Saffran, & Newport, 1998). In order to investigate how participants would treat the familiarization streams when these are stripped of prosodic cues, to establish the magnitude of the rule learning effects, and to see how statistical computations interact with prosodic cues from different levels of the prosodic hierarchy, a third experiment was carried out.

4.4 Experiment 3: Rule learning without prosody

The statistics carried out on the results of Experiments 1 and 2 relied on the assumption that if participants failed to use the prosodic cues and consequently did not extract any kind of regularities from the speech stream, their performance should have been at chance level. However, the syllabic structure of the familiarization stream, that was the same in both previous experiments, contained transitional probabilities that could have influenced participants' performance. Listeners have been shown to assign word-boundaries in a sequence of syllables where the TPs between syllables drop, rather than where they increase (Saffran, Aslin and Newport, 1996; Saffran, Newport, & Aslin, 1996; Aslin, Saffran, & Newport, 1998). In the familiarization streams of Experiment 1 and 2, the TPs between phrases were always 1.0 and the TPs within phrases were always 0.3. This means that, if participants use TPs for segmenting the familiarization stream, they should prefer part-phrases that included the high TP and assign the phrase-boundaries either before or after the middle (x) syllable. To test for this possibility, in Experiment 3, the familiarizations streams were stripped of prosodic cues.

4.4.1 Participants

Twenty-eight native speakers of Italian (14 females, mean age 22.3, range 20-26 years) from the subject pool of SISSA – International School of Advanced Studies (Trieste, Italy). Participants reported no auditory, vision or language related problems. They received a monetary compensation.

4.4.2 Procedure

The procedure was identical to that of Experiment 1, except that participants listened to a familiarization stream that was prosodically flat. The experiment followed a between-subjects design and participants were randomly assigned to one of two conditions: (1) phrase-level rule learning without prosody; or (2) sentence-level rule learning without prosody.

4.4.3 Materials

The syllabic structure of the familiarization stream was identical for both conditions and to the ones used in Experiment 1 and 2 (see the Materials section of Experiment 1). However, to see whether participants in Experiment 1 and 2 relied on prosody, or the results emerged due to the structural characteristics of the familiarization streams, in Experiment 3 the prosodic cues were eliminated. Participants in Experiment 3 thus listened to a prosodically flat familiarization stream where the pitch was kept constant at 200Hz and all phonemes were 116 ms long. The test phase was identical to those used in Experiment 1 and Experiment 2. The synthesis of the stimuli was identical to those in Experiment 1 and Experiment 2.

4.4.4. Results

Figure 4.4A presents the percent of correctly chosen novel rule-phrases. Participants who were habituated with the flat stream (condition 1) chose novel rule-phrases over part-phrases on average 41.3% of the cases (*t*-test against chance with equal variance not assumed: *t*(13) = -2.150, *P* = .051). Figure 4.4B presents the percent of correctly chosen novel rule-sentences. Participants who were habituated with the flat stream (condition 2) chose novel rule-sentences over part-sentences on average 39.7% of the cases (*t*-test against chance with equal variance not assumed: *t*(13) = -4.642, *P* < .001). The difference between participants who had to choose between novel rule-phrases and part-phrases (condition 1) or novel rule-sentences and part-sentences (condition 2) was not significant (*t*-test: *t*(26) = .346, *P* = .733).

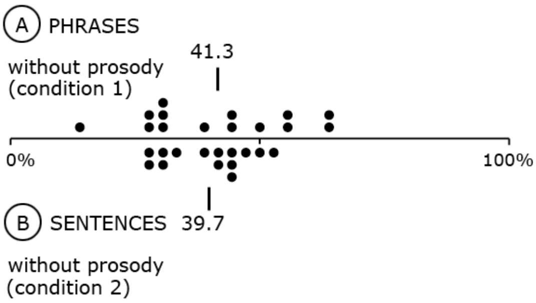


Figure 4.4 Participants' responses in rule learning without prosody (Experiment 3): (A) the average percent of correctly chosen rule-phrases over part-phrases on the word level rule learning without prosody (condition 1); (B) the average percent of correctly chosen rule-sentences over part-sentences on the sentence level rule learning without prosody (condition 2).

When we compare the results of Experiment 1 and 3 we see that participants who were habituated with the stream containing both prosodic cues (final lengthening and pitch declination) chose significantly more novel rule-phrases than participants who were habituated with the flat stream (*t*-test: *t*(26) = 8.070, *P* < .001). Participants who were habituated with the stream containing both prosodic cues chose significantly more also novel rule-sentences than participants who were habituated with the flat stream (*t*-test: *t*(26) = 8.769, *P* < .001). We obtain the same results also when we compare the results of Experiment 2 and 3. Participants who were habituated with the stream containing final lengthening chose significantly more novel rule-phrases than participants who were habituated with the flat stream (*t*-test: *t*(26) = 8.324, *P* < .001). Participants who were habituated with the stream containing pitch declination chose significantly more novel rule-sentences than participants who were habituated with the flat stream (*t*-test: *t*(26) = 7.343, *P* < .001).

4.4.5 Discussion

These results, obtained with prosodically flat familiarization streams, demonstrate that when the familiarization streams were stripped of prosodic cues for phrases (final lengthening) and for sentences (pitch declination), participants could no longer learn the rules for either the phrases or the sentences participants learned in Experiment 1 and 2. Participants in the phrase-level rule learning condition (condition 1) preferred part-phrases to novel rule-phrases. Participants in the sentence-level rule learning condition (condition 2) preferred part-sentences to novel rule-sentences. This means that when prosodic cues for phrase and sentence boundaries were no longer available, participants failed to generalize the long-distance dependency rules and preferred instead syllable sequences that they heard during the familiarization phase. This also means that the generalizations participants made in Experiment 1 and 2 were not simply due to the specific characteristics of the familiarization streams (i.e. the specific syllable combinations used).

The results also suggest that participants were sensitive to transitional probabilities (TPs) between syllables. Listeners have been found to assign constituent boundaries in a sequence of syllables where the TPs between syllables drop, rather than where they increase (Saffran, Aslin and Newport, 1996; Saffran, Newport, & Aslin, 1996; Aslin, Saffran, & Newport, 1998). In the familiarization streams of the experiments reported above, the TPs between phrases were always 1.0 and the TPs within phrases were always 0.3. This means that, if participants were using TPs for segmenting the familiarization stream, they should have preferred part-phrases that included the high TP and assign the constituent boundaries either before or after the middle (x) syllable. The results show that this is the case: participants chose part-phrases over novel rule-phrases (condition 1). These results suggest that when prosodic information is not available, participants use statistical information for segmenting the speech stream. Instead, when we compare the results of Experiments 1 and 2 to the results of Experiment 3, we observe that prosody reversed participants' preference from statistically better-defined part-phrases and part-sentences to prosodically defined rule-phrases and rule-sentences. Final lengthening and pitch declination must be powerful cues to assign a constituent boundary where statistical computations could not assign a boundary (TP 1.0). These results suggest that prosody overrides statistical computations on both the Intonational and the Phonological Phrase level.

4.5 General Discussion

This study reported three experiments that investigated whether: (1) listeners view prosody as hierarchically organized and assign different cues (e.g. duration and pitch) to specific levels of the prosodic hierarchy, (2) listeners use hierarchically structured prosody to both segment the speech stream and group the segmented units hierarchically, and (3) prosody plays a role in drawing generalizations from continuous speech.

The results demonstrate that participants used prosodic cues from different levels of the prosodic hierarchy to learn hierarchically organized structural regularities. In Experiment 1 participants were familiarized with a stream that contained simultaneously prosodic cues to Phonological Phrases (final lengthening) and to Intonational Phrases (pitch declination). In the test phase, participants chose significantly more novel rule-phrases than part-phrases (condition 1), and also significantly more novel rule-sentences than part-sentences (condition 2). This suggested that listeners can keep track of multiple prosodic cues from different levels of the prosodic hierarchy and use these cues to learn hierarchically organized structural regularities. To ensure that participants were indeed relying on both prosodic cues, in Experiment 2, participants were familiarized with either one or the other of the prosodic cues. In the test phase, participants who were familiarized with final lengthening, chose novel rule-phrases significantly over part-phrases (condition 1), however, they did not choose novel rule-sentences over part-sentences (condition 2). This shows that they used final lengthening only for finding phrases. Participants who were familiarized with pitch declination, chose novel rule-sentences significantly over part-sentences (condition 3), however, they did not choose novel rule-phrases over part-phrases. This shows that pitch declination is only used for finding sentences. The findings of Experiment 2 suggest that participants treat prosodic cues from different levels of the prosodic hierarchy separately. The results of Experiment 3, where participants were habituated with prosodically flat streams, show that the findings of Experiment 1 and 2 were not due to any biases caused by the structure of the familiarization streams or possible similarities to words in participants' native language.

In Experiment 1 significant differences emerged between phrase-level rule learning (condition 1) and sentence-level rule learning (condition 2) with prosody. Alternatively to the hypothesis that participants were relying on both prosodic cues, it was possible that participants were only relying on final lengthening and did not use the pitch declination at all (see discussion above): they could have chosen rule-sentences over part-sentences by simply remembering the order of the phrases in the familiarization stream. However, the findings of Experiment 2 demonstrated that when participants were habituated with single prosodic cues, their performance paralleled that of participants who were habituated with both cues simultaneously (Experiment 1). This suggests that these differences were either caused by the fact that there were twice as many instances of phrases in the familiarization stream as there were sentences, and/or that final lengthening is a stronger cue to constituent boundaries than is pitch declination. This latter view is supported by evidence that final lengthening appears to be a more consistent cue to segmentation than the declining pitch contour (de Rooji, 1976; Streeter, 1978; Beach, 1991; for a discussion see Fernald & McRoberts, 1996). In either case, participants segmented and grouped syllables together according to specific prosodic cues online, and they were able to keep track of multiple prosodic cues from different levels of the prosodic hierarchy.

The results of Experiment 1 and 2 are also suggestive of how participants processed final lengthening and pitch declination, two cues that signal constituent boundaries at two different levels of the prosodic hierarchy. In real speech, as well as in our familiarization streams, the final lengthening of the last Phonological Phrase of an Intionational Phrase always coincides with the end of pitch declination. That is, each Intonational Phrase is, in fact, marked by two prosodic cues. Previous studies that have focused solely on speech segmentation have found that duration and pitch declination are either additive in the strength with which they signal boundaries (Streeter, 1978) or are perceived as a single percept (Beach, 1991). However, when we look at participants' performance on sentence-level rule learning in Experiment 1, we see that they did not perform better on sentence-level rule learning (where final lengthening and pitch declined coincided at the sentence final boundary) than on phrase-level rule learning. Similarly, in Experiment 2, where final lengthening was no longer available as a cue to phrase boundaries in the sentence-level rule learning condition (condition 2), participants did not perform significantly worse than participants in Experiment 1 did on sentence-level rule learning. Thus the strength of a constituent boundary is not the sum of the two single prosodic cues (in this case of final lengthening and pitch declination).

Instead, the finding that participants who were familiarized with both cues simultaneously (Experiment 1) did not perform significantly better than participants were familiarized with one cue only (Experiment 2), suggests that listeners know which cue is associated with which level in the prosodic hierarchy and they use the individual cues for finding constituent boundaries at their respective levels only. Participants who were familiarized exclusively with final lengthening could only find phrases and not sentences (Experiment 2 conditions 1 and 2). Similarly participants who were familiarized exclusively with pitch declination could only find sentences and not phrases (Experiment 2 conditions 3 and 4). This may seem surprising if we consider that participants use prosody only for segmenting the speech stream. However, the segregation of the individual prosodic cues may be necessary if we consider that prosody is also used for grouping the segmented units – a process which can only be accomplished if listeners know which prosodic cues signal structural relations at which level in the speech stream (e.g. final lengthening for grouping syllables into phrases and pitch declination for grouping phrases into sentences). By using two distinct prosodic cues to signal structural relations on the phrase-level (signalled by final lengthening) and the sentence-level (signalled by pitch declination), we have shown that participants do not use prosody only for segmenting the speech stream but use it also for finding the structural relations between the segmented units at different levels of the prosodic hierarchy and thus possibly also the syntactic hierarchy.

With respect to rule-learning, the results reported above are in line with the findings of Peña et al. (2002), who demonstrated that statistical computations are powerful enough to segment continuous streams of syllables in short periods of time (see also Saffran, Aslin and Newport, 1996; Saffran, Newport, & Aslin, 1996; Aslin, Saffran, & Newport, 1998), but that subliminal segmentation cues in form of pauses are necessary for extracting higher order structural regularities (i.e. the long distance dependency rules) – an ability that emerges within the first year of life (Marchetto & Bonatti, under review; Marchetto & Bonatti, in prep). However, while the structure of phrases in the study at hand was identical to the structure of words used in Peña et al. (2002), in the familiarization streams the transitional probabilities between phrases were considerably higher than in Peña et al. (TP 1.0 instead of 0.5). Thus, the non-adjacent dependencies (TPs between the first and the last syllable of each phrase and sentence) and the adjacent dependencies (across phrase boundaries) were always 1.0. Because adjacent dependencies are easier to learn than non-adjacent dependencies (cf. Newport & Aslin, 2004; Bonatti et al., 2006 for a discussion), participants familiarized with prosodically flat streams (Experiment 3) preferred part-phrases (that contained the adjacent dependency) to novel rule-phrases (that contained the non-adjacent dependency).

The present findings do not only agree with, but also extend, the results of Peña et al. (2002). Because participants in Peña et al. (2002) only draw generalizations when subliminal 25ms long pauses were introduced between the basic units – the words that contained the long-distance dependencies – the authors argued that prosodic constituent structure (signaled by the subliminal pauses) is a prerequisite for drawing generalizations from continuous speech (Bonatti, Peña, Nespor, & Mehler, 2006). However, on the one hand, systematic 25ms long pauses are not actually present in natural speech and pauses have been found to be unreliable cues for segmentation (cf. Fernand & McRoberts, 1995). The experiments reported above show that participants could draw generalizations also with more natural cues (final lengthening and pitch declination). This enforces the idea that prosodic cues may be necessary for inducing structural generalizations from continuous speech. On the other hand, different prosodic cues occupy different levels of the prosodic hierarchy – i.e. final lengthening at the Phonological Phrase level and pitch declination at the Intonational Phrase level (Nespor & Vogel, 1986; Selkirk, 1984). This means that generalizations of the type shown in Peña et al. (2002) can be drawn on multiple levels of the prosodic hierarchy.

It is important to note that when talking about rule learning, it is possible that participants did not actually learn the long-distance dependency rules either on the phrase or on the sentence levels. Participants could simply have remembered the first and the last syllables of the phrases and the sentences. Endress, Scholl and Mehler (2005) have shown that repetition-based regularities are generalized only at the edges of syllable sequences, suggesting that edges are powerful cues for tackling the speech stream (c.f. Endress, & Mehler, 2009; Endress, Nespor, & Mehler, 2009; Endress, Scholl, & Mehler, 2005). Because all the long-distance dependency rules in experiments 1, 2 and 3 coincided with the prosodic cues signalling phrase and the sentence boundaries, it is possible to successfully complete the task without actually having to compute that the first syllable (A) predicts the last syllable (C) with a probability of 1.0. Thus, because the rules at the phrase-level formed three distinct families (thee different A x C rules) and the rules on the sentence-level formed three distinct families (three different A … C rules) and each rule-family shared distinct initial and final syllables, it is possible that participants' generalized far simpler rules than long-distance dependencies. Regardless of the precise mechanism underlying the rule generalization, the results demonstrate that participants are able to use the prosodic cues for extracting hierarchical regularities from the speech stream.

While these results are the first to demonstrate hierarchical rule learning with cues from different levels of the prosodic hierarchy, the idea that multi-level structure may be acquired from the speech stream is not new. Saffran and Wilson (2003) found that 12-month-old infants are able to segment a continuous speech stream using TPs between syllables and consequently also discover the ordering of the segmented words by using TPs between words. Given the problems with statistical computations as tools for language acquisition (c.f. Yang, 2004), Kovács & Endress (under review) showed that seven-month-old infants can learn hierarchically embedded structures that are based on identity relations of words that followed a syllable repetition ("abb" or "aba", where each letter corresponds to a syllable) that formed sentences based on word repetitions ("AAB" or "ABB", where each letter corresponds to a word). The advantage of prosody, with respect to statistical computations and rule learning is that these latter processes depend on exposure to the speech stream that triggers cognitive processing for structure generalizations (c.f. Peña et al., 2002), whereas prosody can, in theory, occur on single trial learning because it relies on perceptual biases (c.f. Endress, Dehaene-Lambertz, & Mehler, 2007; Endress & Mehler, 2010; Endress, Nespor, & Mehler, 2009; Bion, Benavides & Nespor, in press).

The findings of this study complement this body of research with evidence for hierarchical rule learning with cues that correspond to prosodic cues present in actual speech. On the one hand, prosody provides suprasegmental cues for constituent boundaries at all levels of the prosodic hierarchy. Listeners have been shown to be able to segment speech at Intonational Phrase boundaries (i.e. Watson and Gibson, 2004; Shukla, Nespor, & Mehler, 2007), Phonological Phrase boundaries (Christophe et al., 2003; Christophe, et al., 2004; Millotte, et al., 2008) as well as at Prosodic Word boundaries (Millotte, Frauenfelder, & Christophe, 2007). On the other hand, because different levels of the prosodic hierarchy use at least partially different prosodic cues, they additionally signal how the segmented units relate to each other. For instance, final lengthening is a main signal to the end of Phonological Phrases (Selkirk, 1984; Nespor & Vogel, 1986). In contrast, the declining pitch contour signals Intonational Phrases (Pierrehumbert, & Hirschberg, 1990). Thus, because lower levels of the prosodic hierarchy are exhaustively contained in higher ones (Selkirk, 1984; Nespor & Vogel, 1986), it is possible to determine which Phonological Phrases are contained in any given Intonational Phrase. Because prosody relies on perceptual, rather than computational mechanisms, it may provide a more direct mapping between the speech signal and the hierarchical structure it contains.

Endorsing hierarchical learning with prosody, does not mean that statistical computations are irrelevant to language acquisition. As discussed above, participants, who were habituated with the prosodically flat stream, did use statistical computations for segmentation. However, prosody is a stronger cue to segmentation than transitional probabilities. Previous studies have shown that there is an interaction between statistical computations over syllables and detection of prosodic information. Shukla et al. (2007) demonstrated that statistics is computed over syllables automatically, but prosodic cues (i.e. the declining pitch contour) are used as filters to suppress possible word-like sequences that occur across word boundaries. Similarly, in our experiments, the preference for statistically well-formed phrase-like and sentence-like sequences (Experiment 3) was reverted to a preference for novel rule-phrases and novel rule-sentences when prosodic cues were present in the familiarization stream (Experiment 1 and 2). Final lengthening and pitch declination must thus, be powerful cues to assign a constituent boundary where statistical computations would not assign one (TP 1.0 between phrases). Importantly, our experiment extends the findings of Shukla et al. (2007), who used only Intonational Phrase boundaries, to include final lengthening typical to Phonological phrase boundaries. The findings of the study at hand thus suggest that prosody filters statistical computations on both Intonational and Phonological Phrase levels.

In conclusion, this study investigated whether at least part of the human ability to organize words into phrases and phrases into sentences can be acquired from simply tuning into the acoustic properties of the speech signal. By using duration as a cue to Phonological Phrase boundaries and pitch declination as a cue to Intonational Phrases, it showed that listeners can keep track of prosodic cues from different levels of the prosodic hierarchy, that they perceive prosody as organized hierarchically, and that they are able to use hierarchically structured prosody to acquire hierarchically organized rule-like regularities that mimic the syntactic hierarchy. These findings extend the role of prosody from providing cues to constituent boundaries to a powerful tool for extracting information pertaining to the relation of segmented units in the speech stream. In other words, the information contained in the prosody of the speech signal appears not only rich in cues for discovering hierarchical structure from the acoustic properties of speech, but listeners are capable of extracting one of the core properties of human language from the speech signal alone.

Chapter 6
Conclusions

Over the recent years there has been a gradual shift from representational grammars, like the ‘principles and parameters’ theory (Chomsky, 1981), to seeing the Human Faculty of Language primarily in terms of derivational processes that generate the variety of linguistic structures we observe in the world’s languages. While the specific proposals put forth (c.f. Kayne, 1994; Chomsky, 1995) have considerably simplified the structure of the computational system of grammar (syntax), many researchers have raised concerns about how the differences among the languages emerge (Pinker & Jackendoff, 2005). Following these concerns, the main aim of the present thesis was to test whether the structure and the nature of the Human Faculty of Language might indeed be considerably simpler (Chomsky, 1995; Hauser, Chomsky & Fitch, 2002) as previously thought (e.g. Chomsky, 1957; 1981; Jackendoff, 1997). In particular the studies presented above focused on three major questions: How does systematic grammatical diversity arise? How does the way the grammatical diversity emerges influence the way we conceive the Language Faculty? How are individual languages acquired if all the grammatical diversity is no longer pre-defined in the computational system of grammar?

5.1 The origin of grammatical diversity

The gesture production experiments described in Chapter 2 showed that language-like structure can emerge from outside the computational system of grammar. Just like normally hearing English, Spanish and Turkish speaking adults who had to gesture, instead of using their native language (Gershkoff-Stowe, & Goldin-Meadow, 2002; Goldin-Meadow, So, Ozyürek, & Mylander, 2008), also the Italian (SVO) and the Turkish (SOV) speaking adults in Experiment 1 ordered their gestures in the SOV order. Because adult native speakers do not abandon their native grammar, the results suggest that participants were not using the computational system of grammar to organize their simple gesture strings. The findings in Experiment 2 confirmed the absence of the computational system in improvised gestures : both Italian and Turkish speaking adults failed to produce gesture strings typical to complex sentences in SOV languages. Instead, participants produced complex gesture strings where individual gestures were beaded together linearly without any kind of internal hierarchical structure. Taken together, the results of Experiment 1 and 2 show that participants were not relying on syntax and that the SOV order must have emerged from outside the computational system of grammar.

While the SOV order in gestures does not originate from the computational system of grammar, it is relevant to language because it can become the main grammatical device in new languages that emerge from improvised gestural communication. For example, deaf children born to hearing parents start communicating by using gestures (Goldin-Meadow & Feldman, 1977). These gesture systems, known as homesigns, use the same SOV order (Goldin-Meadow & Mylander, 1998) as the improvised gesture strings of normally hearing adults (Goldin-Meadow, 2005). Importantly, when homesign develops into a new sign language, as happened in the school for deaf in Nicaragua and the Bedouin community in Israel (Kegel, 2008; Senghas, Coppola, Newport, & Supalla, 1997; Shengas & Coppola, 2001; Senghas, Kita, & Özyürek, 2004; Sandler, Meir, Padden, & Aronoff, 2005; Senghas, 2005), the SOV order is used as the primary grammatical device. This indicates that the SOV order that emerged in the gesture production experiments (Experiments 1 and 2) and gesture comprehension experiment

(Experiment 3) is directly linked to the distribution of the SOV order in world's languages.

While the gesture production experiments showed that language-like structures can emerge outside the computational system of grammar, the gesture and speech comprehension experiments (Experiments 3 and 4) unveiled a possible way in which grammatical diversity may emerge among the world's languages. When Italian (SVO) and Turkish (SOV) speaking adults perceived simple gesture strings in all the six logically possible orders of Subject, Object and Verb, they were on average fastest in choosing the correct vignette representing the content of the gesture clip in the Object-Verb orders. In contrast, when Italian and Turkish speaking listened to three word strings in their native language in all the six logically possible orders of Subject, Object and Verb, they were fastest Verb-Object orders. These results show that the cognitive systems responsible for improvised communication and for language, have different word order preferences that match the preferred word orders among the world's languages.

In fact, there is evidence that shows how one cognitive system imposes its preferences on another. Endress and Hauser (2010) showed that participants could learn simple repetition based grammars over syntactic categories (e.g., AAB noun–noun–verb and verb–verb–noun or ABB noun-verb-verb and verb-noun-noun) only if the repetition patterns were syntactically allowed (AAB: Noun-Noun-Verb and Adjective-Adjective-Noun, ABB: Verb-Noun-Noun and Noun-Adjective-Adjective) but not when they were syntactically impossible (AAB: VVN and AAV; ABB: NVV and VAA). This shows that when human adults hear a sequence of nouns and verbs, their syntactic system enforces an interpretation on speech input and, as a result, listeners fail to perceive the simpler repetition pattern (Endress & Hauser, 2010). Taking together the evidence that one cognitive system can enforce its preference on another (Endress & Hauser, 2010) and that individual cognitive systems have specific, and conflicting, preferences for linguistic structure (Experiments 1-4), indicates that it is likely that grammatical diversity among the world's languages may indeed emerge from the struggle between individual cognitive systems trying to impose their preferred structure on human communication.

These kinds of conflicts between the preferences of individual cognitive systems indicate that the Human Faculty of Language, just like other complex biological systems, must have evolved through evolutionary tinkering (Jacob, 1977), where evolution took older pre-existing cognitive abilities to enhance the human cognitive abilities This process of ‘‘recycling” has been shown in different cognitive domains. For example, in an imaging study on mental arithmetic, Knops, Thirion, Hubbard, Michel, and Dehaene (2009) showed that participants recycle brain areas used for spatial attention – an evolutionarily older cognitive ability – when engaging in mental arithmetic – a newer cognitive ability for which evolution has not yet dedicated specific brain mechanisms. In terms of the Human Faculty of Language, the results suggest that the computational system of grammar was recycled into the Language Faculty to provide human language with an enhanced capacity for signalling who did what to whom through linguistic structure. However, the computational abilities themselves must pre-date the recycling process, rather than evolved specificially for the Human Language Faculty. Were it otherwise, we would not expect the cognitive systems responsible for improvised communication and language to have conflicting word order preferences.

If such a view of the Human Language Faculty proves correct, we can expect conflicting preferences on all levels of cognitive and neural processing. In fact, this thesis has assumed that the Human Faculty of Language is modular in the broadest sense of term, i.e. on the level of individual cognitive systems responsible for the sounds and signs of language (phonology), the meaning of utterances (semantics) and the structure of words and sentences (morpho-syntax). However, it is possible that encapsulated and modular processing occurs also within these cognitive systems. For example, within phonology, consonants appear to aide word processing whereas vowels serve primarily for syntactic processing (Nespor, Peña & Mehler, 2003; Bonatti, Peña, Nespor & Mehler, 2005, 2007; Havy & Nazzi, 2009; Nazzi, 2005; Nazzi & Bertoncini, 2009; Hochmann, Benavides-Varela, Nespor, & Mehler, submitted). Within semantics, we can draw a line between word classes such as nouns and verbs (Bickerton 1981; 1992; Jakcendoff, 1992); and most importantly within morpho-syntax, we can encapsulate morphological and syntactic processing (Chomsky, 1957; Jackendoff, 1997; Pinker & Jackendoff, 2005; Hauser, Chomsky, & Fitch, 2002). In other words, all parts of grammar that can be individuated in terms of underlying cognitive processes and/or differential neural substrates are possible sources of conflicting preferences and may thus also trigger the grammatical diversity observed among the world’s languages.

5.2 The relationship between word order and morphology

While the speech comprehension experiment (Experiment 4) in Chapter 2 showed that there are specific word order preferences in the computational system of grammar, word order is not the only grammatical device available to human language to signal the function of words. Minimally, the Human Faculty of Language has to include phrase structure, recursion, word order and morphological marking (Pinker & Jackendoff, 2005). The study reported in Chapter 3 compared word order to morphological marking, because, in theory, these two grammatical devices can be used to accomplish exactly the same task – to signal who did what to whom.

The experiments in Chapter 3 contrasted the learning of morphology and word order in a cross-situational learning paradigm. The methodology relied on the recent experimental evidence that shows that both adults (Yu, & Smith, 2007) and infants (Smith, & Yu, 2008) can employ powerful cross-situational statistics to map word meanings onto entities by simply relying on their co-occurrence. The cross-situational learning paradigm used in the study reported above was modified to additionally query the participants on the structural relations according to which the semantic relations depicted in the vignettes mapped to the auditory stimuli. The results show that participants readily learned the non-native word order (VOS) but failed to perform above chance on the morphology rule (Experiment 1A), even when exposed to twice as many instances of the sample sentences in the familiarization phase (Experiment 1B). Consequently, Experiment 3 showed that participants learned some morphology only when they could additionally rely on fixed word order.

Participants' failure to learn morphological marking in a cross-situational learning situation suggests that word order is computationally simpler than morphological marking. Furthermore, while it is possible that increasing the instances of the familiarization items would eventually have led participants to learn the morphological markings, the results of Experiment 1B suggest that additional experience might not be sufficient for learning morphological marking through calculating statistical co-occurrences. Instead, participants only showed some learning of morphology when they could rely on fixed word order (Experiment 2), which suggests that morphological marking of case and agreement may show a strong and almost universal preference for language to have a basic word order.

While on the one hand these results explain why children master word order before they master morphological marking (Hakuta, 1977; Slobin & Bever, 1982), they also suggest why non-configurational languages that rely primarily on morphology are very rare: they are harder to acquire. These results are supported by recent imaging studies that show how languages thought to be non-configurational, like Basque, have a basic word order that facilitates language processing (Erdocia, Laka, Mestres-Missé, & Rodriguez-Fornells, 2009). This suggests that there are considerably fewer non-configurational languages than previously thought and may even be the case that also languages like Tagalog, that have been claimed to be purely non-configurational, may have a basic word order that can be distinguished from the alternative word order configurations by either frequency of occurrence, marked and un-marked prosody, or ease of processing.

While morphological marking may not exist as a primary grammatical device in any of the world's languages, it may complement the SOV order in clearly mapping meaning to sound. The results of the gesture production and comprehension experiments suggest that new languages are born SOV. This is not only the case with signed languages, but appears to be true also for spoken languages. Studies in historical linguistics indicate that during earlier stages of history the SOV order was considerably more frequent among the world's languages than it is today (Dryer, 2005), suggesting that initially all human languages may have been SOV (Newmeyer, 2000). However, there are reasons to believe that SOV is not particularly well suited for the computational system of grammar. For example, almost all SOV languages allow alternative word order configurations (Steele, 1978), suggesting that they are syntactically not stable. The likely reason for this lies in the fact that the adjacent nouns can assume different functions (e.g. a girl can be either the actor or the patient) and it is not always possible to determine their function without morphological marking (Newymeyer, 2000). This would explain why almost all SOV languages have morphology (Dryer, 2005), while SVO languages tend to loose morphological marking of case and agreement (Heine & Kuteva, 2005): morphological marking complements SOV languages in clearly signalling who did what to whom.

5.3 The role of prosody in the acquisition of syntax

The idea that the computational system of grammar does not define all the grammatical diversity among the world languages has consequences also for language acquisition. Because the grammatical diversity is no longer defined by Principles and Parameters in the computational system of grammar (Chomsky, 1980), it has been implicitly assumed that languages cannot be acquired by parameter setting (Chomsky, 1995; Hauser, Chomsky, & Fitch, 2002). Instead, it has been proposed that infants approach the linguistic input with a toolbox that contains the abilities to calculate Transitional Probabilities between syllables (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996), to generalize algebraic rules (Marcus, Vijayan, Bandi Rao, & Vishton, 1999) and a set of perceptual biases to constrain the speech input (Endress, Nespor, & Mehler, 2009). Recent studies have attempted to show that infants may even use Transitional Probabilities (Saffran & Wilson, 2003) and identity relations between syllables (Kovács, & Endress, under review) to extract multi-level regularities from continuous speech. However, there are problems with both statistical computations and algebraic rule generalizations. For example, there are many cases where Transitional Probabilities fail to detect word boundaries (c.f. Yang, 2004) and there is no evidence that Transitional Probabilities signal syntactic relations between segmented units. Secondly, Peña, Bonatti, Nespor, and Mehler (2002) showed that rule generalizations are carried out only over a segmented input, indicating that language learners need first to segment the speech stream in order to be able to generalize grammar-like rules. Furthermore, both statistical computations and algebraic rule generalizations require considerable exposure to the speech stream.

Given these problems, it is surprising that prosody has received so little attention in the context of grammar acquisition. The variation of the acoustic cues such as pitch, duration and intensity, is systematically correlated to the hierarchical structure of syntax (Selkirk, 1984; Nespor & Vogel, 1986) and evidence shows that the majority of syntactic boundaries can be found by relying on prosody alone (Collier & ‘t Hart, 1975; de Rooij, 1975, 1976; Collier, de Pijper, & Sanderman, 1993). Furthermore, infants appear to become sensitive to the variation of all the major prosodic cues during the first year of life (c.f. Soderstrom et al., 2003), can use these cues for segmenting the speech stream (e.g. Gout, Christophe, & Morgan, 2004) as well as for grouping syllables (e.g. Bion, Benavides, & Nespor, in press). This suggests that prosody may not only be used for segmenting the continuous speech stream, but may additionally also be used for finding the hierarchical relations between the segmented units.

The three experiments presented in Chapter 3 investigated whether participants can use hierarchically structured prosodic cues for extracting hierarchical structure from continuous speech. In Experiment 1, participants were familiarized with a stream that contained simultaneously prosodic cues to Phonological Phrases (final lengthening) and to Intonational Phrases (pitch declination). In the test phase, participants chose significantly more novel rule-phrases than part-phrases and also significantly more novel rule-sentences than part-sentences. This suggested that listeners can keep track of multiple prosodic cues from different levels of the prosodic hierarchy and use these cues to learn hierarchically organized structural regularities. In Experiment 2, in order to confirm that participants were relying on both prosodic cues, participants were familiarized with either one or the other of the prosodic cues. The findings of Experiment 2 parallel the findings of Experiment 1 and suggest that participants treat prosodic cues from different levels of the prosodic hierarchy separately. The results of Experiment 3, where participants were habituated with prosodically flat streams, show that when prosodic cues are not available, participants' resort to using transitional probabilities between syllables to segment the continuous speech stream. These findings show that prosody is not only used for segmenting continuous speech, but can also be used for finding the hierarchical relations between the segmented units.

The results of the three experiments suggest a primary role for prosody in language acquisition. When comparing prosody to the other tools infants may use in language acquisition, it is important to point out that prosody is bound to be more efficient than statistical computations (Saffran, Aslin, & Newport, 1996; Saffran, Newport, & Aslin, 1996) and algebraic rule generalization (Marcus, Vijayan, Bandi Rao, & Vishton, 1999), because prosody functions on perceptual processes that do not involve any calculations or generalizations that are carried out on linguistic input.

Secondly, the results show that prosody is a stronger cue to speech segmentation than Transitional Probabilities. This is evident when we consider that participants' preference for part-phrases and part-sentences after being familiarized with a prosodically flat stream (Experiment 3) was reversed to a preference for rule-phrases and rule-sentences when the familiarization stream additionally contained prosodic cues (Experiments 1 and 2). This indicates that prosody does not only filter statistical computations on the Intionational Phrase level (Shukla, Nespor, & Mehler, 2007), but also on the Phonological Phrase level.

Thirdly, the results of the three experiments also show that hierarchically organized prosody facilitates rule generalization. Peña, Bonatti, Nespor and Mehler (2002) showed that rule generalizations occur only when the speech stream was segmented with subliminal pauses. The experiments reported above showed that this is also true with more natural prosodic cues (pitch declination and final lengthening). Furthermore, because participants generalized long-distance dependency rules on the phrase-level and on the sentence-level only when the familiarization streams contained cues for phrase and sentence boundaries, the results also show hierarchical prosodic cues are necessary for generalizing rules on multiple structural levels.

5.4 Concluding remarks

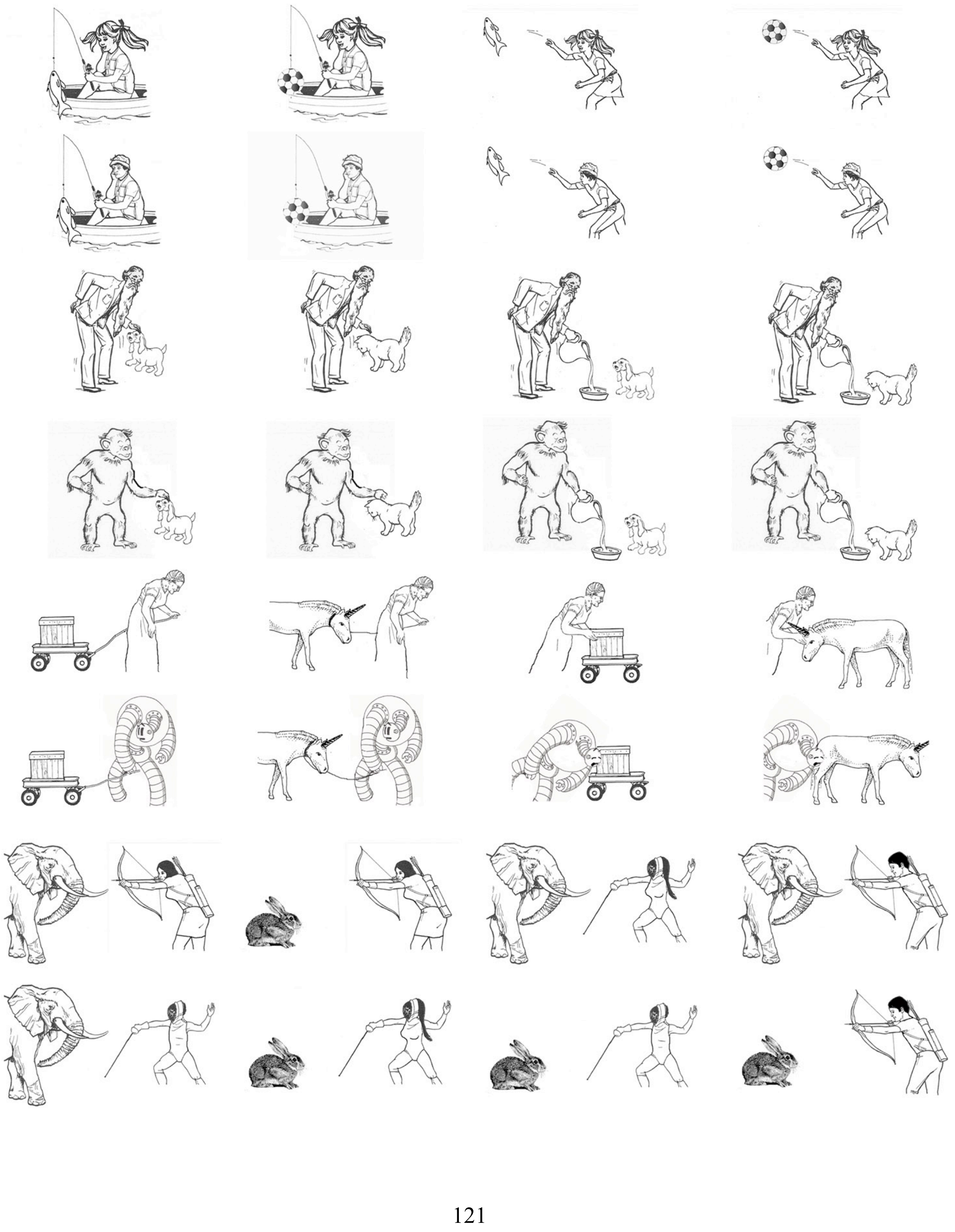
In conclusion, the picture that emerges from the three studies presented above suggests that the Human Faculty of Language is structurally much simpler than previously thought: the grammatical diversity observed among the world languages does not have to be genetically encoded into the structure of the computational system of grammar, but rather, it emerges from the interaction between the individual cognitive systems that make up the Language Faculty. The idea that grammatical diversity can emerge from the conflicts between the specific preferences of individual cognitive systems, or even individual cognitive processes, provides concrete suggestions for further research that will have to be empirically validated. For example, where do the remaining four (of the six logically possible orders) emerge from? Will we be able to explain all the systematic grammatical diversity in terms of the conflicts between the modules in the Language Faculty? What about the strength of the preferences of the individual cognitive systems and how do they affect language change? While many questions still remain unanswered, I believe that the proposal advanced on the previous pages may remove some of the road-blocks on the path of linking linguistic structure to the human cognitive abilities and consequently to the neuro-biological basis of human language.

Appendixes

6.1 Appendix A

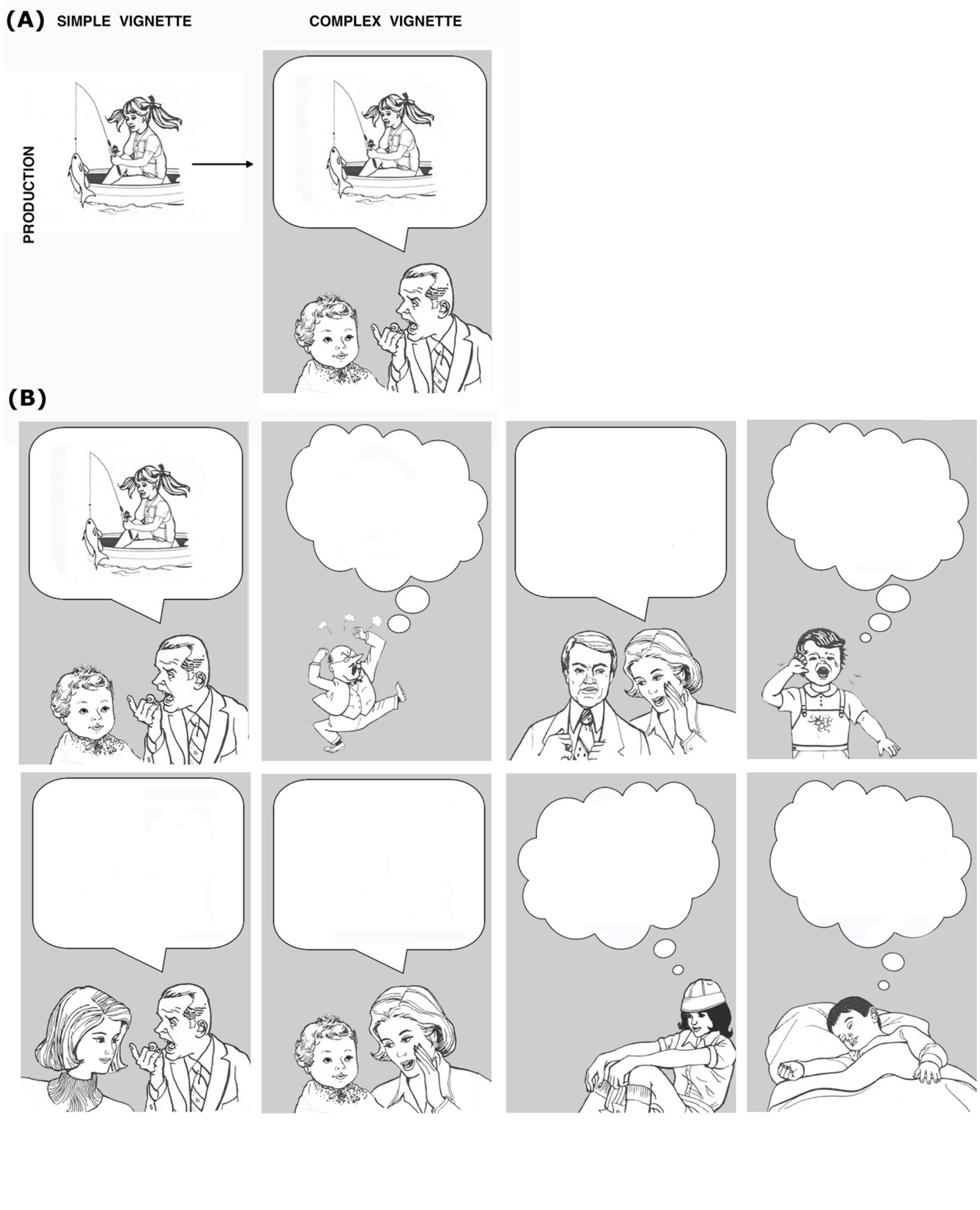
6.1.1 Appendix A1

All the 32 simple vignettes used in the gesture experiments.



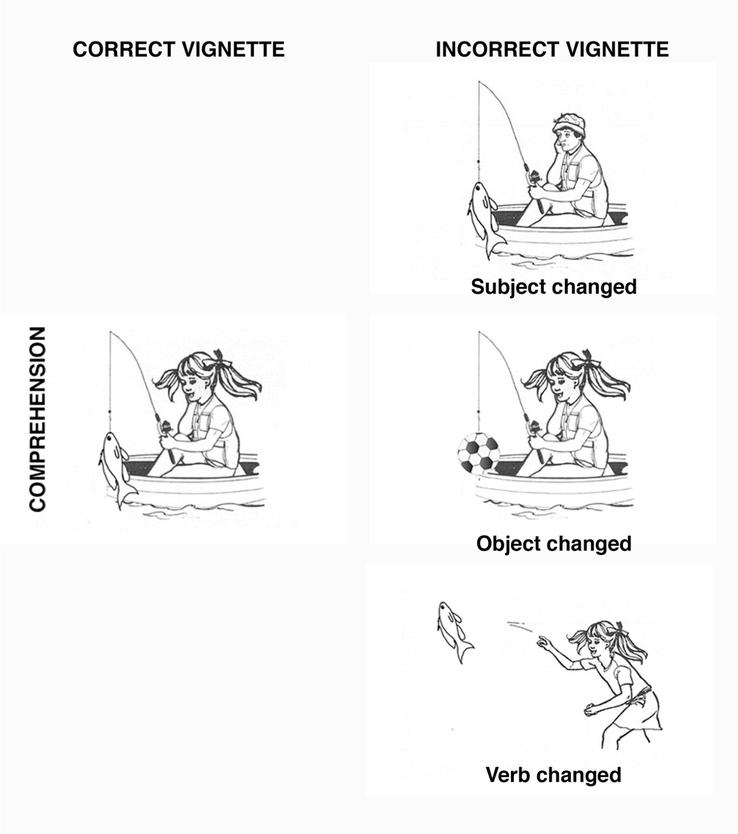
6.1.2 Appendix A2

(A) An example of how a complex vignettes was created by embedding a simple vignette in a complex frame. (B) The eight complex frames into which the 32 simple vignettes were randomly embedded.



6.1.3 Appendix A3

Examples of the correct and incorrect target vignettes in gesture and speech comprehension experiments.



6.1.4 Appendix A4

Turkish and Italian descriptions of the vignettes that were synthesized for the speech comprehension experiment.

1. Kız balığı yakalıyor (Turkish)
   girl fish-ACC catches-PRS-3SG
   La ragazza prende il pesce (Italian)
   girl catch-PRS-3SG fish
   A girl catches a fish (English)

2. Kız topu yakalıyor
   girl ball-ACC catches-PRS-3SG
   La ragazza prende la palla
   girl catch-PRS-3SG ball
   Girl catches a ball

3. Kız balığı atıyor
   girl fish-ACC throw-PRS-3SG
   La ragazza lancia il pesce
   girl throw-PRS-3SG fish
   Girl throws a fish

4. Kız topu atıyor
   girl ball-ACC throw-PRS-3SG
   La ragazza lancia la palla
   girl throw-PRS-3SG ball
   Girl throws a ball

5. Erkek balığı yakalıyor
   boy fish-ACC catch-PRS-3SG

Il ragazzo prende il pesce
boy catch-PRS-3SG fish
Boy catches a fish

6. Erkek topu yakalıyor
boy ball-ACC catch-PRS-3SG
Il ragazzo prende la palla
girl catch-PRS-3SG ball
Boy catches a ball

7. Erkek balığı atıyor
boy fish-ACC throw-PRS-3SG
Il ragazzo lancia il pesce
boy throw-PRS-3SG fish
Boy throws a fish

8. Erkek topu atıyor
boy ball-ACC throw-PRS-3SG
Il ragazzo lancia la palla
boy throw-PRS-3SG ball
Boy throws a ball

9. Adam köpeği okşuyor
man dog-ACC pat-PRS-3SG
Il vecchio accarezza il cane
man pat-PRS-3SG dog
Man pats the dog

10. Adam kediyi okşuyor
man cat-ACC pat-PRS-3SG
Il vecchio accarezza il gatto
man pat-PRS-3SG cat
Man pats the cat

11. Adam köpeği besliyor
man dog-ACC feed-PRS-3SG
Il vecchio nutre il cane
man feed-PRS-3SG dog
Man feeds the dog

12. Adam kediyi besliyor
man cat-ACC feed-PRS-3SG
Il vecchio nutre il gatto
man feed-PRS-3SG cat
Man feeds the cat

13. Maymun köpeği okşuyor
monkey dog-ACC pat-PRS-3SG
La scimmia accarezza il cane
monkey pat-PRS-3SG dog
Monkey pats the dog

14. Maymun kediyi okşuyor
monkey cat-ACC pat-PRS-3SG
La scimmia accarezza il gatto
monkey pat-PRS-3SG cat
Monkey pats the cat

15. Maymun köpeği besliyor
monkey dog-ACC feed-PRS-3SG
La scimmia nutre il cane
monkey feed-PRS-3SG dog
Monkey feeds the dog

16. Maymun kediyi besliyor
monkey cat-ACC feed-PRS-3SG
La scimmia nutre il gatto
monkey feed-PRS-3SG cat

Monkey feeds the cat

17. Kadın arabayı çekiyor
woman carriage-ACC pull-PRS-3SG
La vecchia tira il caretto
woman pull-PRS-3SG carriage
Woman pulls the carriage

18. Kadın atı çekiyor
woman horse-ACC pull-PRS-3SG
La vecchia tira l'unicorno
woman pull-PRS-3SG unicorn
Woman pulls a horse / unicorn

19. Kadın arabayı itiyor
woman carriage-ACC push-PRS-3SG
La vecchia spinge il caretto
woman push-PRS-3SG carriage
Woman pushes a carriage

20. Kadın atı itiyor
woman horse-ACC push-PRS-3SG
La vecchia spinge l'unicorno
woman push-PRS-3SG unicorn
Woman pushes a horse / unicorn

21. Robot arabayı çekiyor
robot carriage-ACC pull-PRS-3SG
Il robot tira il caretto
robot pull-PRS-3SG carriage
Robot pulls the carriage

22. Robot atı çekiyor
robot horse-ACC pull-PRS-3SG

Il robot tira l'unicorno

robot pull-PRS-3SG unicorn

Robot pulls the horse / unicorn

23. Robot arabayı itiyor

robot carriage-ACC push-PRS-3SG

Il robot spinge il caretto

robot push-PRS-3SG carriage

Robot pushes the carriage

24. Robot atı itiyor

robot horse-ACC push-PRS-3SG

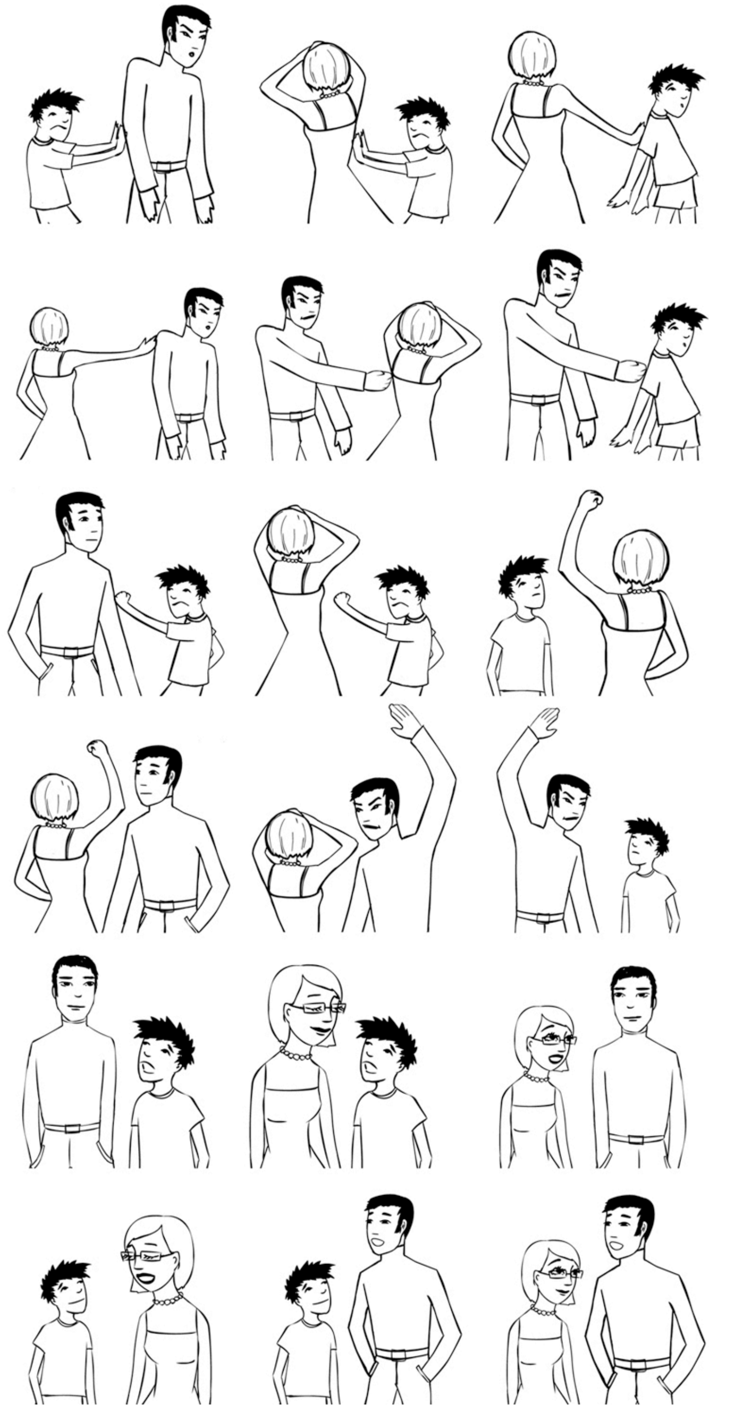
Il robot spinge l'unicorno

robot push-PRS-3SG unicorn

Robot pushes the horse / unicorn

6.2 Appendix B

All the vignettes that were created with a full combinatorial design.



6.3 Appendix C

HABITUATION PHASE
(The same for all experiments)

Phrases	**Sentences**
TO x FE se ko lu	TO x FEMU x GI se se ko ko lu lu
MU x GI se ko lu	BA x PETO x FE se se ko ko lu lu
BA x PE se ko lu	MU x GIBA x PE se se ko ko lu lu

TEST PHASE

Phrases		**Sentences**		
(The same for all experiments)				
RULE	PART	RULE	PART (Exp. 1 & 3)	PART (Exp. 2)
TO x FE mu gi ba	FEMU x se ko lu	BA x PETO x FE mu mu to gi fe ba	BA x PEMU x GI se se ko ko lu lu	GIBA x PETO x se se ko ko lu lu
MU x GI pe to fe	x FEMU se ko lu	MU x GIBA x PE pe mu to to fe fe	MU x GITO x FE se se ko ko lu lu	FEMU x GIBA x se se ko ko lu lu
BA x PE mu to fe	GIBA x se ko lu	TO x FEMU x GI mu pe gi to ba fe	TO x FEBA x PE se se ko ko lu lu	PETO x FEMU x se se ko ko lu lu
* Same "A" and "C" syllables as habituation phrases but a novel "x" syllable that had not occurred in this position before.	x GIBA se ko lu	* Same "A" and "C" syllables as habituation phrases but a novel "x" syllable that had not occurred in this position before.	* Two phrases that occurred in the habituation phase but did not form a sentence.	x PETO x FEMU se se ko ko lu lu
	PETO x se ko lu			x FEMU x GIBA se se ko ko lu lu
	x PETO se ko lu			x GIBA x PETO se se ko ko lu lu
	* The phrases violate the prosodically determined boundary. However, the sequence of syllables did occur in the habitation phase.			* The sentences violate the prosodically determined boundary. However, the sequence of syllables did occur in the habituation phase.

References

Akhtar, N. & Montague, L. (1999). Early lexical acquisition: the role of cross-situational learning. *First Language*, 347–358.

Aldridge, E. (in press). Directionality in word order change in Austronesian languages. In A. Breitbarth, C. Lucas, S. Watts, D. Willis (Eds.), *Continuity and change in grammar*. Amsterdam: John Benjamins.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321-324.

Beach, C.M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of Memory and Language*, 30, 644-663.

Bagou, O., Fougeron, C., & Frauenfelder, U. (2002). Contribution of prosody to the segmentation and storage of "words" in the acquisition of a new mini-language. In B. Bel & I. Marlien (Eds.), *Proceedings of the Speech Prosody 2002 conference* (pp. 59–62). Aix-en-Provence: Laboratoire Parole et Langage.

Baker, M. (2001). Configurationality and polysynthesis. In Haspelmath, M., Ekkehard, K., Wulf, O., & Raible, W. (eds.) *Language Typology and Language Universals*, de Gruyter, Berlin.

Bates, E., Devescovi, A., & D'Amico, S. (1999). Processing complex sentences: A cross-linguistic study. *Language and Cognitive Processes*, 14(1), 69–123.

Bauer, L.M.B. (1995). *The Emergence and Development of SVO Patterning in Latin and French*. Oxford University Press.

Beckman, M. and J. Pierrehumbert (1986) Intonational Structure in Japanese and English. *Phonology Yearbook*, 3, 15-70.

Bickerton, D. (1981). *Roots of language*. Karoma Publishers.

Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7, 173–221.

Bickerton, D. (1992). *Language and species*. Chicago: Chicago University Press.

Bion, R. A. H., Benavides, S., Nespor, M. (in press) Acoustic markers of prominence influence adults' and infants' memory of speech sequences. *Language &*

Speech.

Bock, J., & Warren, R. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, 21, 47-67.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, 5, 341-345.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, 16, 451-459.

Bonatti LL., Peña M., Nespor M., Mehler J. (2006). How to hit Scylla without avoiding Charybdis: comment on Perruchet, Tyler, Galland, and Peereman. *Journal of Experimental Psychology: General*. 135, 314-326.

Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2007). On consonants, vowels, chickens, and eggs. *Psychological Science*, 18, 924-925.

Braine, M. D. S. (1966). Learning the positions of words relative to a marker element. Journal of *Experimental Psychology*.72, 532-540.

Braun, B., Lemhöfer, K., & Cutler, A. (2008). English word stress as produced by English and Dutch speakers: The role of segmental and suprasegmental differences. In *ISCA. Proceedings of Interspeech 2008*.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, 61, 93-125.

Chomsky, N. (2000). Linguistics and brain science. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, Language, Brain*. Cambridge, MA: MIT Press.

Chomsky, N. (1995). *Minimalist Program*. Cambridge, MIT Press.

Chomsky, N. (1986). *Knowledge of Language: Its Nature, Origin, and Use*. Praeger Westport, CT.

Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Mouton de Gruyter.

Chomsky, N. (1980). Principles and parameters in syntactic theory. In N. L. Hornstein & D. Lightfoot (Eds.), *Explanation in Linguistics: The Logical Problem of Language Acquisition* (pp. 32–75). London and New York: Longman.

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

Chomsky, N., & Lasnik, H. (1977). Filters and control. *Linguistic Inquiry*, 8, 425–504.

Christophe, A., Nespor, M., Guasti, M. T., & van Ooyen, B. (2003). Prosodic structure and syntactic acquisition: the case of the head-direction parameter. *Developmental Science*, 6, 211-220.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological Phrase boundaries constrain lexcial access I. Adult data*. Journal of Memory and Language*, 51, 523-547.

Clarke, E.V. (1998). Morphology in language acquisition. In Spencer, A. & Zwicky, A.M. (eds.) *The Handbook of Morphology*. Blackwell, Oxford.

Collier, R., & 't Hart, J. (1975). The role of intonation in speech perception. In A. Cohen, & S.G. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 107-121). Heidlerberg: Springer-Verlag.

Collier, R, de Pijper, J.R., & Sanderman, A.A. (1993). Perceived prosodic boundaries and their phonetic correlates. *Proceedings of the DARPA Wordkshop on Speech and Natural Language* (pp. 341-345). Princeton, NJ, March 21-24.

Collins, C. (1997), *Local Economy*, MIT Press, Cambridge, MA.

Cooper, W.E., & Paccia-Cooper, J. (1980). *Syntax and Speech.* Cambridge: MA:Harvard University Press.

Cooper, W.E. & Sorensen, J.M. (1977). Fundamental frequency contours at syntactic boundaries. *Journal of the Acoustical Society of America*, 62, 682-692.

Cooper, W.E. & Sorensen, J.M. (1981). *Fundamental Frequency in Sentence Production*. New York: Springer-Verlag.

Cutler, A., Dahan, D., & van, W., Donselaar. (1997). Prosody in the comprehension of spoken language: a literature review. *Language and Speech*, 40, 141–201.

Cutler, A. (1993). Phonological cues to open- and closed class words in the processing of spoken sentences. *Journal of Psycholinguistic Research*, 22, 133–142.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, 2, 133–142.

Davis, G. & Driver, J. (1998). Kanizsa subjective figures can act as occluding surfaces at parallel stages of visual search. *Journal of Experimental Psychology: Human Perception and Performance,* 24, 169–184.

de Rooij JJ. (1976). Perception of prosodic boundaries. *IPO Annual Progress Report*, 11, 20–24.

Dryer, M.S. (1989). Large linguistic areas and language sampling. *Studies in Language*, 13(2), 257-292.

Dryer, M.S. (2005). The order of subject, object and verb. In Haspelmath, M., Dryer, M.S., Gil, D., & Comrie, B. (Eds.). *The World Atlas of Language Structures*, Oxford, Oxford University Press, 330-333.

Dutoit, T. (1997). *An Introduction to Text-to-Speech Synthesis*. Dordrecth, The Netherlands: Kluwer Academic Publishers.

Dutoit, T., Pagel, V., Pierret, N., Bataille, F., & Van Der Vreken, O. (1996). The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes, *Proc. ICSLP'96*, Philadelphia, vol. 3, 1393-1396.

Endress, A.D., Dehaene-Lambertz, G., Mehler, J. (2007) Perceptual Constraints and the Learnability of Simple Grammars. *Cognition*, 105(3), 577-614.

Endress, A.D. & Hauser, M.D. (2010). Syntax-induced pattern deafness. *Proceedings of the National Academy of Sciences of the United States of America*. 106(49), 21001-21006.

Endress, A.D. & Mehler, J. (2009). Primitive Computations in Speech Processing. *Quarterly Journal of Experimental Psychology*, 62, 2187–2209.

Endress, A.D. & Mehler, J. (2010). Perceptual Constraints in Phonotactic Learning. *Journal of Experimental Psychology: HP&P.* Vol. 36(1), 235-250.

Endress, A.D., Nespor, M. & Mehler, J.(2009). Perceptual and Memory Constraints on Language Acquisition. *Trends in Cognitive Science*, 13, 348-353.

Endress, A.D., Scholl, B.J., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*. 134, 406-419.

Epstein, S.D. & Hornstein, N. (1999). *Working Minimalism*, MIT Press, Cambridge, MA.

Erdocia, K., Laka, I., Mestres-Missé, A., & Rodriguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word order language: Behavioral and electrophysiological evidences from Basque. *Brain and Language*, 109, 1-17.

Erguvanli, E. (1984). The function of word-order in Turkish grammar. *Publications in Linguistics* (Vol. 6). University of California.

Everett, D. (2005). Cultural constraints on grammar and cognition in Pirahã: Another look at the design features of human language. *Current Anthropology,* 46, 621–46.

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. Trends in Cognitive Sciences, 8, 307–314.

Fernald, A. & McRoberts, G.W. (1995). Prosodic bootstrapping. A critical analysis of the argument and the evidence. In: J. Morgan & K. Demuth (Eds.), *Signal to syntax*. Hillsdale, NJ: Erlbaum.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science,* 303, 377-380.

Fitch, W.T., Hauser, M.D. & Chomsky, N/ (2005). The evolution of the language faculty: Clarifications and implications. *Cognition*, 97, 179–210

Fodor, J. A. (1983). The modularity of mind. Cambridge, MA: Bradford Books. MIT Press.

Frazier, L., & Rayner, K. (1988). Parametrizing the language processing system: Left- vs. right-branching within and across languages. In J. Hawkins (Ed.), *Explaining Language Universals*. Oxford: Basil Blackwell.

Gentner, D., & Boroditsky, L. (2009). Early acquisition of nouns and verbs: Evidence from Navajo. In V. C. Mueller Gathercole (Ed.), *Routes to Language: Studies in Honor of Melissa Bowerman* (pp. 5–32). Taylor & Francis: New York.

Gentner, T. Q., Fenn, K. M., Margoliash, D., & Nusbaum, H. C. (2006). Recursive syntactic pattern learning by songbirds. *Nature*, 440, 1204-1207.

Gershkoff-Stowe, L., & Goldin-Meadow, S. (2002). Is there a natural order for expressing semantic relations? *Cognitive Psychology*, 45, 375–412.

Gervain, J., & Werker, J.F. (2008). Frequency and prosody boostrap word order: A cross-linguistic study with 7-month-old infants. In: *The 33rd Boston University Conference on Language Development*, Boston, MA.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler J. (2008). Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cognitive Psychology, 57*, 56-74.

Gervain, J., Macagno, F., Cogoi, S., Peña, M., Mehler, J. (2008). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences of the United States of America*, 105(37), 14222-14227.

Goldin-Meadow, S. (1982). The resilience of recursion: A study of a communication system developed without a conventional language model. In E. Wanner & L. Gleitman (Eds.), *Language Acquisition: The State of the Art* (pp. 51–78). Cambridge University Press: Cambridge.

Goldin-Meadow, S. (2005). Watching language grow. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2271–2272.

Goldin-Meadow, S., & Feldman, H. (1977). The development of language-like communication without a language model. *Science*, 197, 401–403.

Goldin-Meadow, S., & Mylander, C. (1983). Gestural communication in deaf children: Non-effect of parental input on language development. *Science*, 221, 372–374.

Goldin-Meadow, S., & Mylander, C. (1998). Spontaneous sign systems created by deaf children in two cultures. *Nature*, 391, 279–281.

Goldin-Meadow, S., So, W. C., Ozyürek, A., & Mylander, C. (2008). The natural order of events: How speakers of different languages represent events nonverbally. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 9163–9168.

Goudbeek, M., Cutler, A., & Smits, R. (2008). Supervised and unsupervised learning of multidimensionally varying non-native speech categories, *Speech Communication*, 50, 109-125.

Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological Phrase boundaries constrain lexical access II. Infant data. *Journal of Memory and Language*, 51, 548-567.

Green, T.R.G. (1979). The necessity of syntax markers: Two experiments with artificial languages. *Journal of Verbal Learning and Verbal Behavior*, 18, 481-496.

Greenberg, J. H. (1978). *Universals of Human Language*. Syntax. Stanford, California: Stanford University Press.

Greenberg, J. H. (1963). *Universals of Languages*. Cambridge, MIT Press.

Haider, H. (2000). OV is more basic than VO. In P. Svenonius (Ed.), *The Derivation of VO and OV* (pp. 45–67). Amsterdam: Benjamins.

Hauser, M. D., Chomsky, N., & Fitch, T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.

Harris, Z.S. (1955). From phoneme to morpheme. *Language*, 31, 190-222.

Havy, M., & Nazzi, T. (2009). Better processing of consonantal over vocalic information in word learning at 16 months of age. *Infancy*, 14, 439-456.

Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge

University Press: Cambridge.

Hay, J., & Diehl, R. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, 69, 113-122.

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. Chicago: The University of Chicago Press.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and Phonology, Vol 1: Rhythm and Meter* (pp. 201–260). San Diego: Academic Press.

Hayes, J.R. & Clark, H.H. (1970). Experiments in the segmentation of an artificial speech analog. In J. R. Hayes (Ed.), *Cognition and the Development of Language.* New York: Wiley.

Heine, B. & Kuteva, T. (2005). *Language Contact and Grammatical Change.* Cambridge: Cambridge University Press.

Hirst, D. (1993). Detaching intonational phrases from syntactic structure. *Linguistic Inquiry*, 24, 781-788.

Hirsh-Pasek, K., Kemler Nelson, D., Jusczyk, P. W., Wright, K., Druss, B., & Kennedy, L. J. (1987). Clauses are perceptual units for young infants. *Cognitive Psychology*, 24, 252–293.

Hochmann, J-R. Endress A.D., & Mehler, J. (2010). Word frequency as a cue to identify function words in infancy. *Cognition*, 115, 444-457.

Hochmann, J.R., Azadpour, M. Mehler J. (2008). Do humans really learn AnBn artificial grammars from exemplars? *Cognitive Science*, 32, 1021-1036.

Hochmann, J.R., Benavides-Varela, S., Nespor, M., & Mehler, J. (submitted). Consonants and Vowels, Different Roles in Early Language Acquisition.

Holmberg, A., & Platzack, C. (1995). *The Role of Inflection in Scandinavian Syntax*. Oxford University Press.

Hróarsdóttir, T. (2000). *Word Order Change in Icelandic: From OV to VO*. Amsterdam: John Benjamins.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing unpredictable variation: The roles of adult and child learners in language formation and change. *Language Learning and Development*, 1, 151–195.

Inkelas, S., & Zec, D. (1990). *The Phonology-Syntax Connection*. Chicago: The University of Chicago Press.

Jackendoff, R. (1992). *Semantic Structures*. Cambridge: MIT Press.

Jackendoff, R. (1997). *The Architecture of the Language Faculty*. Cambridge: MIT Press.

Jacob, F. (1977). Evolution and tinkering. *Science*, 196, 1161–1166.

Jansen, B., Lalleman, J., & Muysken, P. (1981). The alternation hypothesis: Acquisition of Dutch word order by Turkish and Moroccan foreign workers. *Language Learning*, 31, 315–336.

Jusczyk, P. W. (1998). Dividing and conquering the linguistic input. In M. C. Gruber, D. Higgins, K. Olson, & T. Wysocki (Eds.), *CLS 34*, Vol.2: The panels (pp. 293–310). Chicago: University of Chicago.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress pattern of English words. *Child Development,* 64, 675– 687.

Jusczyk, P. W., Hirsh-Pasek, K., Kemler Nelson, D., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252–293.

Jusczyk, P. W., Hohne, E., & Mandel, D. (1995). Picking up regularities in the sound structure of the native language. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 91–119). Timonium, MD: York Press.

Kayne, R. S. (1994). *The Antisymmetry of Syntax*. Cambridge, MIT Press.

Kayne, R. S. (2004). Antisymmetry and Japanese. In J. Jenkins (Ed.). *Variation and Universals in Biolinguistics* (Vol. 62, pp. 3–35). Elsevier Science.

Kegl, J. (2008). The case of signed languages in the context of pidgin and creole studies. In S. Kowenberg & V. Singler (Eds.), *The Handbook of Pidgin and Creole Studies*. Wiley-Blackwell.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The headturn preference procedure for testing auditory perception. *Infant Behavior and Development*, 18, 111–116.

Kiparsky, P. (1996). The shift to head-initial VP in Germanic. In H. Thrainsson, J. Peter, & S. Epstein (Eds.), *Comparative Germanic Syntax*. Kluwer.

Klatt, D.H. (1974). The duration of [s] in English words. *Journal of Speech and Hearing Research*, 17, 51-63.

Klatt, D.H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, 59, 1208-1221.

Klibanoff, R.S. & Waxman, S.R. (2000). Basic level object categories support the acquisition of novel adjectives: evidence from pre-school aged children. *Child Development*, 71(3), 649–659.

Klima, E. & Bellugi, U. (1979). *The Signs of Language.* Cambridge University Press.

Knops, A., Thirion, B., Hubbard, E. M., Michel, V., & Dehaene, S. (2009). Recruitment of an area involved in eye movements during mental arithmetic. *Science*, 324, 1583–1585.

Kovács, Á.M. & Endress, A.D. (under review). Seven-month-olds learn hierarchical "grammars".

Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception and Psychophysics*, 31, 279–292.

Kuno, S. (1973). *The Structure of the Japanese Language*. Cambridge, MIT Press.

Kucera, H., & Francis, N. (1967). *A computational Analysis of Present-day American English*. Providence, RI: Brown University Press.

Langus, A. & Nespor, M. (2010). Cognitive systems struggling for word order. *Cognitive Psychology*, 60, 291–318.

Lehiste, I. (1970). *Suprasegementals*. Cambridge: MIT Press.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa*, 7, 107-122.

Lehiste, I, Olive, J.P., & Streeter, L. (1976). Role of duration in disambiguating syntactically ambiguous sentences. *Journal of the Acoustical Society of America*, 60, 1199-1202.

Lehmann, W. P. (1973). A structural principle of language and its implications. *Language*, 49, 42–66.

Lehmann, W. P. (1978). The great underlying ground-plans. In W. P. Lehmann (Ed.), *Syntactic Typology* (pp. 3–55). University of Texas Press.

Leinonen, M. (1980). A closer look at natural serialization. *Nordic Journal of Linguistics*, 3, 147–159.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. Wiley.

Li, C. N. (1977). *Mechanisms of Syntactic Change*. Austin: University of Texas Press.

Li, Y. H. A. (1990). *Order and Constituency in Mandarin Chinese*. Dordrecht: Kluwer Academic Publishers.

Li, C. N., & Thomson, A. (1974). An explanation of word order change SVO?SOV.

Foundations of Language, 12, 201–214.

Light, T. (1979). Word order and word order change in Mandarin. *Journal of Chinese Linguistics*, 7, 149-180.

MacWhinney, B., Osmán-Sági, J., & Slobin, D. I. (1991). Sentence comprehension in aphasia in two clear case-marking languages. *Brain and Language*, 41, 234–249.

Marchetto, E., & Bonatti L.L. (in prep). Infants' discovery of words and grammar-like regularities from speech requires distinct processing mechanisms.

Marchetto, E., & Bonatti L.L. (under review). Finding Words and Rules in a Speech Stream at 7 and 12 Months.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule-learning in seven-month-old infants. *Science*, 283, 77-80.

Marslen-Wilson, W., & Tyler, L. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1–71.

Mazuka, R., Itoh, K., Kiritani, S., Niwa, S., Ikejiru, K., & Naito, K. (1989). Processing of Japanese Garden Path, center-embedded, and multiply-left embedded sentences: Reading time data from an eye movement study. *Annual Bulletin of the Research Institute of Logopedics and Phoniatrics*, 23, 187–212.

Millotte, S., Frauenfelder, U.H., & Christophe, A. (2007). Phrasal prosody constraints lexical access. *AmLap – 13th Annual Conference on Architectures and Mechanisms for Language Processing*, Turku, Finland.

Millotte, S., Rene, A., Wales, R., & Christophe, A. (2008). Phonological Phrase boundaries constrain the online syntactic analysis of spoken sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43, 874-885.

Miyagawa, S. (1996). Word Order Restrictions and Nonconfigurationality. *Proceedings of Formal Approaches to Japanese Linguistics 2*, *MIT Working Papers in Linguistics 29*, 117-142.

Moeser, S.D. & Bregman, A.S. (1972). The role of references in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior*, 11, 759-769.

Moeser, S.D. & Bregman, A.S. (1973). Imagery and language acquisition. *Journal of Verbal Learning and Verbal Behavior*, 12, 91-98.

Morgan, J.L., Meier, R.P., Newport, E.L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology,* 19, 498-550.

Morgan, J. L., Swingley, D., & Miritai, K. (1993). Infants listen longer to speech with extraneous noises inserted at clause boundaries. *Paper presented at the Biennial Meeting of the Society for Research in Child Development*, New Orleans, LA.

Moro, A. (2000). *Dynamic Antisymmetry.* Linguistic inquiry monograph series (Vol. 38). Cambridge, MA: MIT Press.

Morse, P. A. (1972). The discrimination of speech and nonspeech stimuli in early infancy. *Journal of Experimental Child Psychology*, 13, 477–492.

Murphy, R.A., Mondragon, E., & Murphy, V.A. (2008). Rule learning by rats. *Science,* 319, 1849–1851.

Muysken, P. (1988). Are creoles a special type of language? In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (pp. 285–302). Cambridge University Press: Cambridge.

Nagata, H. (1981). Effectiveness of word order and grammatical marker as syntactic indicators of semantic relations. *Journal of Psycholinguistic Research*, 10(5), 471-486.

Nagata, H. (1983). Effectiveness of word order and grammatical markers as syntactic indicators of semantic relations in opaque input conditions. *Journal of Psycholinguistic Research*, 12(2), 157-169.

Nagata, H. (1984). Effectiveness of word order over grammatical markers as a syntactic indicator of semantic relations in an opaque partial description situation. *Journal of Psycholinguistic Research*, 13(4), 281-293.

Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, 98, 13-30.

Nazzi, T., & Bertoncini, J. (2009). Consonant specificity in onset and coda positions in early lexical acquisition. *Language and Speech* 52.

Nazzi, T., Kemler Nelson, D. G., Jusczyk, P. W., & Jusczyk, A. M. (2000). Six month olds_ detection of clauses embedded in continuous speech: Effects of prosodic well-formedness. *Infancy*, 1, 123–147.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue & Linguaggio*, 2, 201-227.

Nespor, M., Shukla, M., van de Vijver, R., Avesani, C., Schraudolf, H., & Donati, C. (2008). Different phrasal prominence realizations in VO and OV languages. *Lingue e Linguaggio,* 2, 1-29.

Nespor, M., & Vogel, I. (2008). *Prosodic Phonology*. Berlin: Mouton de Gruyter. 1st edition 1986. Dordrecht. Foris.

Nevins, A., Pesetsky, D., & Rodrigues, C. (2009). Piraha Exceptionality: A Reassessment. *Language,* 85, 355-404.

Newman, A.J., Supalla, T., Hauser, P., Newport, E.L., & Bavelier, D. (2010). Dissociating neural subsystems fro grammar by contrasting word order and inflection. *Proceedings of the National Academy of Sciences of United States of America*, 107, 7539-7544.

Newmeyer, F. J. (2000). On the reconstruction of 'Proto-world' word order. In Knight, C., Hurford, J. R., & Studdert-Kennedy, M. (Eds.). *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*. Cambridge, Cambridge University Press.

Newport, E., L., & Aslin, R., N. (2004). Learning at a distance. I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, 48, 127–162.

Nooteboom, S.G., Brokx, J.P.L., Rooij, J.J. de (1978). Contributions of prosody to speech perception. In W.J.M. Levelt & G.B. Flores d'Arcais (Eds.), *Studies in the perception of language* (pp. 75-107). Chichester: John Wiley & Sons.

Odlin, T. (1989). *Language Transfer: Cross-linguistic Influence in Language Learning*. University of Cambrdige Press: Cambridge.

Oller, D.K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, 54, 1235-1246.

O'Shaughnessy, D. (1979). Linguistic features in fundamental frequency patterns. *Journal of Phonetics*, 7, 119-145.

Peña, Bonatti, Nespor, & Mehler (2002). Signal driven computations in language processing, *Science*, 298, 604-607.

Perruchet, P., Tyler, M. D., Galland, N., & Peereman, R. (2004). Learning nonadjacent dependencies: No need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133, 573–583.

Pierrehumbert, J., & Hirschberg, J. (1990). The meaning of Intonational contours in the interpretation of discourse. In P. Cohen, J. Morgan, & M. Pollack (Eds.), *Intentions in communication*. Cambridge, MA: The MIT Press.

Pinker, S., & Jackendoff, R. (2005). The faculty of language: What's special about it? *Cognition*, 95, 201–236.

Price, P.J., Ostendorf, M., Shattuck-Hufnagel, S., & Fong, C. (1991). The use of prosody in syntactic disambiguation. *Journal of the Acoustical Society of America*, 90, 2956-2970.

Rooij, J.J. de (1975). Prosody and the perception of syntactic boundaries. *IPO Annual Progress Report*, 10, 36-39.

Rooij, J.J. de (1976). Perception of prosodic boundaries. *IPO Annual Progress Report*, 11, 20-24.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35, 606-621.

Saffran, J.R., Aslin, R.N., & Newport, E.L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1925-1928.

Saffran, J. R. & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, 4, 273-284.

Sandler, W., Meir, I., Padden, C., & Aronoff, M. (2005). The emergence of grammar: Systematic structure in a new language. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 2661–2665.

Scott, D.R. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America*, 71, 996-1007.

Selkirk, E. (1984). *Phonology and Syntax: The Relation Between Sound and Structure.* Cambridge, MA: The MIT Press.

Selkirk, E. (1996). The prosodic structure of function words. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. (pp. 187–213). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Senghas, A. (2005). Language emergence: Clues from a new Bedouin sign language. *Current Biology*, 15(12), 463-465.

Senghas, A., & Coppola, M. (2001). Children creating language: How Nicaraguan Sign Language acquired a spatial grammar. *Psychological Science*, 12(4) 323-328.

Senghas, A., Coppola, M., Newport, E. L., & Supalla, T. (1997). Argument structure in Nicaraguan sign language: The emergence of grammatical devices. In E. Hughes & A. Greenhill (Eds.), *Proceedings of the Boston University Conference on Language Development* (pp. 550–561). Cascadilla Press: Somerville.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: evidence from an emerging sign language in Nicaragua. *Science*, 305, 1779-1782.

Shattuck-Hufnagel, S. & Turk, A.E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193-247.

Shukla, M., Nespor, M., & Mehler, J. (2007). An interaction between prosody and statistics in the segmentation of fluent speech. *Cognitive Psychology*, 54, 1-32.

Singleton, J. L., & Newport, E. L. (2004). When learners surpass their models: The acquisition of American sign language from inconsistent input. *Cognitive Psychology*, 49, 370–407.

Smith, L.B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 333-338.

Slobin, D.I., & Bever, T.G. (1982). Children use canonical sentence schemas: A crosslinguistic study of word order and inflections. *Cognition*, 12, 229-265.

Soderstrom, M., Seidl, A., Kemler Nelson, G., & Jusczyk, P.W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49, 249-267.

Steedmna, M. (1990). Syntax and intonational structure in a combinatory grammar. In G.T.M. Altmann (ed.), *Cognitive Models of Speech Processing: Psycholinguistic and Computational Perspectives* (pp. 457-482). Cambrdige, MA: MIT Press.

Soderstrom, M., Seidl, A., Kemler Nelson, D.G., & Jusczyk, P.W. (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, 49, 249-267.

Staal, J. F. (1967). *Word Order in Sanskrit and Universal Grammar*. Dordrecht, The Netherlands.

Steele, S. (1978). Word order variation: A typological study. In J. H. Greenberg, C. A. Ferguson, & E. A. Moravcsik (Eds.), *Universals of Human Language. Syntax*

(Vol. 4, pp. 585–623). Stanford, CA: Stanford University Press.

Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America,* 64, 1582-1592.

Sun, C. F., & Givón, T. (1985). On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implications. *Language*, 61, 329–351.

Taylor, A. (1994). The change from SOV to SVO in ancient Greek. *Language Variation and Change*, 6, 1-37.

Thiessen, E.D. & Saffran, J.R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*. 39, 706-716.

Tsujimura, N. (1999). *The Handbook of Japanese Linguistics*. Malden, MA: Blackwell Publishers.

Umeda, N. (1977). Consonant duration in American English. *Journal of the Acoustical Society of America*, 61, 846-858.

Vennemann, T. (1974). Analogy in generative grammar: The origin of word order. In Proceedings of the eleventh international congress of linguists (1972) (pp. 79–83).

Vennemann, T. (1976). Categorial grammar and the order of meaningful elements. In A. Juilland (Ed.), Linguistic Studies Offered to Joseph Greenberg on the Occasion of his Sixtieth Birthday (pp. 615–634). Saratoga, California: Anma Libri.

Vaissiere, J. (1974). On French prosody. *Quarterly Progress Report, M.I.T*, 114, 212-223.

Vaissiere, J. (1975). Further note on French prosody. *Quarterly Progress Report, M.I.T*, 115, 251-262.

Watson, D., & Gibson, E. (2004). The relationship between Intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, 19(6), 713–755.

Wightman, C.W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P.J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707-1717.

Yang, C. (2004). Universal Grammar, statistics or both. *Trends in Cognitive Sciences*, 8, 451-456.

Yoshida, K.A., Iversen, J.R., Patel, A.D., Mazuka, R., Nito, H., Gervain, J., Werker, J.F. (in press). The development of perceptual grouping biases in infancy: A Japanese-English cross-linguistic study. *Cognition*.

Yu, C., & Smith, L.B. (2007). Rapid Word Learning Under Uncertainty via Cross-Situational Statistics. *Psychological Science*, 18(5), 414-420.