DISCOVERING WORDS AND RULES FROM SPEECH INPUT: AN INVESTIGATION INTO EARLY MORPHOSYNTACTIC ACQUISITION MECHANISMS

Candidate:
Erika Marchetto

Supervisor:
Luca Lorenzo Bonatti

Thesis submitted for the degree of Doctor Philosophiae
in Cognitive Neuroscience
Trieste, December 2009

Scuola Internazionale di Studi Superiori Avanzati –
International School for Advanced Studies

Jury

Anne Christophe, Laboratoire de Sciences Cognitives et Psycholinguistique (LSCP), EHESS / CNRS / DEC-ENS, Paris, France.

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.

Tim Shallice, Cognitive Neuroscience Sector, International School for Advanced Studies (SISSA/ISAS), Trieste, Italy.

Acknowledgments

This PhD thesis is the result of both my work and the generous support I have received from many people along the doctoral years.

First, I would like to thank my advisor, Luca Lorenzo Bonatti, and the director of the Language, Cognition and Development Lab, Jacques Mehler, for providing me an excellent scientific environment. They offered me the most adequate technological tools and all the facilities I needed to conduct the infant studies. I'm really grateful to them for their monetary support that allowed me to participate to several international conferences and meetings – a really great opportunity to present and discuss my studies at different stages of their preparation.

I am extremely grateful to all the members of the Language, Cognition and Development Lab at SISSA for all their scientific advices, for the insightful comments they made on my work, and for the friendly and collaborative atmosphere we had in the lab: Jacques Mehler, Luca L. Bonatti, Marina Nespor, Marcela Peña, Agnes M. Kovács, Judit Gervain, Ansgar D. Endress, Erno Téglás, Juan Manuel Toro, Alan Langus, Mohinish Shukla, Damir Kovacíc, Silvia Benavides, David Maximiliano Gómez, Jean-Remy Hochmann, Luca Coletti, Mahan Azadpour, Ricardo Bion Hoffman.

Special thanks to Juan Manuel Toro, Agnes M. Kovács, Ansgar D. Endress, Judit Gervain, and Alan Langus for very critical discussions of my experiments, and for helping me with the writing of the thesis and of other scientific reports. A big thank to Agnes M. Kovács for her friendship and emotional support.

I would like to thank Alessio Isaja for his technical support, as well as Marijana Sjekloca and Francesca Gandolfo for helping recruiting infant participants.

While designing and conducting the studies presented in the thesis, I could benefit from discussions with different people. I am grateful to all of them. In particular, I would like to thank Janet Werker, Richard Aslin, Anne Christophe, Lila and Henry Gleitman.

I am deeply indebted with all the infants participating to my studies and their wonderful parents, most of them coming to the laboratory several times with their kids at different ages -- some of them also coming back with their second or even third child. Thanks to them for contributing to the advance my work, and for believing in scientific research.

I'm grateful to Valerie Lesk, Padhu, Andrea della Chiesa, and Raffaella Ida Rumiati for trusting my motivation, and directing my work in a fruitful direction.

A special thank to Gianluca for staying beside me, with love and patience, from the first page I read to take my first exam at the university, to the last page I wrote for this PhD thesis.

Summary

To acquire language proficiently, learners have to segment fluent speech into units – that is, words -, and to discover the structural regularities underlying word structure. Yet, these problems are not independent: in varying degrees, all natural languages express syntax as relations between nonadjacent word subparts. This thesis explores how developing infants come to successfully solve both tasks. The experimental work contained in the thesis approaches this issue from two complementary directions: investigating the computational abilities of infants, and assessing the distributional properties of the linguistic input directed to children.

To study the nature of the computational mechanisms infants use to segment the speech stream into words, and to discover the structural regularities underlying words, I conducted seventeen artificial grammar studies. Along these experiments, I test the hypothesis that infants may use different mechanisms to learn words and word-internal rules. These mechanisms are supposed to be triggered by different signal properties, and possibly they become available at different stages of development. One mechanism is assumed to compute the distributional properties of the speech input. The other mechanism is hypothesized to be non-statistical in nature, and to project structural regularities without relying on the distributional properties of the speech input.

Infants at different ages (namely, 7, 12 and 18 months) are tested in their abilities to detect statistically defined patterns, and to generalize structural regularities appearing inside word-like units. Results show that 18-month-old infants can both extract statistically defined sequences from a continuous stream (Experiment 12), and find internal-word rules only if the familiarization stream is segmented (Experiments 13 and 14). Twelve-month-olds can also segment words

from a continuous stream (Experiment 5), but they cannot detect word-straddling sequences even if they are statistically informative (Experiments 15 and 16). In contrast, they readily generalize word-internal regularities to novel instances after exposure to a segmented stream (Experiments 1-3 and 17), but not after exposure to a continuous stream (Experiment 4). Instead, 7-month-olds do not compute either statistics (Experiments 10 and 11) or within-word relations (Experiments 6 and 7), regardless of input properties. Overall, the results suggest that word segmentation and structural generalization rely on distinct mechanisms, requiring different signal properties to be activated --that is, the presence of segmentation cues is mandatory for the discovery of structural properties, while a continuous stream supports the extraction of statistically occurring patterns. Importantly, the two mechanisms have different developmental trajectories: generalizations became readily available from 12 months, while statistical computations remain rather limited along the first year.

To understand how the computational selectivities and the limits of the computational mechanisms match up with the limitations and the properties of natural language, I evaluate the distributional properties of speech directed to children. These analyses aim at assessing with quantitative and qualitative measures whether the input children listen to may offer a reliable basis for the acquisition of morphosyntactic rules. I choose to examine Italian, a language with a rich and complex morphology, evaluating whether the word forms used in speech directed to children would provide sufficient evidence of the morphosyntactic rules of this language. Results show that the speech directed to children is highly systematic and consistent. The most frequently used word forms are also morphologically well-formed words in Italian: thus, frequency information correlates with structural information -- such as the morphological structure of words. While a statistical analysis of the speech input may provide a small set of words occurring with high frequency, how learners come to extract structural properties from them is another problem. In accord with the results of the infant studies, I propose that structural generalizations are projected on a different basis than statistical computations.

Overall, the results of both the artificial grammar studies an the corpus

analysis are compatible with the hypothesis that the tasks of segmenting words from fluent speech, and that of learning structural regularities underlying word structure rely on statistical and non-statistical cues respectively, placing constraints on computational mechanisms having different nature and selectivities in early development.

Contents

CHAPTER 6: ON THE NATURE OF GENERALIZATIONS AND STATISTICAL COMPUTATIONS: ARTIFICIAL GRAMMAR STUDIES WITH 12- AND 18- MONTH-OLD INFANTS 93

List of Figures

List of Tables

CHAPTER 1

FROM LINGUISTIC SIGNAL TO LINGUISTIC COMPETENCE

Language acquisition is taken for granted because it usually happens so effortlessly. Children normally start understanding and producing their first words and utterances much before the preschool age, and they do so accurately and without experiencing any particular difficulty. In spite of the easiness with which young learners face the task, acquiring a language is a nontrivial problem: language is a rich and complex system consisting of thousands of elements, which combine together to generate a virtually infinite number of novel structures.

There are at least three facts about language acquisition that make it an intriguing problem to study. First, children are not explicitly told what the linguistic elements and the rules governing the creation of novel structures are. Instead, they have somehow to figure out both of them, and they have to do so just by listening to the speech signal. Second, all normal children pass the same milestones in linguistic development, independently from their specific linguistic background. They succeed at acquiring language within their first years of life, and they seem to be in a privileged situation, as they learn language even better than adults, although they possess limited cognitive resources (Newport, 1990). Third, only human beings succeed at acquiring language and at using it productively. The other animals possess communications systems, but none of them is a system that conveys an unlimited set of brand-new structures using a limited set of elements – as instead language does (Hauser, Chomsky, & Fitch, 2002).

How language is acquired and why it is a faculty unique to humans are some of the most fascinating questions addressed by developmental psychologists, psycholinguists and linguists. The latter elaborated the most sophisticated formal theories to account how this unique competence arises specifically in humans, so accurately, effortlessly and without over instruction in all (healthy) children that are exposed to a linguistic input from birth on.

In the view advocated by Noam Chomsky, language acquisition is the result of an innate, hard-wired predisposition to evolve language into a final, mature state. This innate form of knowledge constraining language acquisition, or *Universal Grammar* (Chomsky, 1957; 2000) is assumed to be encoded in the biological endowment of our species. It is hypothesized to be the force determining why children converge on the same grammar as the other members of their linguistic community, independently from the specific language they are exposed to. Importantly, such grammar would allow language learners to go beyond the input they had been exposed to, and to generate novel structures. Universal Grammar is assumed to act as a *constraint* to the learners' expectations about what forms possible grammatical structures can take.

However, it is also clear that, no matter how much of knowledge the child has through Universal Grammar, language has to be learned. Phonology, lexicon, and grammar vary from language to language, and they must be learned on the basis of linguistic experience. In other words, both endowment and learning contribute to language acquisition, the result of which is extremely sophisticated body of linguistic knowledge. Consequently, both must be taken in account, explicitly, in a theory of language acquisition.

Controversies arise when it comes to the relative contributions from innate knowledge and experience-based learning. Some researchers approach language acquisition by characterizing the scope and limits of innate principles of Universal Grammar that govern the world's languages (Marcus, 1998; Pinker, 1984). Others tend to emphasize the role of experience and the child's domain-general learning ability (Tomasello, 2000; Gómez & Gerken, 2001; Mintz, 1997). Such a controversy stems from the division of labor between endowment and learning: things that are built in need not be learned, and things that can be garnered from experience need not be built in.

Assessing infants' early abilities may illuminate our understanding of the cognitive mechanisms involved in language acquisition, of their nature and their limitations. Conversely, documenting the properties of the linguistic input would help establishing to what extent the learning situations offer the optimal conditions for language acquisition. The empirical investigations of the present thesis will endorse both approaches to the purpose of understanding how infants extract both words and structural regularities inherent in words from fluent speech. The next chapter will describe in what sense words may contain structural information, and it will offer a general plan of the experiments reported in the thesis.

CHAPTER 2

WORDS AND MORPHOSYNTACTIC RULES IN A SEA OF SOUNDS

The auditory world provides a rich source of information to be acquired by the developing infant. However, such information is not readily available from the speech input. To understand the difficulty of isolating the relevant information to consider, and the amazing complexity of the system to learn, imagine for a while to be a young infant listening to the following speech sequence:

(1) *…theunfriendlyboyswiththebrownhairareplayingwiththeredballs…*

In order to make some sense of this sentence, you will have to break the continuous stream up into its constituent units, such as words. This task is as necessary as difficult to achieve, given that word boundaries are not clearly marked in speech by salient and obvious cues (Cole, & Jakimik, 1980).

After you will solve this problem, you will have identified distinct elements, such as:

(2) *the, unfriendly, boys, with, the, brown, hair, are, playing, with, the, red, balls*

But language does not merely consist of a sequence of words; instead, it is a system that combines elements together into structures. For instance, morphemes organize together to generate words, and words combine together to generate sentences. The organization of units into structures is governed by rules. A *productive* use of language, i.e., the ability to generate and understand an infinite number of structures from a finite set of elements, demands learners to master the rules of a language (Chomsky, 1957; Hauser, Chomsky, & Fitch, 2002).

Hence, the next problem you will have to solve is to map structural relations into the linguistic units. In English, structural relations - both *between* and *within* words - are expressed through grammatical morphemes, i.e., word sub-parts that mark grammatical distinctions and relations. For instance, in sentence (1), the subject "the boys" and the verb "are playing" are syntactically related each other through the suffix "-s", which is added to the stem "boy" to mark plurality; accordingly, the auxiliary "to be" takes its plural form, i.e. "are" ("the boys are playing", underline font marks number agreement). Similarly, the verb "to play" takes the inflection "-ing" to express progressive aspect together with the auxiliary "are" ("are playing"). Therefore, infants must discover that words are not the ultimate units to consider. Instead, words should be analyzed into smaller pieces expressing morphosyntactic relations – thus, such smaller pieces have to be taken into primary consideration in order to acquire the morphosyntactic rules of a language.

Like English, most of the world's languages express morphosyntactic processes as relations between grammatical morphemes, i.e., as pieces of information internal to word structure. However, languages differ in the degree to which they rely on morphology to carry grammatical information. For instance, English is a language with a rather simple morphology, as it makes use of a few grammatical morphemes to express morphosyntactic relations. Other languages have richer morphological systems than English; they express much more syntactic information via grammatical morphemes, which are "fused" together in a single word. For example, in Italian the sentence in (1) would become:

(3) I ragazzi antipatici con i capelli marroni stanno giocando con le palle rosse.
the.PL.MASC unfriendly.PL.MASC boy.PL.MASC with
the.PL.MASC brown.PL.MASC hair.PL.MASC
play.PRES.CONT.3PL.MASC. with the.PL.FEM red.PL.FEM
ball.PL.FEM
"the unfriendly boys with the brown hair are playing with the red balls".

Italian has morphological markers not only to indicate number, tense and mood, but also gender and person. In the example (3), the noun ("ragazzi", "boys") should simultaneously express number and gender in agreement with the determiner ("i", "the"), as well as number together with the auxiliary ("stanno", "are"). Hence, in Italian one form of a morpheme can simultaneously encode several meanings, such as in the previous example, where the suffix "-i", added to the stem "ragazz-", determines both gender and number[1].

Some languages express morphosyntactic relations with an even more extreme degree of fusion between several morphemes within each word, thus yielding words with a very rich and complex structure. For instance, in Mohawk, an Amerindian language, several morphological markers are added to the word stem and "fused" together to generate one single word, expressing alone the idea that would be conveyed in an entire sentence in a language like English (which instead makes a rather limited use of grammatical morphemes), for example:

(4) Washakotya'tawitsherahetkvhta'se

---

[1] Most European languages have a somewhat fusional morphology; for example Spanish, French, German, and Russian.

(He made the thing that one puts on one's body ugly for her)
"He ruined her dress"

In other languages, words also contain several morphemes, but instead of "fusing" several morphemes together – and making word boundaries difficult to identify --, the grammatical rules combine (or "agglutinate") lots of easily separable morphemes of different types (nouns, verbs, affixes, etc.). That is, smaller morphemes, each generally having one meaning or function and retaining its original form and meaning during the combination process, are glued together, such as in the next example (in Finnish)[2]:

(5) juoksentelisinkohan
"run-erratic motion-conditional-I-question-casual"
"I wonder if I should run around (aimlessly)"

The examples in (3) – (5) illustrate the typologically variability of the world's languages. They vary along a continuum in the extent to which they rely on morphology to mark grammatical distinctions and relations. In essence, they range from languages that use only single morphemes and almost do not have an internal compositional structure in terms of word morphemes (the extreme case is Chinese, which uses only some bound morphemes), to languages that form words by affixing a given number of morphemes (the extreme case is represented by Amerindian languages, such as Mohawk, see example 4). In spite of this variability, we may want to conclude that natural languages express (with various degrees of complexity) morpho-syntactic processes as relations between word sub-parts. Such relations often appear as dependencies between distant sub-parts, both between words (as in "*are* play*ing*") and within words (as in "*un*friend*ly*").

Then, to acquire morphosyntactic rules, infants should be able not only to

---

[2] Other languages having an agglutinative morphology are Turkish, Hungarian, or Korean.

identify words from fluent speech, but also to analyze them into smaller pieces, and to discover the rules determining the legal combinations of morphemes into words. Such rules often require to track dependencies between nonadjacent word sub-parts. In addition to discovering relations between (nonadjacent) word sub-parts, language productivity demands to generalize the structural regularities lying below the internal composition of words. Importantly, infants should derive both words and the rules governing their structure just by listening to the speech stream, without relying on consistent or overt cues indicating the relevant information to consider, and without receiving any explicit instruction.

What computational abilities assist infants in attaining such linguistic knowledge? An increasing interest for the early linguistic sensitivity yielded a growing body of evidence attesting early computational abilities. While several studies established what mechanisms infants use to segment the speech input into words, very little is known about of the nature and the limits of the learning mechanisms infants recruit to acquire linguistic structure. Moreover, despite the relevance of structural information carried by words in their compositional structure, very little attention has been devoted to understanding whether infants can generalize structural information within words, and whether they use mechanisms of the same or different nature to accomplish the two distinct (but related) tasks of identifying word and of discovering their internal structure.

The aim of the present thesis is precisely to investigate the early acquisition mechanisms accounting for the discovery of words and of their compositional structure. I will explore this issue from two complementary directions: investigating the innate dispositions of infants, and assessing the distributional properties of the linguistic input directed to children. To investigate the nature of the computational mechanisms infants use to segment the speech stream into words and to discover the structural regularities underlying words, I conducted a series of artificial grammar studies with infants at different ages (7, 12 and 18 months). Early linguistic abilities are of particular relevance as they provide a window into the computational skills that emerge during development to learn language in all its complexity. To understand how the computational selectivities and the limits of the computational mechanisms match up with the limitations and the properties of natural language, I evaluated the distributional properties of speech directed to children. These analyses aimed at assessing whether the input children listen to may offer a reliable basis for the acquisition of morphosyntactic rules. I examined Italian, a language with a rich and complex morphology, evaluating whether the word forms used in speech directed to children would provide sufficient evidence for the morphosyntactic rules of this language.

The thesis is organized as follows. In Chapter 3, I will review and discuss the existing literature about the computational mechanisms infants recruit to segment the speech flow into units, to acquire grammar-like structures, and to generalize simple patterns to novel items. While researchers agree on the statistical nature of the mechanisms involved in word segmentation, theoretical controversies exist about the nature and the limits of the mechanisms accounting for the acquisition of syntax. In Chapter 4, I will discuss the matter of debate between the two most relevant approaches (i.e., the "statistical learning" and the "rule-based" approach), and I will present in great details a third approach, i.e. the "MOM hypothesis". This hypothesis postulates that mechanisms of different nature are required to extract words and to acquire morphosyntactic rules. In Chapters 5 and 6 I will describe two series of artificial grammar studies carried out with infants at 7, 12 and 18 months, aiming at investigating the MOM hypothesis, and predicting that distinct computational components -- having different sensitivities and developmental trajectories -- are required to accomplish word segmentation and structural generalizations. In Chapter 7, I will examine the distributional properties of the linguistic input by analyzing Italian child-directed speech corpora. These analyses are aimed at assessing whether the input young children listen to may offer them a reliable basis for the acquisition of Italian morphology. Finally, in Chapter 8 I will summarize the main findings of the empirical work of the thesis, and I will discuss their broader implications for language acquisition theories.

CHAPTER 3

LEARNING WORDS AND RULES FROM SPEECH: THEORIES AND EMPIRICAL FINDINGS

Chapter 2 pointed out *what* infants have to learn by listening to the linguistic input, i.e., words and the rules governing their compositional structure. The present chapter is aimed at discussing *how* humans face the problem of identifying words from the speech flow, and the one of discovering grammatical structure. Several studies interrogated about the challenges young learners are faced with to break the speech stream into words, and to discover the structure underlying them. Along with these investigations, researches speculated about the existence of mechanisms assisting infants both in segmenting words from the speech flow, and in acquiring linguistic structure. This chapter is devoted both to presenting the data currently available about how infants segment the speech into words and acquire linguistic structure, and to discussing the different theoretical approaches accounting for the nature of the language acquisition mechanisms.

The chapter is organized as follows. The first part will describe what challenges pose the problem of identifying words from fluent speech, and what is already known about how young infants may solve the segmentation problem. Special attention will be given to the "statistical learning approach", as it represents an important perspective taken up by the experimental work of the thesis.

While much progress has been made to understand how infants and adults achieve the segmentation of the speech stream into words, our knowledge about how infants come to discover linguistic structure is still incomplete. The second part of the chapter is devoted to presenting the most relevant studies investigating the early acquisition of linguistic structure. A critical discussion about the mechanisms that may explain how infants generalize grammar-like structures will follow the literature review.

Theoretical controversies exist about the nature and the limits of the mechanisms assisting infants in acquiring language in all its complexity. At the heart of the debate stands the hypothesis that the same mechanisms involved in word segmentation can also account for the acquisition of linguistic structure. The last part of the chapter is devoted to introducing the matter of controversy between the most relevant approaches on language acquisition, as it is central to the theoretical framework of the thesis; the debate will be taken up and widely discussed in Chapter 4.

3.1 The challenge of discovering words from fluent speech

Identifying word boundaries in continuous speech is not an easy task: words are not consistently delimited by silences (Cole, & Jakimik, 1980), as shown by the sound spectrogram in Figure 1.

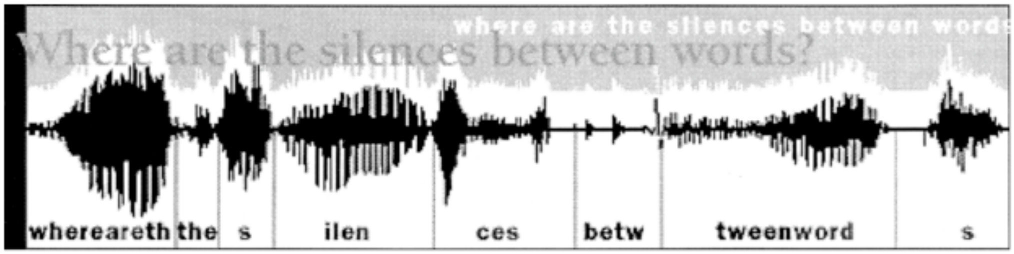


Figure 1: A speech waveform of the sentence "Where are the silences between words?". The height of the bars indicates loudness, and the x-axis is time. This example illustrates the lack of consistent silences between word boundaries in fluent speech. The vertical gray lines represent quiet points in the speech stream, some of which do not correspond to word boundaries. Infants must determine where one word ends and the next word begins, without access to obvious acoustic cues. Image and caption adapted from Saffran (2003).

Given the lack of overt marks delimiting words, the problem of segmenting the speech flow into words is potentially very challenging for infants. However, all normal children solve the segmentation task without experiencing special difficulty. How can they succeed? In order to answer this question, we might consider two facts. First, the speech input contains different forms of information that can be useful markers to word boundaries. Second, young infants are equipped with computational abilities allowing them to exploit many of the available segmentation cues. The input properties and the learner's abilities represent the two factors contributing to language acquisition; the next two sections will discuss each of them.

3.1.1 From the input side: the informative richness of the signal

The existence of cues other that silences marking word boundaries has been extensively investigated. Phonologists and acoustic-phoneticians proposed that different cues could be potentially useful to delimitate words, for example allophonic variation, lexical stress, vowel harmony, and prosodic features such as duration, pitch and energy (see Jusczyk, 1999, for a review). Several studies showed that infants are sensitive to acoustic correlates of word boundaries since birth (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Christophe, Mehler, &

Sebastián-Gallés, 2001), and that, by 9 months of age, they are able to consider phonotactic regularities (Friederici & Wessels, 1993; Mattys & Jusczyk, 2001; Sebastián-Gallés & Bosch, 2002), prosodic patterns (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Houston, & Newsome, 1999; Morgan, 1996), stress patterns and allophonic variation (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Jusczyk, Hohne, & Bauman, 1999) as reliable cues to word boundaries.

However, all the above-mentioned cues pertain to the sound organization of a language, thus they are language specific. For example, in English /l/ and /r/ are distinct phonemes, as they are used in different words (for instance, "led" and "red"), while in other languages (such as in Chinese) they are allophonic variations of the same phoneme. Similarly, most of the English content words are stressed in their first syllable, while French words are generally stressed in their last syllable. As the abovementioned cues are all language specific, to make use of them infants must already know something about the sound patterning of their native language, particularly with respect to correlations between sound patterns and word boundaries. To put it in another way, they have to learn something about the sound pattern of words in their language before applying this knowledge to detect the words occurring in speech. Therefore, the question is how infants might achieve the initial segmentation, identifying a subset of word-like units from which to glean other language-specific regularities.

A possible solution is that infants possess one or more adaptable learning strategies. One ability that infants may bring to bear on word segmentation is the ability to exploit the statistical regularities of language to bootstrap initial word forms, from which other regularities can be discovered (Saffran, Aslin, & Newport, 1996; Aslin, Saffran, & Newport, 1998). The idea behind this proposal is that a great deal of information regarding the basic units and grammatical constructions of a language is provided by the patterns in a corpus from that language: what elements regularly occur together, what elements cannot be interrupted, and the like. This approach is called *"statistical learning"*, a term adopted from a related literature in computational linguistics (Charniak, 1993) and used both by structural linguistics (Harris, 1955) and by nativist perspectives

(Chomsky, 1957; Chomsky, 1980), along with proposals using distributional analysis as part of the process of language acquisition (Morgan, Meier, & Newport, 1987; Rumelhart, & McClelland, 1986; Seidenberg, 1997; Saffran, Aslin, & Newport, 1996; Bates & Elman, 1996). In the next paragraph, I will present one type of such linguistic regularities infant may take into account to solve the initial segmentation problem, namely *transitional probability*.

3.1.1.1 Transitional Probability

The linguistic code is a rich source of information, as we can empirically appreciate if we consider that words merely consist of *predictable sequences of sounds*. Several authors have proposed that over a corpus of speech there are measurable statistical regularities that distinguish recurring sound sequences that comprise words from the more accidental sound sequences that occur across word boundaries (Harris, 1955; Hayes, & Clarke, 1970; Charniak, 1993; Brent & Cartwright, 1996).

The statistical regularity expressing the probability of one syllable given the occurrence of another syllable is termed *transitional probability* (henceforth, TP) or conditional probability between syllable sequences. The TP of a syllable y given another syllable x is expressed by this formula:

$$\mathrm{TP}\,(x \rightarrow y) = \frac{\mathrm{frequency}(xy)}{\mathrm{frequency}(x)}$$

TP is a statistics that refers to the relation more than the frequency with which one element follows another, as it adjusts for the base rate of the first event or element. For example, if we consider the word sequence *pretty#baby*, we can describe its statistical structure as follows. Because the syllable *pre* precedes a small set of syllables in English, the probability that *pre* is followed by *ty* is quite high. However, because the syllable *ty* occurs word-finally, it can be followed by any syllable that can begin an English word. Thus, the probability that *ty* is followed by *ba,* as in *pretty#baby*, is low. This difference in sequential probabilities is a clue that *pretty* is a word, and *tyba* is not. Hence, the TP from one sound to the next will generally be higher when the two sounds follow one another within a word, whereas TPs spanning a word boundary will be relatively low.

As the linguistic input is statistically informative, it provides a basis from which learners can derive its units and structures. The question is whether adults and children can use information about TPs to discover word boundaries in a continuous speech stream, when no other cues are available. The next paragraph reviews the most relevant studies that have been conducted to establish whether learners possess the ability to track TPs, and to capitalize on them to learn the statistical structure of a linguistic input.

3.1.2 From the learner's side: The use of TPs in segmenting a continuous speech stream

In their seminal work, John Hayes and Herbert Clark (1970) showed that adults are able to segment a continuous stream, identifying its recurrent patterns (i.e., words) just by computing TPs. Remarkably, they can do so simply by listening to the language without the aid of other segmentation markers than TP, reinforcements, referents, or without the need of a teacher. Hayes and Clark hypothesized the existence of a "clustering" mechanism aiding the segmentation task. Participants to Hayes and Clark's experiments were exposed to an artificial language, designed to be unfamiliar and not to contain extraneous segmentation cues (such as stress pattern or semantic information). Adopting this methodology, they faced their adult participants with a situation that closely resembles the one the child encounters when s/he first listens to language: s/he perceives a phonological structure, s/he listens for the statistical properties of the speech input, and s/he identifies where the word boundaries are in the speech.

Apart from the analogy with adults, are young infants sensitive to probabilistic cues? Do they use them to accomplish the word segmentation task? The answer to these questions arrived almost thirty years after the studies by Hayes and Clark, and was provided by the experiments conducted in 1996 by

Saffran, Aslin and Newport. The reason for this delay was the widespread belief concerning the "unlearnability" of language (Bates & Elman, 1996). This belief followed as a logical conclusion from the influential work of Noam Chomsky, who argued that, as the language faculty is codified in the human genetic program, learning is a useless scientific construct (Chomsky, 1980). This assumption was based on the argument that infants attain a linguistic knowledge that is "perfect", and they do so in spite of the "imperfect" data available from everyday language use.

Saffran, Aslin and Newport (1996) proved that young infants are capable of computing adjacent TPs from a linguistic input, and they use these simple statistics to learn sequences of syllables in rapidly presented streams of speech. In their study, 8-month-old infants were exposed to a miniature artificial language, consisting of four nonsense sequences of three syllables (e.g., "bidaku", "padoti", "tupiro", etc.) repeated several times for two minutes. A sample of the speech stream is the orthographic string:

…bidakupadotigolabubidakutupiro…

The four sequences were characterized by perfect adjacent TP among adjacent syllable; that is, given a certain syllable of a sequence, it predicted with TP of 1.00 the following syllable. For instance, in the sequence /bidaku/, the syllable /bi/ always predicted the syllable /da/, and /da/ always predicted /ku/. Instead, TPs across the boundaries of two sequences were equal to .33. That is, the last syllable of a sequence predicted with equal probability the occurrence of the first syllable of any of the three other sequences. The four sequences can be statistically defined as "words", because they have high adjacent TPs between syllables of a sequence, and low TPs between syllables spanning the boundaries of two sequences. The artificial stream was synthesized to be continuous and monotonous, in order not to provide any acoustic cues to segmentation. The only signal to word boundaries was given by the statistical structure of the stream, since syllables within a word followed each other with a TP of 1, while syllables spanning word boundaries had a TP of 0.33.

After having exposed the 8-month-olds to the two-minute stream, Saffran and colleagues presented infants with sound sequences that could be either words (like "bidaku" and "padoti"), or sequences that occurred in the stream but spanned the word boundaries, for example "kupado", which contains the last syllable of the word "bidaku" and the first two syllables of "padoti". They called such sequences *part-words.* Part-words had the "wrong" word structure in statistical terms, since they contained a drop in word-internal TPs. Saffran and collaborators compared the infants' performance on words against part-words, and the results showed that the short exposure to the artificial language was enough for 8-month-olds to discriminate between them. To succeed at this task, listeners would have had to track the statistical properties of the input, and to group syllables into word-like units (in order to distinguish them from sequences spanning word boundaries): this is what Saffran and collaborators concluded from their results.

In the stream used by Saffran and collaborators, each of the four words appeared equally often in the stream, thus trisyllabic sequences that formed the words had not only higher TPs than part-word, but they also occurred more frequently than part-words. Hence, the results showing that infants could discriminate words from part-words are compatible with two interpretations: infants could have used TPs among adjacent syllable, or they could have used the frequency of co-occurrence of the syllables in that sequences to succeed in the task. The design of the study by Saffran et al. (1996) cannot discern which of the two types of information infants actually performed. In a follow-up study, Aslin et al. (1998) documented that infants used precisely TPs –and not other statistics- to solve the segmentation task (Aslin, Saffran, & Newport, 1998). To control for the possible influence of frequency in determining the results, they created a stream in which two of the words occurred twice as often as the other two. As a result, the syllable sequences across boundaries of frequent words occurred with the same frequency as the two less frequent words. By equating the frequency of less frequent words and part-words, but not their TPs (all words had adjacent TPs of 1.00, frequent part-words of .50), and asking infants to discriminate between them in the test phase, they could clarify that only the difference of TPs accounted for the segmentation of the their continuous streams.

3.2 Beyond speech segmentation: the acquisition of linguistic structure

Sensitivity to adjacent TPs could help infants to learn recurrent linguistic patterns and to identify the words of a language, as suggested by Saffran et al.'s results. However, mastering a language does not entail only to acquire its lexicon. Instead, it also requires representing structural relations between sub-lexical morphemes or syntactic word classes. Such structural relations determine the combinations of language units into larger constituents; thus, they govern the astronomical variety of words and sentences any natural language user can produce and understand. In spite of the importance of learning the rules combining elements into structures, how infants come to derive them is still poorly understood.

The next sections will review the data currently available examining aspects of the language learner's task that are different from (and more complex than) word segmentation. The first aspect concerns the ability to track dependencies between distant elements (Section 3.2.1). The second aspect involves the ability to learn abstract categories (e.g., determiner, noun, verb, etc.; Section 3.2.2). The third aspect concerns the ability to abstract and generalize grammatical relations (Section 3.2.3). All these three aspects are related to language learner's ability to organize linguistic elements into grammatical structures.

3.2.1 Learning dependencies between nonadjacent elements

All natural languages have a system of rules for constructing novel words and sentences (Chomsky, 1957; 1980) Remarkably, in most of the world's languages, the structural information expressing the morphosyntactic rules involves the relation between *nonadjacent* elements (see Chapter 2). For instance, in the sentence *"the kids with the red hair are using the sand in the garden to play"*, the final verb ("to play") does not refer to its adjacent constituent ("in the garden") but rather to the initial one ("the kids"). Similarly, dependencies between the auxiliary and inflectional morphemes are expressed through distant elements, as in the smaller portion of this sentence "*are using*" (underline font marks the morphemes to consider). This is also the case of dependencies involving number agreement, for example *"the kids with the red hair are* […]". In morphology, comparable phenomena occur. For instance, words in some languages[3] are created by adding suffixes to the word morpheme, e.g., *"uninteresting", "ex-girlfriends*" (for additional examples, see Chapter 2). Some verbs in Italian are created in the same fashion by adding simultaneously a prefix and a suffix to the adjectival morphological root, e.g., the verb "*imbiancare*" ("to paint"), where neither the adjective *"imbianco"* nor the verb *"biancare"* exist. In Semitic languages, words may be built from a consonant pattern, with varying vowel patterns inserted between the consonants to mark time or number. For instance, in Arabic morphemes are defined by three consonants, to which various affixes (prefixes, suffixes and infixes) can be attached to create a word. For example, the tri-consonant "*k-t-b"* represents the concept of writing. Some of the ways in which "*k-t-b*" is turned into real Arabic words are *kataba* ("to write"), *kaatib* ("writing"), *kitaab* ("a book"), or *kattaab* ("author.")

As such, learning dependencies between nonadjacent words, morphemes or segments is fundamental in acquiring the morphology and the syntax of the language. The question is whether infants are able to detect a structure requiring to track distant elements.

---

[3] As stated in Chapter 2, languages across the world vary in their morphological typology. Some of them mark grammatical distinctions and relations of complex morphology through morphological markers; others have a poor morphological system, and their words almost do not have an internal compositional structure in terms of word morphemes. Languages of the first typology face learners with the problem of analyzing words into their internal structure.

3.2.1.1 Can infants learn nonadjacent dependencies?

In the last decade, researchers have begun investigating how infants might learn nonadjacent sequential dependencies, as they have an important role in most of the world's languages - specifically, in expressing morpho-syntactic relations.

Santelmann and Jusczyk (1998) tested infants with their natural language (i.e., English) and showed that, by 18 months, they are able to track nonadjacent dependencies over as many as three intervening morphemes (thus, suggesting some processing limitations correlated to the distance between the dependent morphemes). Toddlers at this age were able to distinguish phrases like "*is running*" from "*can running*" and "*is quickly running*" from "*can quickly running*". In contrast, younger children (at 15 months) did not appear to track the co-occurrence of the morphosyntactic sub-elements, suggesting that the knowledge about this type of dependencies between morphemes develops sometime between 15 and 18 months.

Gómez (2002) replicated the Santelmann and Jusczyk's (1998) findings with 18-month-olds using an artificial language paradigm. She exposed 18-month-old infants to artificial languages containing strings of three nonsense words separated by silences of 250 ms, like "pel-wadim-jic", or "vot-kicey-rud". The first and the last elements of the sentences were always the same (that is, their pair-wise transitions were held constant), while the middle element varied. This variation was aimed at investigating the role of variability of the intervening element in the detection of predictable (or invariant) structures of nonadjacent elements. Gómez was interested in studying the conditions and the constraints in the perception of remote dependencies because she observed that, in natural languages, the intervening categories are often open-class items comprising much larger sets than the function morphemes associated with nonadjacent structure (for instance, all the verbs can potentially appear between the auxiliary "*is*" and the inflection "*-ing*"). This pattern would translate into lower TPs between adjacent syllables than between distant syllables, and would probably configure the elements from the smaller sets as prominent or less variable. Gómez hypothesized that the set size of the middle elements might aid learning of the "invariant" surrounding structure. Possibly, this learning would occur via the same statistical computation mechanisms involved in the detection of adjacent TPs. To test this hypothesis, she manipulated the size of the set from which she drew the middle element (set sizes could be of 2, 12, or 24 elements), while holding frequency of exposure to particular nonadjacent dependencies constant. She found that 18-month-olds were able to acquire the nonadjacent dependency between the first and the last elements, but only when the intervening element came from the largest set of possible items (i.e, set size 24).

In a following study, Gómez and Maye (2005) extended these results to languages with different set sizes (12, 18 and 24 middle elements) and tested infants of 12, 15 and 17 months. Their results replicated the main findings of the original study with a slightly younger age group (i.e., 17-month-olds) and a more subtle variability manipulation: 17-month old infants failed to learn nonadjacent dependencies in the 12 set size, but succeed in the 18 set size condition. In contrast, 12-month-olds did not appear able to track the nonadjacent dependencies neither in the 18 nor in the 24 set size conditions. However, by 15 months of age infants begin to show sensitivity to this structure: they attend to nonadjacent dependencies only under the condition of maximal variability (set size 24). Overall, the results of Gómez and Maye suggest that the ability to track dependencies between remote elements gradually develops from 15 to 18 months (thus, confirming the period of acquisition proposed by Santelmann and Jusczyk, 1998).

Taken together, the results of Gómez (2002) and Gómez and Maye (2005) show that young learners can acquire remote dependencies from segmented strings of three elements, but only under the condition of high variability of the intervening element. A pattern with high variability in the occurring element translates into low adjacent TPs, high nonadjacent TPs, and configures the remote elements as much more frequent than the middle ones. Hence, it is possible that the detection of the first and the last items may have been favored by their distributional properties (both nonadjacent TPs and frequency of occurrence).

Overall, these studies suggest that nonadjacent dependencies are acquired later – and, possibly, with more difficulty - than adjacent ones[4]. Further research is needed to understand the limits, the constraints and the neural substrate of the computation of nonadjacent dependencies[5]. Still, to acquire linguistic structure, not only do language users have to identify and track remote elements, but also to derive an abstract knowledge about the rules governing the combination of distant elements. The next section will focus on the latter issue, that is, the learners' generalization abilities, and it is aimed at reviewing some studies investigating to what extent young infants are capable of extracting abstract categories and rules from the linguistic input.

3.2.2 The abstraction of linguistic categories

The ability to abstract over categories is fundamental to linguistic productivity. For example, English-speaking children need to learn that the determiners "the" and "a" precede nouns and not verbs, whereas auxiliaries like "was" and "is" precede verbs, but not nouns. Despite the relevance of category abstraction in language acquisition, how young learners master this kind of knowledge is currently poorly understood.

Among the few studies investigating whether young learners can "bootstrap" some aspects of syntax considering the distributional properties of the input, Gómez and Lakusta (2004) addressed the issue of abstract category-based generalizations, in which each variable in a rule can be filled with all members of the relevant category. For example, in the (simplified) template of a transitive sentence noun–verb–noun, all nouns can fill the first and the last

---

[4] A recent neuroimaging study conducted with newborns suggests that distant relations are processed differently compared to close relations (Gervain et al., 2008a).

[5] For a discussion about how prior experience influences sensitivity to nonadjacent statistical regularities, see Lany & Gómez, 2008.

position. The two nouns in such a sentence need not be (and in general are not) identical; instead, what is repeated in the "noun–verb–noun" template are the *categories* to which the items in the sentence belong. In their studies, Gómez and Lakusta exposed 12-month-old infants to auditory structures of the forms *aX* and *bY*, where *a* and *b* were elements of the form vowel, consonant, consonant (e.g., "alt", "ush"), X were disyllabic elements and Y were monosyllabic elements. In the test phase, infants could generalize abstract rules (for example, "*a* is followed by a disyllabic", "*b* is followed by a monosyllabic") to pairs containing novel *Xs* and *Ys*. At the basis of such generalization stands the ability to notice and to abstract the functional elements differentiating the lexical categories, and the ability to extent such properties to novel word-like elements. Importantly, in the study by Gómez and Lakusta distributional cues were correlated with phonological cues; hence, whether multiple, convergent cues are required for learners to abstract categories is still an open question.

3.2.3 Acquiring rules and generalizing them to novel structures

Although sensitivity to distributional information is necessary for tracking sequential information in speech, language learners must ultimately abstract beyond the specific linguistic input they listened to, and generalize to novel structures. As such, the ability to abstract rules and to generalize them is fundamental to linguistic productivity.

3.2.3.1 Generalization of simple rules from patterns containing repetition

Do infants possess a mechanism for learning rules, that is, open-ended abstract relations for which they can substitute arbitrary items? Marcus et al. (1999) asked this question and investigated whether young infants are able to generalize their knowledge about the rules governing the sound combinations. In their studies, infants were first presented with syllables (such as ''ga'' or ''ti''), organized into triads that follow either the sequence ABA (e.g., ''ga ti ga'') or ABB (e.g., ''ga ti ti''). Infants were then tested using new syllables (e.g., "wo", “fe”), which were arrayed in either the familiarized or novel patterns (‘‘wo fe wo’’ versus ‘‘wo fe fe’’, in abstract terms, ABA versus ABB). Seven-month-old infants could readily acquire these abstract rules and could generalize them to novel triads, as evidenced by their successful discrimination between the familiar and novel patterns when presented using new syllabic patterns. Building on their results, Marcus and colleagues concluded that infants are capable of extracting algebraic rules that represent relationships between placeholders (variables), such as “the first item X is the same as the second (or third) item J”.

To what extent is the ability to generalize abstract patterns specific for language? Saffran et al. (2007) tackled this question, and tested infants’ generalization in the visual domain. The results showed that infants were able to detect simple rules including a repetition from different elements of a visual category (e.g., dogs, cats). In contrast, Marcus et al. (2007) showed that infants fail to learn simple rules when they are implemented using non-linguistic stimuli, including the same kind of tones and visual shapes about which infants successfully learn in statistical learning tasks. However, infants are able to subsequently generalize, or to transfer, that rules to a different domain (tones, timbres, animal sounds) if they if they first hear those rules instantiated in sequences of speech. This latter result suggests that speech can facilitate rule-learning in domains where infants may otherwise not acquire rules (Marcus, Fernandes, & Johnson, 2007).

How to reconcile the findings from the studies of Saffran et al. (2007) and of Marcus et al. (2007)? It is possible the task performance may have been facilitated by materials that infants can readily represent and/or categorize. On this view, the animal pictures used by Saffran et al. and the syllables used by Marcus et al. could have been readily categorized as token of the same category, thus they could have been represented as elements of the triads. From this representation, infants may have abstracted the general pattern of same/different elements. In contrast, infants could not categorize tones, timbres and animal sounds. The reasons why speech and dog pictures are easy to represent while acoustic or animal sounds are not are still unknown. One possibility is that both speech and animal pictures are highly familiar or salient to infants. Concerning speech, recent data provided evidence that human beings are profoundly interested from birth in paying attention to it (Peña et al., 2003; Vouloumanos & Werker, 2004, 2007). This preference would help newborns to orient selectively to spoken language and would account for the saliency and familiarity of speech for older infants.

3.2.3.2 A specific mechanism for detecting repetition

Overall, the above mentioned studies show that infants are capable of detecting regularities from sequences containing elements that are identical or different, and they generalize such regular patterns to novel instances. The detection of structures involving repeating elements has been proposed to be a "perceptual primitive", i.e., a specific configuration automatically processed and detected as a result of the way perceptual systems function (Endress, Dehaene-Lambertz, & Mehler, 2007; Gervain et al., 2008a). A perceptual primitive is assumed to be a highly specialized mechanism, operating over specific patterns; it is recruited rather effortlessly, it does not depend on learning and, as such, it is not influenced by the distributional properties of the linguistic input (for instance, frequency of occurrence and co-occurrence of elements in the language, see Tunney & Altmann, 2001). According to this view, elements that are repeated have a special status.

In agreement with this hypothesis, several studies with adult participants attested the salience of repetition structure during transfer (Gómez, Gerken, & Schvaneveldt, 2000; Endress, Scholl, & Mehler, 2005). The findings of these experiments suggest that the work by Marcus et al., alleging rule-learning in young infants, could be interpreted as a highly constrained process driven by perceptual primitives operating over repetition-based sequences, rather than a formal, algebraic process operating over all patterns equally well. There is another reason to suspect that the detection of repetition-based pattern may depend on lower level mechanism: cotton-top tamarins can learn grammars including repetitions (Hauser, Weiss, & Marcus, 2002), and even bees can learn identity and non-identity relations (Giurfa, Zhang, Jenett, Menzel, & Srinivasan, 2001). Since non-humans are endowed with such perceptual primitive, it is unlikely that such operator is among the specific computational capacities that make language possible only in humans.

3.2.3.3 Infants can generalize words in abstract patterns

The ability to perceive and project structural generalizations appears to be required in natural language processing, and it is not fully captured by studies in which variables must be filled with particular physical tokens (such as the syllables in Marcus et al.'s, 1999, experiments). Several studies attempted to study whether infants could generalize a fairly complex artificial grammar to a new vocabulary. For instance, Gómez and Gerken (1999) trained infants with an artificial language consisting of strings of three up to six syllables. The strings could contain different syllables (e.g., "vot-pel-jic-rud-tam") or a repeating syllable, either adjacent (e.g. "pel-tam-rud-rud", in abstract terms "_ _ A A") or nonadjacent (e.g., "pel-tam-pel-jic", in abstract terms "A _ A _). The results showed that infants could discriminate new strings produced by the training grammar from strings produced by another grammar[6]. This generalization was probably based on the learning of the specific ordering of internal pair-wised combinations (at the first-order level), as well as on specific patterns of repetitions[7].

Saffran and Wilson (2003) extended this line of research, and asked how infants might approach learning tasks consisting of multiple levels of information. Twelve-month-olds listened to a continuous speech stream in which the words were ordered via a finite-state grammar. The infants were thus

---

[6] See Experiment 3, Gómez & Gerken (1999).

[7] It should be noted that Gómez and Gerken (1999) did not explain their results in terms of "algebraic" rule-based processing. Instead, they proposed that the transfer could be due to "complex associations" or to the detection of the repetition pattern, which may or may not be related to abstractions in grammar; later experiments showed that the latter interpretation is probably correct (Gómez et al., 2000; Tunney & Altmann, 2001).

presented concurrently with a word segmentation task and a syntax-learning task, which required them to abstract the permissible word ordering and to generalize it to novel sentence-like strings. The results suggested that infants could perform two sequential tasks: first, they segment words contained in a continuous stream, subsequently they discover syntactic regularities relating the words and they apply this knowledge to discern grammatical novel sentences (that is, sentences respecting to the permissible order of words) from ungrammatical ones. Several other studies conducted with adults and children suggested that statistical learning may play a role in the acquisition of syntactic rules (Saffran, 2001; Thompson, & Newport, 2007).

3.2.4 Does the speech directed to young children provide a reliable basis for the acquisition of linguistic structure?

The studies presented so far concerned the mechanisms human learners use to acquire linguistic structures. Infants seem to possess remarkable computational abilities to track nonadjacent dependencies, to generalize linguistic patterns, and to derive linguistic categories. However, one indisputable fact about language acquisition (thus, including the acquisition of linguistic structure) is that language has to be learnt. Several studies considered the problem of syntax acquisition from a complementary perspective, asking not only what abilities assist infants in acquiring linguistic rules and structures, but also to what extent the linguistic input directed to children may contain sufficient distributional information to support such abstract learning.

The studies examining the distributional properties of language directed to children with the purpose of assessing the potential contribution of the linguistic input in the acquisition of grammatical categories (e.g., Noun, Determiner, Verb, etc.) arrived at opposite conclusions. Some of them found that the input is informative and, in virtue of its statistical properties, it may offer learners a reliable basis to form a representation of phrase structure. For example, Gervain et al. (2008b) examined corpora of child directed speech and found that the distribution of function words (like *the, of, his*, etc.) and of content words (like *mammy, mirror, drink*) correlates with the order in which words appear in a given language (i.e., whether the ordering sequence is "determiner-noun", like in *"the mirror"*, or "noun-determiner", like in Japanese *"Taroo ga tegami o kaita"*, "Taroo a letter wrote"). Thus, frequency information is a useful predictor of the category membership (Noun or Determiner) and thus may allow the differentiation between the two categories. Similarly, Mintz et al. (2002) found that the input directed to young children contains distributional information, such as pairs of words frequently occurring with one word position intervening (such as "you X to", where X stands for a verb). Thus, the input offers a reliably basis to the process of constructing the grammatical categories of (at least) nouns and verbs. In contrast, other studies showed that the input does not unambiguously support the representation of linguistic structures (Lidz, Waxman, & Freedman, 2003; Yang, 2004). The authors of these latter studies interpreted their findings as evidence that the distributional analysis might be useful to acquire some -- but not all -- aspects of language. Thus, any such analysis must serve the innate predispositions of the learners, rather than determining alone the acquisition of linguistic structures.

Overall, we may want to conclude that understanding how linguistic structures are acquired requires balancing the structure derivable from the surface input with the structure inherent in the learner.

3.3 Summary and theoretical controversies

This chapter looked at the studies investigating what kind of abilities young learners possess for segmenting words and for acquiring linguistic structure. It has been shown that 8-month-old infants can use their sensitivity to adjacent statistical dependencies to discover word boundaries. At 18 months, children can track and learn more difficult statistical dependencies, involving relationships between nonadjacent elements. Moreover, at 7 months infants are capable of generalizing abstract rules underlying sequences of repeating elements (both adjacent and nonadjacent). At 12-months, they can abstract linguistic categories, and generalize sequences of words after having analyzed he distributional properties of a linguistic input.

Overall, we can believe that young learners possess remarkable computational abilities supporting them in the difficult tasks of identifying words from fluent speech, and of learning and generalizing morphosyntactic structures and linguistic categories. Researches interrogated themselves about the *nature* of such abilities, asking whether the same statistical mechanism involved in learning distributional information (and nonadjacent dependencies in particular) can also account for the acquisition of syntax. They disagree about the role that statistical learning should play in language acquisition in all its complexity. Chapter 4 deals with this theoretical controversy.

CHAPTER 4

MECHANISMS OF LANGUAGE ACQUISITION: STATISTICAL COMPUTATIONS, GENERALIZATION, OR BOTH?

Language exhibits statistical structure (see Section 3.1.1 and 3.2.4 on Chapter 3); however, the importance of this type of structure in determining language productivity is controversial. According to one approach, a mechanism dedicated to computing statistical information may also account for the acquisition of morphosyntactic structure. In contrast, another approach maintains that a rule-based mechanism needs to be postulated to explain the acquisition of grammar.

Whether or not a single mechanism suffices for the acquisition of rule-based linguistic structures is a matter of current debate. The first part of this chapter is aimed at presenting the matter of controversy and the theoretical approaches that have been proposed to account for the learning of linguistic structure. The second part is dedicated at introducing the "MOM hypothesis", a recent proposal that reconciles the tenets of rule-based approach while accommodating evidence for statistical learning. The MOM hypothesis and the experimental work testing its prediction will be presented in great details as they provide both the theoretical premises and the empirical basis for the studies reported in Chapters 5 and 6 of the present thesis.

4.1 Statistics, Grammar, or both?

4.1.1 The "rule-based" and the "associative learning" approach

Consider the following sentence: "Colorless green ideas sleep furiously". We would all agree that language users could immediately recognize that the sentence is well formed. Now consider this sentence: "Ideas colorless sleep furiously green"[8]. We would all predict that language users would not consider it as well formed. This is not a surprising conclusion. However, in a sense, there is something surprising in it. We should notice that both sentences are entirely meaningless and hence extremely unlikely: the probability that the words composing them have previously occurred in this order is close to zero. Therefore, we might conclude that the linguistic knowledge must go beyond the statistical properties of language and must include *rules* allowing learners to generate and/or understand an unbounded number of linguistic structures. This syllogism is the famous argument used by Chomsky to argue that frequency of occurrence and conditional probability cannot be the basis for grammatical/syntactical language knowledge. Instead, he took the fact that meaningless sentences are nevertheless instantly and effortlessly recognized as grammatical as evidence that linguistic knowledge goes beyond what learners heard from the linguistic environment, and allow them to generate and

---

[8] These two sentences are classical examples taken from Chomsky, 1957.

understand brand-new linguistic structures. As a consequence, linguistic knowledge must involve rules. This insight became part of the foundation of modern linguistic theory, and research focused on how the child converges on the rules and other components of grammar using a combination of deductive (non-statistical) reasoning and their innate linguistic dispositions (Chomsky, 1957; Pinker, 1994).

In contrast to this view, known as the "rule-based approach", the "associative learning approach" claims that language exhibits structure at multiple levels, each of which has its own statistical character. The "colorless" grammatical sentence is less puzzling when one looks beyond transition probabilities to other information that is used in comprehension. For example, words fall into general types (e.g., "green" is a property or adjective, and "sleep" is an action or verb) that exhibit characteristic distributions. The "colorless" sentence conforms to these distributions in English, whereas "Ideas colorless sleep furiously green" does not (Allen, & Seidenberg, 1999). This view is taken up by the statistical learning approach and by connectionism, which both considered the learnability problem from the point of view of the statistical structure of natural languages with respect to the human ability to learn associations among items.

The interest in the statistical structure of language and in its contribution to learning inspired a large body of work. Several studies showed that infants and young children incorporate statistical cues when performing several linguistic tasks, such as when learning about the sounds of a language (Maye, Werker, & Gerken, 2002), acquiring a vocabulary (Saffran, Aslin, & Newport, 1996; Graf Estes, Evans, Alibali, & Saffran, 2007), and extracting the structures in which words occur (Tomasello, 2000; Mintz, 2003). These findings complement evidence from adults demonstrating the use of statistical information in comprehending and producing utterances, suggesting that similar mechanisms may underlie the learning and use of language in all of its aspects, that is, including grammar (Seidenberg, MacDonald, & Saffran, 2002; Seidenberg, 1997). If this were the case, then a single, general-purpose statistical mechanism, applied to different levels of the linguistic corpus, would explain how language users acquire both words and morphosyntactic rules (Elman, 1999; Elman, 2001; Altmann, 2002). This approach assumes that the same statistical learning mechanism involved in the identification of words has potentially a role in the acquisition of morphology and syntax. This assumption stands at the heart of the associative theories of learning, and generates opposite predictions than the rule-based approach (according to which, instead, the acquisition of grammar requires mechanisms of non-statistical, deductive nature).

4.1.2 A third, conciliatory approach: The "MOM hypothesis"

Although the research inspired by the distributional learning approach establishes that statistical information is used in language acquisition, the extent to which acquisition of abstract structure (not obviously mirrored in the surface statistics of the input) can be explained in these terms is not yet known. Then, how can we account for the discovery of both statistical regularities and abstract structures?

One possibility could be that both statistical processes (based on frequency and distribution of elements in language) and non-statistical (deductive) processes are involved in language processing. Statistical learning may be limited to solve problems such as learning the sounds of a language and building a lexicon. In contrast, grammar may require other non-statistical procedures. This hypothesis, postulating that at least two distinct mechanisms are involved in language acquisition, is known as the "More-than-One-Mechanism Hypothesis", henceforth "MOM hypothesis" (Peña, Bonatti, Nespor, & Mehler, 2002; Bonatti et al., 2006).

Peña and collaborators conducted a set of experiments aimed at investigating the MOM hypothesis. The next paragraph will present these studies in great details, as the main findings of Peña et al. inspired the rationale of the infant studies reported in this thesis (see Chapters 5 and 6), and the material I used to test infants was adapted and simplified from the set of stimuli designed by Peña et al.

4.1.2.1 The MOM hypothesis: from theory to data

Before presenting the theoretical stances of Peña et al., it is necessary to present the material they created to carry on their studies. They designed an artificial language containing trisyllabic nonsense "words" with an AXC form (capital letters stand for consonant-vowel syllables). In these nonsense words, the first syllable always predicted the third syllable, thus instantiating the rule "if A, then C". The three A-C frames they used were: /pu - ki/, /be - ga/, /ta - du/. The middle syllable, instead, varied. The intervening syllable was chosen from a pool of three possible syllables, i.e., /li/, /fo/ and /Ra/. Adjacent TP within each word was .33 and nonadjacent TP was 1.00. TP spanning word boundaries was .33 (see Figure 2).

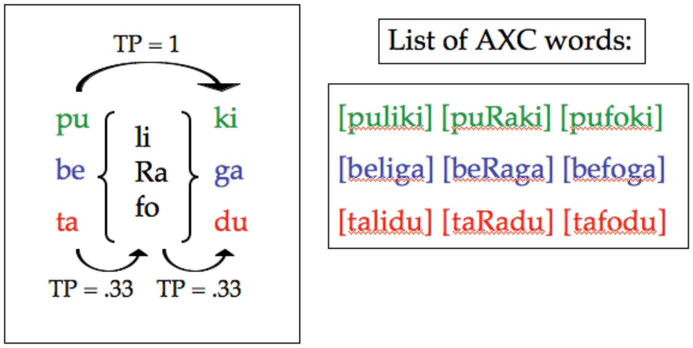


Figure 2: The figure represents the statistical structure of the nonsense words used by Peña et al. (2002). Colors are used to identify the three A-C frames, and to group the three words generated by crossing each frame with the three middle syllables. On the right side is reported the list of nine resulting words.

This material, implementing a remote dependency between the first and the last syllables of words, mimics morphosyntactic constructions present in real languages (for a discussion of how morphosyntactic rules are realized through nonadjacent dependencies, see Chapter 2 and Section 3.2.1 on Chapter 3). Importantly, the first and the last syllables of words were different, thus they did not include reduplication of a syllable as a facilitating cue to abstract structure learning, as it was suggested to account for the learning of the ABA sequences used by Marcus et al. (see Section 3.2.3.2 on Chapter 3 for a discussion of the possibility that Marcus et al. were probably tapping into a different learning mechanism, namely sensitivity to structures containing repetitions).

Here is sample (orthographically transcribed) of the stream concatenating the nine words of the artificial language used by Peña et al. (henceforth, "AXC language"):

...pulikibeRagatafodubeligapuRakitalidupufokitaRadubefogatalidupuRakitafodu...

Peña and collaborators first asked whether participants could track nonadjacent relations from the AXC-language, having adjacent TPs uniformly equalized both within and between words, but high TPs only among the distant syllables of words. They exposed adults to ten minutes of the continuous AXC stream and they tested whether participants preferred the "words" of the language to trisyllabic items that appeared in the stream but spanned word boundaries, containing a part of a word and a part of another word (henceforth, *part-words*). The part-words were either formed by the last syllable of one word and the first two syllables of the following word, i.e., they had an CA'X structure ([kitaRa], [kitafo], [gapufo], and [dubeRa]), or by the last two syllables of one word and the first syllable of the following word, i.e., they had an XCA' structure ([likita], [lidube], [Radube], [Ragapu], [fogapu]).

Results are presented in Figure 3; participants, asked to judge which item seemed to them more likely a word of the imaginary language, significantly preferred words to part-words. This preference indicates that participants could take advantage of nonadjacent statistical dependencies between syllables to segment a continuous stream and identify words.

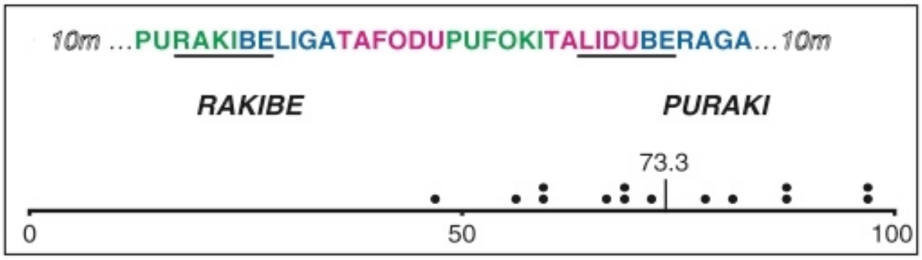


Figure 3: The figure represents the results of the first experiment by Peña et al., testing word segmentation after exposure to a continuous stream. In this and in the next figures, the dots over the line at the bottom of each frame represent individual scores; the number above the vertical mark indicates the general mean. In this experiment, each dot represents the percentage of choices for words (against part-words) of individual subjects averaged across items.

The ability to compute nonadjacent dependencies could, in principle, aid learners also to acquire structural regularities. The AXC language also contain structural information, as its words respect the rule "if A, then C, with an intervening X". In a second experiment, Peña and collaborators asked whether participants exposed to the continuous AXC stream could only identify words, or they could also learn structural regularities. They familiarized participants with the same ten-minutes continuous stream as in their first experiment. However, in the test phase, they tested whether participants preferred part-words against novel items that never appeared in the stream but respected the dependency between the first and the last syllable of words. They called such items "*rule-words*". Rule-words had the same A and C as words, but the intervening X syllable was a syllable that appeared in the stream but never in the middle position, i.e., it could be an A or a C syllable of the other words. An item like /pubeki/ complies with the generalization "If there is /pu/ in the first position, then there will be /ki/ in the last position"; but, because /be/ never occurred in that position during familiarization, the frequency of this trisyllabic token is zero. Thus, rule-words were similar to words in their structure, but differed in their surface form. In contrast, part-words appeared in the stream, thus they had a familiar surface form, but they violated the word structure.

The results are presented in Figure 4. Participants failed to choose rule-words over part-words. Thus, learners exposed to the continuous AXC stream could extract words by computing nonadjacent TPs, but could not use this information to discover that all words conformed to a common structure.

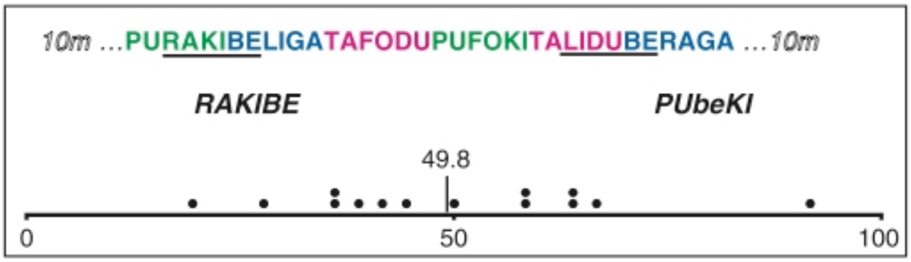


Figure 4: The figure represents the results of the second experiment by Peña et al., testing generalization after exposure to a continuous stream. In this experiment, each dot represents the percentage of choices for rule-words (against part-words) of individual subjects averaged across items.

Why? A possibility suggested by Peña et al. is that finding words is a distinct task from finding structural information about words: it may require a different computation and it may act upon a specific input. They conjectured that a change in the signal would induce a change in computation, hypothesizing that the relevant factor would be the way in which the input stream is packaged. To test this hypothesis, they inserted subtle pauses of 25 ms between words, so that segmentation cues would relieve learners from computing probabilities and making them able to capture the generalizations they would otherwise ignore. They tested participants with part-words versus rule-words – exactly the same contrasts used in their second experiment. They predicted that the stream containing segmentation cues - even if only subliminal - would prompt participants to acquire its structure.

Results are illustrated in Figure 5. Participants preferred rule-words to part-words, even though rule-words never appeared in the familiarization stream while part-words did.

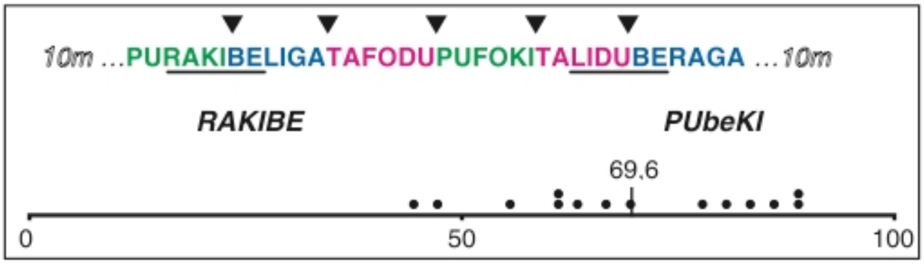


Figure 5: The figure represents the results of the third experiment by Peña et al., testing generalization after exposure to a subliminally segmented stream. Pauses of 25 ms at the edges of words are indicated by the black triangles above the first line. In this experiment, each dot represents the percentage of choices for rule-words (against part-words) of individual subjects averaged across items.

Peña et al. interpreted this result as evidence that the introduction of a minor change in the signal (i.e., the presence of silent gaps between words) induced learners to spontaneously generalize to structural information. Indeed, participants were persuaded that rule-words were more likely to be the words of the imaginary language than part-words, even if they never occurred in the stream, while part-words did -- but with lower frequency than words. Indeed, in a stream where all words occur with the same frequencies, part-words occur half as frequently as words. However, Peña and collaborator ran another experiment to control for the absolute frequency of syllable co-occurrence (Peña et al. 2002, footnote 16). They constructed a stream where each family of words contained one high-frequency item and two items occurring half as frequently. As a consequence, part-words spanning the boundaries of high-frequency words were as frequent as low-frequency words. Participants were tested with part-words and low-frequency words, thus with couples equalized in their frequency of occurrence. Results showed that participants preferred low-frequency words against part-words. Therefore, the authors concluded that the preference was not directed by the items' absolute frequency; rather, it was determined by the computation of distant TPs among syllables.

However, there is another possibility to explain why learners could extract words by computing nonadjacent TPs from the continuous stream (Experiment 1), but failed to discover the word structure (Experiment 2). We may think that extracting words is easier than discovering word structure, as words are items that learners listened to during familiarization. It is then possible that the discovery of structural regularities, being a difficult task, requires a larger amount of experience with the linguistic input to be accomplished, as hypothesized by the statistical learning approach (Seidenberg, MacDonald, & Saffran, 2002). Would an increased amount of exposure led participants to capture the word structure? To test the possibility, Peña et al. performed a fourth experiment, where participants were exposed to a 30-minutes continuous stream (i.e., three times the duration of the original stream) and were tested with part-words and rule-words (as in Experiments 2 and 3).

Results are presented in Figure 6. After such a long familiarization, participants preferred part-words against rule-words, that is, they showed the *opposite* performance as predicted by the statistical learning approach to account for the acquisition of structure. To put it another way, a massive exposure did not help retrieving information about the structure of words. In contrast, the result conforms to the prediction of Peña et al.: an increased exposure helped consolidating statistically occurring patterns, rather than inducing the projection of generalizations. In other terms, the same statistical computations that took advantage of a greater experience with the linguistic input did not give rinse to grammatical structure.

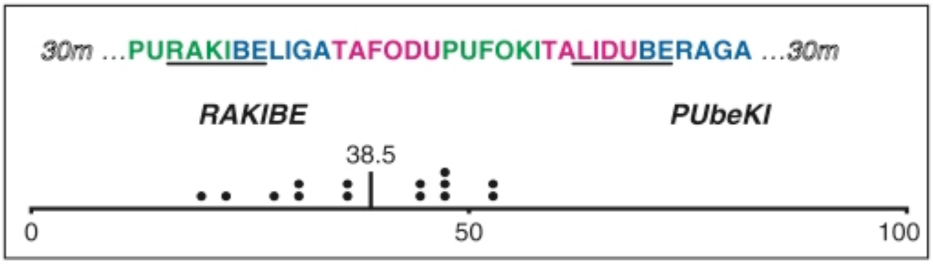


Figure 6: The figure represents the results of the fourth experiment by Peña et al., testing generalization after exposure to 30-minutes long continuous stream, that is, three times longer than the stream of their second experiment (see Figure 4). In this experiment, each dot represents the percentage of choices for rule-words (against part-words) of individual subjects averaged across items.

What then might trigger computations that induce learners to project structural regularities? Peña et al. proposed that the mechanism responsible for generalizations is not statistical in nature – as it does not benefit from a massive exposure to the input and it does not suffer from a reduced experience with it. In contrast, it can be activated after a reduced exposure to few exemplars, as it can quickly extract regularities from them, and it rapidly projects such regularities to novel instances. However, the generalization mechanism requires *segmentation cues* present in the signal to be activated. This is because the presence of bracketing cues may help participants to segment the speech flow, without any need of exploiting TP to isolate words. To test these hypotheses, Peña and collaborators reduced familiarization to a subliminally segmented stream by a factor of five, and allowed only 2 minutes of exposure. Participants were tested with part-words against rule-words (as in Experiment 3). Results showed that participants preferred rule-words to part-words, thus indicating that a brief exposure sufficed for structural generalizations to be computed (see Figure 7). This finding is compatible with the original hypotheses of Peña et al.: generalizations arose rapidly, providing the same performance as with a five times longer exposure, but this happens only when subliminal segmentation cues were available.

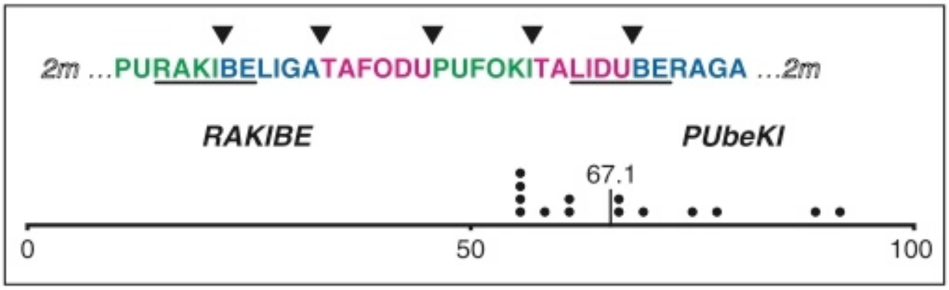


Figure 7: The figure represents the results of the fifth experiment by Peña et al., testing generalization after a short exposure to a subliminally segmented stream; 25ms silences at the edge of words are represented by the black triangles above the first line. In this experiment, each dot represents the percentage of choices for rule-words (against part-words) of individual subjects averaged across items.

Peña et al. took the failure in generalizing after a long familiarization and the success in projected structural regularities as evidence that two different computational processes that can be triggered by subtle difference in the signal. One process is biased toward the discovery of statistical patterns, and its role is confined to the individuation of occurring patterns. It operates slowly over a continuum; it needs to gather evidence to identify the components and to break to flux into its units. The other mechanism is oriented toward the discovery of the grammatical structure of the tokens discovered after breaking the continua. It acts quickly on discrete items, thus it requires silent gaps making the speech stream discontinuous.

The assumption that two, or more mechanisms are involved in the discovery of statistically occurring patterns and in the generalization of structural information stands at the basis of the “MOM hypothesis”. According to it, a plurality of mechanisms (at least two) can exist to assist altogether the language acquisition process. The proposals that distinct mechanisms are needed to account for language learning is well known in other areas of language, especially in the debate concerning how the English past tense is mentally represented. The so-called “dual route models” posit that both associationist and non-associationist mechanisms are required to store the irregular past tense or words (via some forms of pattern-association processes) and to form the regular past tense or regular inflected words (via by rule-based procedures: Pinker, & Ullman, 2002; Prasada, & Pinker, 1993; Pinker, & Prince, 1988; Marcus, et al., 1992).

However, the dual-route models are mostly concerned with the kinds of representations necessary to account for language competence, while the “MOM hypothesis” is concerned with the learning mechanisms needed to attain such representations. For example, a model like Pinker’s (1991) “rule-and-exceptions” model is compatible with the view that both rules and exceptions are learned by virtue of a single, inductive, learning mechanism, because it focuses more on the final result of the learning process than on the mechanism generating it. In contrast, Peña et al.’s model focuses on the nature of the learning mechanisms that language learners (and, more specifically, adult learners) can recruit *on-line* to extract such representations. Hence, the MOM hypothesis represents a theoretical account on the nature of the mechanisms involved in language processing, and offers specific predictions on the expected performances on different linguistic tasks.

4.2 Contrasting views on the role of nonadjacent dependencies in the acquisition of linguistic structure

There is an extremely important result collected by Peña et al. on the learning of nonadjacent dependencies: adults could track remote dependencies and could use them to isolate words from a continuous speech stream, but the same statistical information could not aid them to discover and project structural regularities to novel instances. Instead, generalizations quickly arose after exposure to an overtly segmented stream. Only the presence of silences at the edges of words induced learners to capture structural regularities between nonadjacent elements they would otherwise ignore. Remarkably, a long familiarization did not help retrieving information about the structure of words. This result is extremely relevant as it contradicts the general predictions of associative learning theories.

Such theories assign to statistical learning a primary role in determining the acquisition of morphosyntactic structures, and assume that the same mechanisms computing adjacent dependencies may be themselves capable of a broader range of computations, among both adjacent and non-adjacent elements. Hence, mechanisms of the same nature can in principle account both for the extraction of a lexicon and for the acquisition of morphosyntactic regularities. For instance, one possibility for syntax learning is that higher-order TPs build on lower-level ones, such that first-order TPs are embedded in knowledge of second-order ones - where first-order TPs require knowledge of the immediately preceding element and second-order TPs require knowledge of the previous two elements. In this view, learners first chunk adjacent dependencies, then form a higher-order relation between the chunked pair and the next element in sequence. This is what several studies on nonadjacent dependency learning in infants suggested to account for the early acquisition of linguistic structure (Gómez, 2002; Gómez & Maye, 2005; Gómez & Lakusta, 2004; Lany & Gómez, 2008).

In contrast to this view, the MOM hypothesis assumes that second-order TPs do not build upon first-order TPs. Rather, they may enter into different computations: they may serve to isolate words from a continuous stream, or to discover word structure after exposure to a segmented stream. This latter computation is assumed to determine the discovery syntactic rules from the linguistic input without exploiting its distributional information. Therefore, this approach postulates that the statistical learning mechanism involved in the identification of high TP linguistic patterns (i.e., words) does not have a role in the acquisition of morphology and syntax. Rather, another mechanism, sensitive to other properties than statistics, is required to explain the acquisition of linguistic structure.

Even if, in principle, one single mechanisms accounting for language acquisition in all its complexity may suffice for acquiring both the lexicon and the grammar of a language, computing statistical regularities over a linguistic input may be a nontrivial task. To learn the patterns of language and to find consistent non-adjacent regularities, learners might have to keep track of the probabilities relating all the syllables one away, two away, three away, etc. If such a device were to keep track of regularities among many types of elements—syllables, features, phonemic segments, and the like—this problem grows exponentially. In principle, statistical computations on adjacent and nonadjacent TPs would lead to an unmanageable explosion in the number of computations that must be performed to do the learning (Chomsky, 1965, 1980; Wexler & Cullicover, 1980). However, non-adjacent regularities in natural languages take only certain forms, thus the range of possibilities is actually restricted. Still, are language learners capable of extracting not only adjacent regularities, but also nonadjacent ones from a continuous speech stream? The next section deals with this question.

4.2.1. Can learners acquire nonadjacent dependencies from a continuous speech stream?

To better understand the limits and the constraints of the statistical learning capacity, Newport and Aslin (2004) asked whether language learners are capable of extracting nonadjacent ones from a *continuous* speech stream. They familiarized adult participants with an artificial language constructed in a way so that learners could extract word-like groupings only by computing various types of non-adjacent regularities; adjacent elements were controlled so that no grouping occurred if only adjacent regularities were computed. Specifically, the stream contained 20 nonsense words, created by five regular "word frames" (i.e., sets of pairs of nonadjacent syllables occurring with a nonadjacent TP of 1.00) and four middle syllables. The same middle syllables could occur inside all five nonadjacent word frames (thus, adjacent TP within the sequences was equal to .25). Participants to the studies of Newport and Aslin showed no evidence of learning the nonadjacent syllable pairs. Different variations of the experimental task, including increased exposure, simplified language structure, and explicit learning procedure, did not yield any significant result. In contrast, participants readily acquired regularities among nonadjacent segments – both consonants and vowels --, suggesting that adult learners perform their computations on segments rather than on syllables. What Newport and Aslin (2004) concluded from their results is that adult learners are unable to acquire statistical groupings based on syllables, but they succeed in computing remote dependencies over consonants or vowels.

The results by Peña et al. (2002) and by Newport and Aslin (2004) seem to provide contradictory evidence on nonadjacent dependency learning, the former showing that learners could accomplish segmentation on the basis of nonadjacent TPs computations, the latter finding that such learning could not occur over syllables, unless specific cues where present in the signal. The two cues that have been considered as possible factors accounting for the difference between these two set of studies were the specific phonetic and phonological properties of the syllables used to create the material, and the presence of silences between words.

The effect of prior information about the phonetic and phonological properties of words in native language was one of the main criticism to the Peña et al.'s results. Seidenberg, MacDonald, and Saffran (2002) suggested that there was a potential confound in the material they used, in that the first and third syllables began with plosives and the intervening syllable began with a continuant. Thus, responses could have been driven by sound regularities rather than the structure of the language. Onnis et al. (2005) removed this confound and found no evidence for either segmentation or generalization in such AXC languages. However, Peña et al. (2002) already discussed this point in their footnote 17, arguing that in adults such features do play a role in facilitating word extraction, although they are not sufficient to account for the projection of structural generalizations (see also Bonatti, Peña, Nespor, & Mehler, 2006 for a more through investigation of phonological effects).

Concerning the presence of silences between words, it is important to note that Peña et al. used them to mark word boundaries, as did other studies investigating nonadjacent dependency learning in infants and in adults (see Gómez, 2002, and Gómez and Maye, 2005; see also Section 3.2.1.1. on Chapter 3). Both of them reported successful learning of remote dependencies (placed *within* words in Peña et al, and *between* words in the other studies), while Aslin and Newport did not. Then, the presence of silence gaps, bracketing the speech signal into units, seems to be crucial to determine such kind of learning.

We may speculate about the role of silences in aiding nonadjacent dependency learning. It is possible that a stream already parsed would aid learners in discovering which patterns they should attend to, and would help them focusing their attention to the relevant units to track, i.e., the elements appearing at the initial and final positions of the strings (words-like units in Peña et al., sentence-like units in Gómez, 2002, and Gómez & Maye, 2005). A continuous stream would, in contrast, maintain the difficulty of learning nonadjacent syllable regularities from continuous streams, as attested in several studies (Peña et al., 2002; Endress, & Bonatti, 2007).

4.3 Summary

This chapter presented the controversy between the most relevant approaches that have been proposed to account for the acquisition of grammar-like structures: the rule-based approach, the associative learning approach, and the MOM hypothesis. The three perspectives differ in the relevance they attribute to statistical computations in explaining the acquisition of linguistic structures. The MOM hypothesis appears to be an intriguing proposal as it preserves the main tenets of the rule-based approach, while maintaining the need for statistical computations.

Theoretical controversies exist also about the nature of the mechanism accounting for the acquisition of linguistic structure in early development. The MOM hypothesis may provide some insights about how infants may begin to master linguistic structure, and on the developmental course of this process. It may allow investigating whether infants possess and recruit distinct mechanisms, one dedicated to isolating words from continuous speech, the other one at generalization structural dependencies within words from segmented stream (as adult do). The next two chapters will present a series of artificial grammar studies I have conducted with young infants, seeking to address these hypotheses.

CHAPTER 5

INVESTIGATIONS INTO EARLY MORPHOSYNTACTIC ACQUISITION MECHANISMS: ARTIFICIAL GRAMMAR STUDIES WITH 7- AND 12- MONTH-OLD INFANTS

To achieve language proficiency, language learners must find the building blocks of speech – words - from a speech flow where no clear signs of segmentation are marking word boundaries. In addition, to use language productively, they have to discover and master the rules governing the legal combinations of words (see Chapter 3). Yet, the problems of finding the basic elements of language and that of identifying the structural information that makes language productive are not independent. Indeed, natural languages realize morphosyntactic dependencies as relations within words. As such, many languages express syntax as relations between nonadjacent word subparts (see

Chapter 2). Recent evidence suggests that adult learners recruit at least two distinct mechanisms to segment words and to learn structural information within them (Peña et al., 2002; see Chapter 4). Instead, whether infants also use separate mechanisms to identify words from speech and to learn morphosyntactic relations inside word structure is not known.

Precisely to investigate the nature and the limits of the mechanisms young learners possess to learn words and within-word regularities, I conducted a series of artificial grammar learning studies with infants. In these experiments, I test the hypothesis that infants possess and recruit different mechanisms when processing language. One mechanism is assumed to be statistical in nature, and it may contribute infants to segment words from continuous speech. The other mechanism is assumed to project generalizations after exposure to a bracketed linguistic input, and it may be contribute at acquiring linguistic structures. Importantly, I investigated whether these two mechanisms are triggered by different signal properties, as predicted by the "MOM hypothesis" (Peña et al., 2002; see Section 4.1.2 on Chapter 4).

The chapter is organized as follows. The first part is aimed at reviewing the evidence attesting that preverbal infants can compute simple statistics (namely, adjacent TPs), and can extract and generalize simple rules. The existing literature will be discussed with respect to the problem of how infants acquire morphosyntactic structure – a problem for which very little data is currently available. Theories disagree on the nature of the mechanisms accounting for language acquisition in all its complexity: the second part of the chapter takes up the controversy and presents the MOM hypothesis as an alternative view on the nature of the mechanisms assisting infants in word segmentation and in word structure acquisition. The third part of the chapter is devoted to presenting the series of artificial grammar learning experiments with 7- and 12-month-olds investigating whether young infants possess and use distinct computational mechanisms to solve the tasks of isolating words from fluent speech and of discovering word structure. Finally, the results will be discussed in light of the attested literature on statistical computation and generalization abilities in infants.

5.1 Learning words and within-word rules: theoretical questions and experimental hypotheses

5.1.1 Open issues in morphosyntactic acquisition

In the last decade, a growing interest for the abilities assisting infants in acquiring the words and the rules of their native language determined a better understanding of the early linguistic computational skills (see Chapter 3). For instance, we know that, at 8 months, infants are able to track distributional properties, such as adjacent TPs among syllables; such sensitivity may help young learners segmenting a continuous artificial stream into its components (Saffran, Aslin, & Newport, 1996; Aslin, Saffran, & Newport, 1998) and, in principle, it may also help them extracting real words from real speech. At a later age, i.e., at around 18 months, they can also track nonadjacent relations between discrete, word-like items (e.g., Gómez & Gerken, 1999; Gómez, 2002; Gómez & Maye, 2005); accordingly to some authors, the ability to learn relations among distant elements may aid young children acquiring syntactic structure (see Section 4.2 on Chapter 4 for contrasting views on this possibility). In addition to the sensitivity to TPs, infants are capable of extracting and generalizing structural information both between adjacent (Gómez & Lakusta, 2004; Gómez & Gerken, 1999) and nonadjacent (Marcus et al., 1999) relations (the latter requiring the repetition of the distant element; see Section 3.2.3.1 and Section 3.2.3.2 on Chapter 3); the ability to generalize structural information and to apply it to novel instances has been proposed to account for the productive use of language.

Even if much progress has been done to understand in what kind of tasks such computational abilities would aid infant learning, our understanding of the *nature* and the *limits* of the learning mechanism infants possess to acquire linguistic structure is still poor. For example, the studies investigating the ability to track dependencies between remote elements familiarized infants with lengthy pauses between elements (Gómez, 2002; Gómez & Maye, 2005); thus, we do not know whether infants are also able to compute TPs among nonadjacent elements in fluent speech -- an ability that could help them extracting structural relations -- and if so, at what age. Moreover, it is not known whether infants can track nonadjacent elements *within* a linguistic unit – such as a word.

Asking whether infants can learn structures between nonadjacent elements inside a word unit is a question of particular relevance, as in natural languages many morphosyntactic processes are expressed as relations between distant morphemes both between and within the words (see Chapter 2, and Section 3.2.1 on Chapter 3). Thus, the ability to analyze words in their sub-parts and to recognize a rule-governed structure represents a *requisite* for syntax acquisition. Even if language acquisition demands discovering both words and morphosyntactic rules from the speech input, still attaining both kinds of learning may represent a difficult task, since it requires to perform computations of different types across the same set of units - that is, learners should detect different properties on the same patterns of sub-lexical units. We do not know whether infants are able to compute different types of information from the same speech input to the purpose of accomplishing the two distinct (but related) tasks of discovering words and their compositional structure. And, if so, we do not know whether they attain both types of learning at the same age, or else, if developmental delays between the two abilities should be expected.

A crucial issue lying below the surface of these problems is the nature of the mechanisms responsible for infants' linguistic abilities (see Chapter 4 for a broad discussion on this issue). One possibility is that a single mechanism, computing statistical relations over speech, helps infants finding words and syntactic regularities (Elman et al., 1996; Elman, 1999; Elman, 2001; Altmann, 2002; see Section 4.1.1 on Chapter 4). Possibly, infants may compute several kinds of statistical relations over the same stream, and bootstrap lexicon and grammar in the same fashion (Bates, & Goodman, 1999; Elman, 2004; Marchman & Bates, 1994). Another possibility is that building a lexicon from a continuous stream and finding the structure of lexical items rely on different mechanisms, governed by specific properties of the input signal (Peña et al., 2002; Bonatti, 2008). Several studies attested that adult learners do recruit two distinct mechanisms to solve different linguistic tasks (see Section 4.1.2.1 on Chapter 4). In contrast, data on how infants may solve both the word segmentation and the word structure generalization tasks are less complete. Whether they can compute different kinds of statistics (such as first- and second-order TPs) and they use them also to acquire morphosyntactic regularities is not clear yet. The experiments reported in this chapter aim to explore whether infants – as adults- possess and recruit distinct mechanisms for accomplishing word segmentation and for discovering structural regularities within word.

5.1.2 Mechanisms assisting early morphosyntactic acquisition: A hypothesis

The "MOM hypothesis" may provide insights about how infants begin to master within-word grammatical relations. Furthermore, it may offer an explanation both to the data attesting the infants can compute TPs to identify words (e.g., Saffran et al., 1996) and to the data showing that they can generalize abstract patterns and use them to recognize novel structures as grammatical (e.g., Marcus et al., 1999). Remarkably, it may allow addressing the issues of whether infants possess and recruit distinct mechanisms, one dedicated at isolating words from continuous speech, the other one at generalization structural dependencies within words from segmented stream (as adult do), whether the two mechanisms are sensitive to different signal properties, and whether they efficiently available at the same age, or else whether they have different developmental trajectories.

The MOM hypothesis generates specific predictions about the expected performances on word segmentation and on structural generalizations. First, if two mechanisms devoted to different computations and triggered by differences in the signal exist, then the nature of the linguistic stream may activate either one or the other. Thus, infants should be able to project generalizations after exposure to a segmented stream, but should fail if exposed to a continuous stream. In contrast, they should successfully extract words from the continuous stream on the basis of statistical cues, but should fail to generalize.

Second, if the generalization mechanism is an effective tool for morphological and syntactic acquisition given only scant input (Endress & Bonatti, 2007; Bonatti, 2008), then infants, like adults, should be able to find abstract relations within words and between nonadjacent word subcomponents after a brief exposure to the input (that is, without the need of extensive experience with it).

5.1.3 Plan of the studies

The artificial grammar studies with infants reported in this chapter were conducted using a modified head-turn preference procedure. This procedure is one of the most used procedures to study infant speech perception and processing abilities, and to establish how these abilities change as a function of experience and/or development. This procedure has been of considerable interest and of vast application because it allows testing pre-verbal infants on several auditory and linguistic tasks (such as language discrimination, word segmentation, category perception, etc.) in a non-invasive fashion (Werker, Polka, & Pegg, 1997).

The studies were conducted with infants of 12 and 7 months. These ages were chosen ad they correspond to the achievement of important milestones in the linguistic domain. Infants at 7 months have been shown to be capable of extracting words from running speech (Jusczyk, 1999) and one cue they consider to identify sound sequences contained in a stream are adjacent TPs (Saffran et al., 1996). At the same age, they can also learn and generalize abstract patterns involving repetition (Marcus et al., 1999). In addition, at 12 months are already well zoomed into their native language, starting fixating their phonology, associating words with meanings, and so on (Werker & Tees, 1984). To date, it is not known whether infants would also display sensitivity to morphosyntax by their first year of life, i.e., before revealing the ability to track dependencies between distant words (Santelmann & Jusczyk, 1998; Gómez, 2002; Gómez & Maye, 2005).

In order to study whether infants recruit distinct mechanisms, one to segment words from a continuous stream, the other one to project structural regularities, I adapted the Peña et al.'s AXC artificial languages to infants. Such languages are suitable for testing both word identification and structure learning because the nonsense words contained in them can be statistically defined by their adjacent and nonadjacent TPs, but they can also be structurally defined in virtue of the nonadjacent dependency between the first and the last syllable. Importantly, the generalizations are inside words -- and not between words, as in the other studies investigating the learning of structural regularities did (for instance, Gómez, 2002; Gómez & Maye, 2005; Marcus et al., 1999). Moreover, words contain structural information between two different nonadjacent syllables. As mentioned in Chapter 3, in most natural languages grammatical rules involve dependencies between nonadjacent elements, which usually are different morphemes. As such, grammatical rules in real language are usually not cued by the identity of the physical stimuli (as in Marcus et al.'s studies; see Section 3.2.3.1 on Chapter 3). For this reason, the exposure to AXC streams may better simulate morphosyntactic structures, and thus allow one to investigate the onset of morphosyntactic sensitivity in infants.

Along eleven experiments, I test both word and word-structure identification after exposure to either a continuous or a segmented stream, where the only structural relations were among nonadjacent syllables.

In Experiments 1-3, I investigate the generalization abilities in 12 month-olds. To the purpose of controlling for the possible effect of phonetic and phonological features in influencing generalizations, familiarization words and test items varied across the three experiments. In Experiment 4, I test whether 12-month-olds can generalize after being exposed to a continuous stream. In Experiment 5, I assess whether they can extract words from the continuous stream by computing TPs.

In Experiments 6 and 7, I investigate whether generalization is also available to younger infants, i.e., at 7 months. In Experiments 8 and 9, I test whether 7-month-olds can learn words from segmented streams. In Experiments 10 and 11, I ask whether they can extract words from continuous streams by computing TPs.

5.2 Experiment 1: generalizations after exposure to a segmented stream in 12-month-olds

5.2.1 Method

5.2.1.1 Participants

Sixteen 12-month-old infants (7 girls; mean age: 12 months 25 days; age-range: 12 months 9 days - 13 months 15 days) participated in Experiment 1 and were retained for analysis. All participants to this and to the other infant studies reported in the thesis were healthy, full-term babies, with a gestational age > 37 weeks and Apgar scores > 7 one and five minutes after birth. Their parents reported no hearing or vision problems present at and/or prior to their participation to the experiments. An additional 16 infants participated in the experiment but were excluded from analysis due to the following reasons: because of excessive fussiness during familiarization or during test phase, proving average looking times for less than 12 test trials (12), or because they exceeded maximum looking time criteria, looking longer than 65 cumulative seconds in more than two test trials (4).

All infants participating to this and to the other infant studies were tested at the Language, Cognition and Development Laboratory at SISSA, Trieste, Italy. A parent for each participant gave informed consent prior to participation. The study was approved by the Ethics Committee of SISSA.

5.2.1.2 Material

5.2.1.2.1 Familiarization stream

To the purpose of testing young infants, I adapted and simplified the AXC-language used by Peña et al. in their adult studies. An artificial speech stream was created by pseudo-randomly concatenating four nonsense words, with the constraint that the same word could not occur twice in a row. The words of this language consisted of sequences of three Consonant-Vowel (CV) syllables, with nonadjacent TP of 1.00, and adjacent TP of .50. Adjacent TPs across word boundaries were also .50. The language contained two word couples with identical first and last syllables (e.g., /ba/-/so/ and /li/-/fe/), differing only in their middle syllable that was pooled from a set of two syllables (e.g., /ga/ and /mu/; see Table x). Thus, the language had two A-C frames characterized by minimal variability. During the stream, words were repeated 48 times and were separated by 200ms silences. Pauses of this duration (that is, not subliminal) were inserted between words to compensate for the difference in acoustic properties between sounds presented via high-quality headphones (as Peña et al. did with adult participants) and a stream played in a testing booth affected by minor sound reverberation and background.

5.2.1.2.1Test items

The test items were either "non-words" or "rule-words". *Non-words* were novel trisyllabic sequences composed of syllables that appeared in the stream, but never in that position. Non-words were four in total. Two of them had a CBA' structure (where C and B stand for the third and second syllable of one A-C frame, and A' for the first syllable of the other frame). The other two had an A'CB structure (see Table 1). *Rule-words* were novel trisyllabic sequences in which the A and C syllables were the same as in words, but the middle syllable varied: it was a syllable that appeared in the stream, but never in that position (i.e., it was the initial of final syllable of the other A-C frame). Rule-words were four in total (see Table 1). Thus, two non-words shared their initial syllable with words and two did not. Non-words had 0 frequency, as well as both adjacent and nonadjacent TPs equal to 0. Rule-words shared the initial and the final syllables with words; they had 0 frequency, 0 adjacent TPs, but a nonadjacent TP of 1.00.

The material was synthesized with the Mbrola speech synthesizer (Dutoit, Pagel, Bataille, & Vreken, 1996), using the FR-2 diphone database[9], setting flat

---

[9] Pilot tests with native participants revealed that Italian native speakers find synthesized speech with the fr2 diphone data base more intelligible than speech synthesized with the

prosody to sound monotonous, 116 ms phoneme length and 200 Hz pitch. Words' mean length was 696 ms. In order to avoid direct cues to word onsets, the familiarization stream was synthesized with increasing and decreasing amplitude ramps in the first and last 5 s, respectively. The familiarization stream lasted 2 m 52 s.

Words	**Rule-Words**	**Non-Words**	
		Experiment 1	**Experiment 2**
/**ba**mu**so**/	/**ba**li**so**/	/so**ga**li/	/so**ga**li/
/**ba**ga**so**/	/**ba**fe**so**/	/fe**mu**ba/	/fe**mu**ba/
/**li**mu**fe/**	/**li**ba**fe**/	/**li**soga/	/muba**fe**/
/**li**ga**fe**/	/**li**so**fe**/	/**ba**femu/	/gali**so**/

Table 1: Items used to compose the familiarization stream and used in the test phase of Experiments 1 and 2. Underlined font indicates the syllables that define the A-C frames. Boldface font indicates those syllables appearing in the same position as in familiarization words.

5.2.1.3 Procedure

Infants were tested in a modified version of the head-turn preference procedure (Kemler-Nelson, Jusczyk, Mandel, & Myers, 1995). They sat on their caretaker's laps, in a dimly lit, quiet room, with three monitors positioned at their front and sides. The caretakers listened to masking music and were instructed not to interact with infants during the experiment (see Figure 8).

---

available MBROLA Italian diphone data bases. For this reason, I decided to use fr2. Obviously, all phonemes selected to create the auditory material also exist in Italian.

Figure 8: The experimental setup of the head-turn preference paradigm used in all the infant studies reported in the thesis. The picture is taken from the top of the left corner of the testing booth. The three monitors at the center and at the sides are in light blue just to show their arrangement inside the booth. During the experiment, they would otherwise be in black, with only exception of the colored visual attractor (see Procedure). The camera recording the infant's looking behavior is located above the central monitor; it is not visible from the picture as it is hidden behind the black curtain.

During familiarization, a visual stimulus (a recorded movie of a moving hand) attracted infants' attention towards the center, while the speech stream was played. After familiarization, infants were tested in 16 test trials, i.e., two trials for each of the 8 test stimuli (4 non-words, 4 rule-words) presented in pseudo-random order, that is, with the following constraints: *(i)* the same item could not be immediately repeated in the next trial, and *(ii)* a maximum of three test items of the same type (rule-words or non-words) could occur in three subsequent trials. The order of item presentation was counterbalanced as a between participants factor. Each trial started with the visual attractor (i.e., the moving hand as in familiarization) appearing at the center. Once infants attended to it and fixated it for a continuous period of 1.5 s, the moving hand disappeared from the center and reappeared on one of the side monitors. As infants stably oriented towards it (defined as a 45° head turn toward the visual attractor), the test item started playing repeatedly from the loudspeaker from the corresponding side. The sound file was repeated with a 500 ms ISI, and continued until infants looked away for 2 s consecutively or looked up to 65 s cumulative. Afterwards, a new test trial began.

An Apple G5 controlled by PsyScope X (http://psy.ck.sissa.it/) ran the experiment. A camera hidden behind the center monitor was recording infants' looking behavior thus allowing the experiment to control online the experimental procedure, that is, starting and ending test trials depending on the infant's looking behavior. The camera also allowed checking the infants' comfort during session, and providing a permanent record of his/her looking behavior.

5.2.2 Results

Videotapes of infants' looking behavior were coded off-line. Looking times were averaged across all trials of the same Test Item Type (Non-word, Rule-word). Average looking times shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type, were excluded from further analysis. With these criteria, the overall excluded data amounted to 2.9% of the total (7 out of 243 LTs data points). A repeated measure ANOVA with Test Item Type (2 levels: Non-word, Rule-word) as within-participant factor was conducted, using mean looking times as dependent measure. Infants looked longer while listening to non-words than to rule-words ($M_{\text{Non-words}} = 8.89$s, SE = 0.64; $M_{\text{Rule-words}} = 7.67$s, SE = .68, $F(1, 15) = 7.97$, $p < 0.02$; Figure 9).

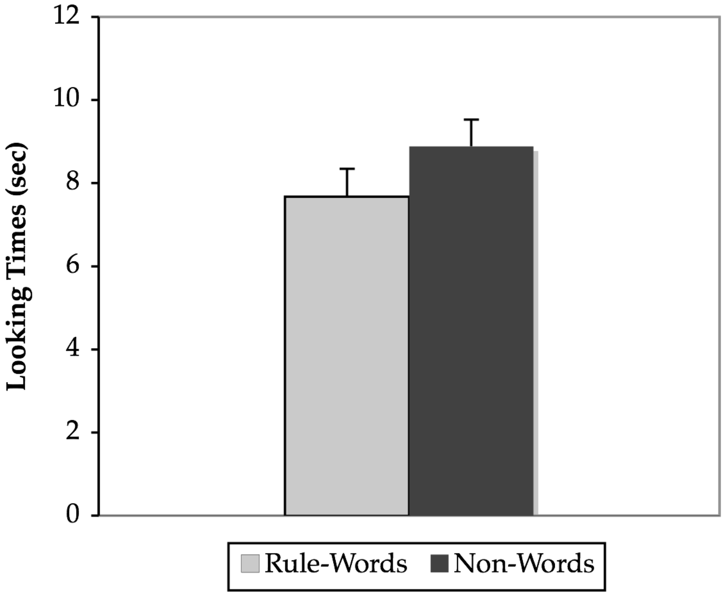


Figure 9: Mean looking time(s) and SE for the test items of 12-month-olds listening to rule-words and non-words after a segmented familiarization stream.

5.2.3 Discussion

Given that neither rule-words nor non-words appeared during familiarization, the infants' preference for non-words may indicate that they could learn the structural dependency between the first and the last syllable of words, and generalized it to novel sequences containing it (i.e., rule-words). Given that rule-words had the same nonadjacent TP as words, while non-words did not share any statistical information or structural property with words, longer LTs for non-words than for rule-words may be interpreted as a novelty preference. Importantly, infants could learn the structural dependencies after being exposed to a segmented stream, as predicted by "MOM hypothesis", and as Peña et al. found with adult participants.

While the result suggests that infants prefer legally constructed unheard items to non-legally constructed unheard ones, another alternative hypothesis may explain it. It is possible that infants reacted to local mismatches between syllable positions, and not to the presence of the nonadjacent relation between first and last syllables. This possible interpretation may be suggested by the fact that the experiment compares infant's reactions to rule-words and to non-words. By definition, rule-words have both the A and C syllables identical to words, but non-words do not: if they did, they would also be rule-words (or words). Thus, infants may have reacted to a partial mismatch between rule-words, or familiarization words, and non-words. According to this explanation, instead of representing abstract classes of items defined by syllables in initial and final positions, infants may have only extracted words during the familiarization phase. Then, in the test phase they may have only monitored one single syllable (the first, or the second, or the last) and noticed that the syllable in that position differed between non-words and familiarization words. This alternative explanation of the result of Experiment 1 requires no sensitivity to rule-words in order to account for the differential interest infants allocated to rule-words and non-words.

This explanation is partially made implausible by the construction of the material of Experiment 1. Two of the four non-words used in the test phase had their initial syllable identical to that of words (and hence of rule-words), whereas two non-words did not, having their middle syllable identical to those of words instead. Therefore, if infants only monitored the first syllable of test items, looking for partial matches between test items and their memories of familiarization words, they should find surprising only the non-words whose initial syllable did not match that of familiarization words, but they should find no difference between those non-words with initial syllables identical to familiarization words and rule-words. This argument predicts that infants should look longer at non-words than at rule-words only for the two non-words whose first syllable differed from that of words. By the same argument, if infants monitored only the second syllable of test items and looked for differences with familiarization words in that position, they should look longer only at the non-words with a middle syllable different to that of words, but not so at the non-words having a middle syllable identical to that of familiarization words. To test this possibility, I assessed whether there was any difference between the trials where infants listened to non-words sharing the first syllable with words and those in which they maintained the middle one. There was no hint of a difference among them, t(15) = .26, p = .80, ns, paired t-test.

Therefore, the construction of the material of Experiment 1 and its results exclude that infants merely attended to the first or the second position of the test items. However, Experiment 1 does not exclude that infants only monitored differences between test items and the *last* syllables of familiarization words, perhaps because of recency effect induced by their rehearsal of words in memory. To control for this possibility, I ran Experiment 2. In it, the familiarization and the rule-words were maintained identical to Experiment 1, but two novel non-words were created. These new non-words, instead of having the same first syllables as words, retained the same last syllables; therefore, they had a BAC′ structure. Differential looking behaviors when listening to non-words than to rule-words would indicate sensitivity to the structural properties of rule-words. In contrast, absence of difference in looking behavior to non-words having last syllables identical to that of words and those not having it would exclude that the results of Experiments 1 were due to the fact that infants simply monitored the last syllable of items, looking for mismatches between familiarization and test items.

5.3 Experiment 2: controlling for the effect of test item structure

5.3.1 Method

5.3.1.1 Participants

Sixteen 12-month-old infants (8 girls; mean age: 12 months 29 days; age-range: 12 months 14 days - 13 months 15 days) participated in Experiment 2 and were retained for analysis. An additional 14 infants took part but were excluded from analysis because of excessive fussiness.

5.3.1.2 Material and Procedure

Infants were familiarized with the same stream used in Experiment 1. After familiarization, they listened to the same test items as in Experiment 1, except for the fact that the two non-words with AC'B structure were replaced with novel non-words having BAC' structure (see Table 1). Otherwise, procedure and data analysis were identical to Experiment 1.

5.3.2 Results and discussion

In Experiment 2, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the mean computed for each Test Item Type amounted to 2.7% of the total (6 out of 221 data points). The repeated measure ANOVA revealed that, as in Experiment 1, infants looked longer while listening to non-words than to rule-words ($M_{\text{Non-words}}$ = 11.76s, SE = .99; $M_{\text{Rule-words}}$ = 9.50s, SE = 1.13, F(1, 15) = 4.58, $p < 0.05$; Figure). Furthermore, as in Experiment 1, there was no difference between those trials in which infants listened to non-words with last syllable identical to that of words and those with the middle syllable identical to that of words, t(15) = .395, p = .699, ns, paired t-test. Thus, Experiment 2 shows that infants did not look longer at non-words simply because their last syllable is different from that of words or rule-words.

Taken together, the results of Experiments 1 and 2 show that infants do not pay attention to one single syllable position within the test items. Experiment 1 shows that the initial or middle syllables alone do not suffice, and Experiment 2 shows that the final syllable alone does not suffice either. Rather, infants categorized as structurally identical only those novel sequences retaining both their initial *and* final syllables identical to those of words, suggesting that they projected structural generalizations on the basis of nonadjacent relations within words, or word classes (defining a word class as a set of items sharing the same initial and final syllables).

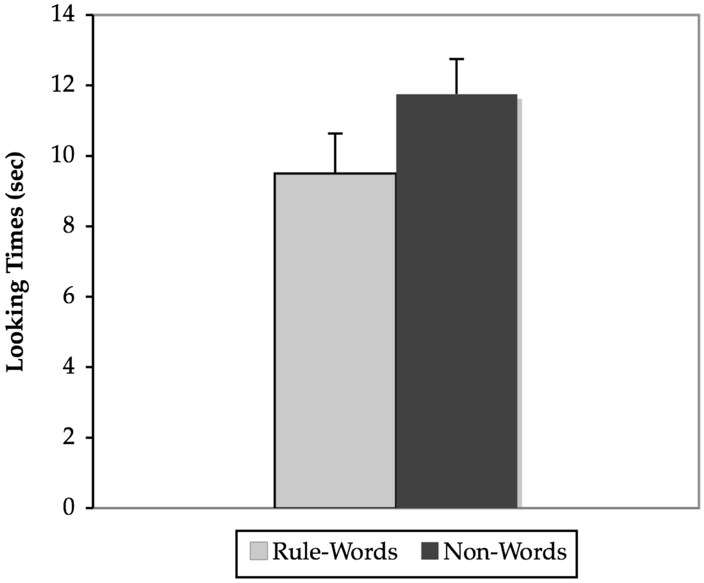


Figure 10: Mean looking time(s) and SE for the test items of 12-month-olds listening to rule-words and non-words after exposure to a segmented familiarization stream. Non-words of Experiment 2 retained either the middle or the last syllable of familiarization words to control for the possible effect of test item structure in determining generalization. See also Table 1.

However, another possible alternative explanation of the results of Experiments 1 and 2 may hold that infants reacted to some phonetic or phonological features of the material used in these experiments. Specific phonetic and phonological features have been shown to influence the adults’ performance: more specifically, words beginning and ending with a stop consonant are easier to segment than words that begin and end with other types of consonants, possibly because they are favored by the statistical distribution of phonemes within Italian words (Onnis, Monaghan, Richmond, & Chater, 2005; Endress & Bonatti, 2007).

Even if the original stream was designed taking care to include syllables with varied phonological patterns in the stream of Experiments 1 and 2 (thereafter Stream 1), a new AXC was created language in order to assess the role of phonological, phonetic or phonotactic factors (to which infants are sensitive; e.g. Mattys, Jusczyk, Luce, & Morgan, 1999) and Experiment 3 was ran. Because in an artificial language experiment "words" and "non-words" are only so because of the probabilistic relations among syllable sequences, the novel familiarization stream (thereafter, Stream 2) was constructed in such a way that the probability relations among syllables would transform the rule-words of Experiments 1 and 2 into non-words of Experiment 3 (see Table 2, page 67). Due to the statistical structure of words, rule-words and non-words -the former two retaining the same first and last syllables, the latter having adjacent and nonadjacent TPs of 0 with respect to words-, it was impossible to cross both types of test items while maintaining the aforementioned statistical properties.

If infants' preferences after familiarization to Stream 1 were induced by some of its low-level features, or by some aspects of the test items irrelevant to their structure, then their looking behavior with respect to the rule-words and non-words after a familiarization to Stream 2 should change. If, instead, infants still look longer at non-words, then the hypothesis that they reacted to a structural property common to rule-words and absent in non-words would be strengthened.

5.4 Experiment 3: controlling for the effect of phonetic and phonological features of familiarization words

5.4.1 Method

5.4.1.1 Participants

Sixteen 12-month-old infants (5 girls; mean age: 12 months 18 days; age-range: 12 months 2 days – 13 months 2 days) participated in Experiment 3 and were retained for analysis. An additional 9 infants took part but were excluded from analysis due to the following reasons: excessive fussiness (8) or exceeding the maximum looking time criteria (1).

5.4.1.2 Material and procedure

A novel familiarization stream was synthesized by using the same criteria used to construct Stream 1, except that the nonadjacent probability relations among syllables was such that the rule-words of Experiment 3 would become non-words in the present experiment. Due to this constraint, non-words had a BB'A and a BAA' structure. Moreover, four novel rule-words were synthesized, constructed exactly with the same criteria as in Experiments 2. Table 2 presents the words, rule-words and non-words of Stream 2. Procedure and data analysis were otherwise identical to Experiment 2.

Words	**Rule-Words**	**Non-Words**
/**fe**li**ga**/ /**fe**ba**ga**/ /**so**li**mu/** /**so**ba**mu**/	/**fe**so**ga**/ /**fe**mu**ga**/ /**so**ga**mu**/ /**so**fe**mu**/	/baliso/ /bafeso/ /libafe/ /lisofe/

Table 2: Items used to compose the familiarization stream and used in the test phase of Experiment 3. Boldface font indicates the syllables that define the A-C frames. Importantly, non-words of Experiment 3 correspond to rule-words of Experiment 2.

5.4.2 Results

In Experiment 3, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type, amounted to 3.3% of the total (8 out of 245 data points). The repeated measure ANOVA revealed that, as in Experiments 1 and 2, infants looked longer while listening to non-words than to rule-words (M $_{\text{Non-words}}$ = 10.37s, SE = 0.795; M $_{\text{Rule-words}}$ = 9.07s, SE = .56, F(1, 15) = 4.73, $p < 0.05$), despite the change in familiarization stream, in test items and in non-words structure (Figure 11).

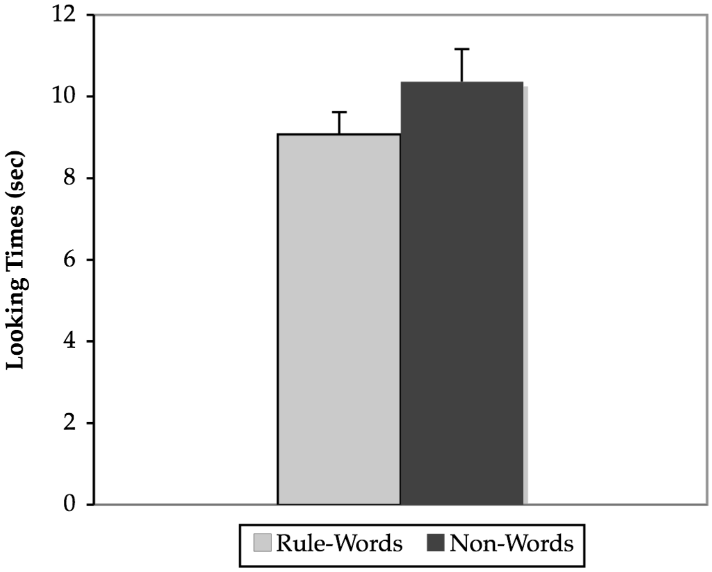


Figure 11: Mean looking time(s) and SE for the test items of 12-month-olds listening to rule-words and non-words after a segmented familiarization stream (Experiment 3). The stream used familiarization words having different phonetic and phonological properties than in the stream used to conduct Experiments 1 and 2.

Because in Experiments 2 and 3 the role of the same test items as rule-words or non-words inverts, in order to better compare the effect of changes in familiarization stream, an ANOVA was conducted by pooling the data of Experiments 2 and 3 together, with Stream (2 levels: Stream 1, Stream 2) as a between-participant factor, and Test Item Type (2 levels: Non-words, Rule-words) as within-participant factor. Infants looked longer while listening to non-words than to rule-words ($M_{\text{Non-words}} = 11.06s$, $SE = 0.64$; $M_{\text{Rule-words}} = 9.29s$, $SE = 0.62$, $F(1, 30) = 8.6$, $p < 0.007$). No other effect was significant. Thus, the role as rule-words or non-words of the test items had an effect, while neither the specific phonetic and phonological characteristics of the familiarization stream did not. An ANOVA pooling the data of Experiments 1 and 3 (in which the familiarization stream also inverts the role of test items, as rule-words of Experiment 1 became non-words of Experiment 3) was conducted, with Stream as a between-participant factor, and Test Item Type as within-participant factor. The results were comparable to the ones of the previous analysis: infants looked longer while listening to non-words than to rule-words ($M_{\text{Non-words}} = 9.63$s, SE = 0.52; $M_{\text{Rule-words}} = 8.37$s, SE = 0.47, $F(1, 30) = 11.66$, $p < 0.002$). No other effect was significant. Taken together, these analyses suggest that phonotactic, phonetic, or phonological features of the test items or of the familiarization streams were irrelevant to explain infants' looking behavior.

5.4.3 Discussion of Experiments 1-3

Overall, the results of Experiments 1-3 indicate that 12-month-olds possess the resources to grasp morphosyntactic rules. They could discover a structural dependency between distant syllables occurring within words, and generalize it to novel instances. Crucially, they could do so after being exposed to little variation of the middle element and to few exemplars instantiating such rule. Experiments 1 and 2 show that infants consider nonadjacent within-word relations. Experiment 3 shows that lower-level factors extraneous to the pseudo-morphological regularity embedded in the familiarization stream do not play a major role to determine infants' preferences. Previous studies familiarizing infants with arbitrary lists of stimuli containing *between-word* nonadjacent relations (Gómez, 2002; Gómez & Maye, 2005), as well as research on the ability to detect morphological relations in natural speech (Santelmann & Jusczyk, 1998; Mintz, 2004) failed to report a sensitivity to nonadjacent elements before 15/18 months (see also Section 3.2.1.1 on Chapter 3). There are various explanations accounting for the differences among these studies and the findings of Experiments 1-3, which will be discussed in Section 5.10.1 – General discussion.

In Experiments 1-3, infants were exposed to a segmented stream, and they succeeded at finding the morphological-like regularity underlying words. Together with this success, the MOM hypothesis also predicts a failure in generalizing structural information when the speech signal does not contain bracketing cues. More specifically, familiarization to a continuous stream should not suffice for infants to detect within-word relations, because generalization should appear only if the segmentation problem has been solved. Yet, at the same time, the same continuous stream will be the object of statistical computations, and so it should suffice for them to find words inside the stream. The next two experiments test these predictions. An additional group of 12-month-olds was exposed to a stream with the same statistical properties as that of Experiment 1, but played without interruptions between words. Because the two streams use to conduct Experiments 1-3 did not yield any significant differences in looking times (see the ANOVAs conducted by pooling together Experiments 1-3 and Experiments 2-3), in the following two experiments 12-month-olds were exposed only to Stream 1.

5.5 Experiment 4: do 12-month-olds generalize after exposure to a continuous stream?

5.5.1 Method

5.5.1.1 Participants

Sixteen 12-month-olds (9 girls; mean age: 12 months 19 days; age-range: 12 months 3 days – 13 months 6 days) participated in Experiment 4 and were retained for analysis. An additional 7 infants participated but were excluded from the analyses due to the following reasons: because of excessive fussiness (4), because they exceeded the maximum looking time criteria (2), and because of equipment failure (1).

5.5.1.2 Material and Procedure

Procedure, test items and data analysis were identical to Experiment 2, except for the familiarization stream, which was a newly synthesized stream containing the same syllable sequences of Stream 1, but without any pause between words. The stream lasted 2 m 14 s.

5.5.2 Results and discussion

In Experiment 4, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the mean computed for each Test Item Type amounted to 3.3% of the total (9 out of 245 data points). Infants looked equally to both test items ($M_{\text{Non-words}}$ = 8.67s, SE = .69; $M_{\text{Rule-words}}$ = 8.79s, SE = 0.55, F(1,15) = .04, p = .85; Figure 12). Thus, unlike Experiments 1-3, infants failed to extract any structural information after familiarization to a continuous stream. Experiment 4 suggests that, just as adults in Peña et al.'s experiments (see Section 4.1.2.1 on Chapter 4), infants need segmentation indexes to capture within-word generalizations.

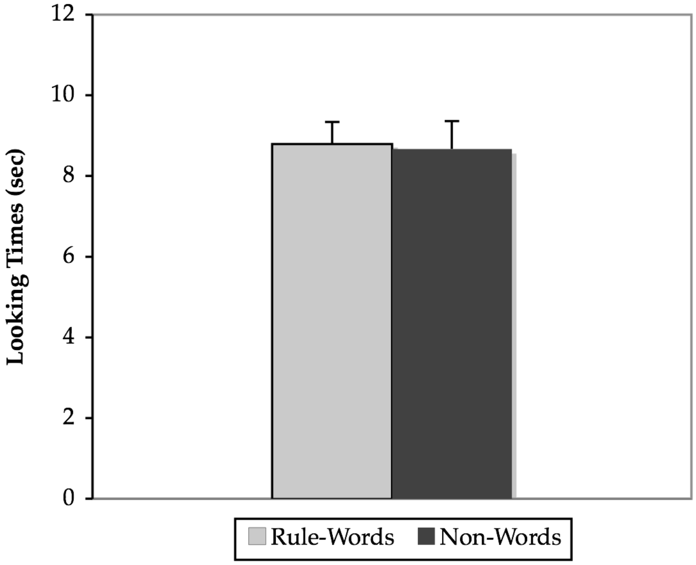


Figure 12: Mean looking time(s) and SE for the test items of 12-month-olds listening to rule-words and non-words after a continuous familiarization stream.

However, it is also possible that infants failed to manifest a differential behavior when listening to the rule-words and non-words during the test phase, not because they could not find rules, but because they could not compute anything from the continuous stream to which they were exposed during familiarization. If this were the case, then after the same familiarization they should also fail to differentiate words from non-words. Experiment 5 tests this alternative explanation.

5.6 Experiment 5: word segmentation after exposure to a continuous stream in 12-month-olds

5.6.1 Method

5.6.1.1 Participants

Sixteen 12-month-olds (8 girls; mean age: 12 months 21 days; age-range: 12 months 2 days – 13 months 13 days) participated in Experiment 5 and were retained for analysis. An additional 8 infants took part but were excluded from analysis because of excessive fussiness.

5.6.1.2 Stimuli and Procedure

Infants were familiarized with the same continuous stream of Experiment 4, but were tested with words and non-words. The non-words were identical to those used in the test phase of Experiment 2 (see Table 1). The experiment and the data analysis were otherwise identical to Experiment 4.

5.6.2 Results and discussion

In Experiment 5, the looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the mean computed for each Test Item Type amounted to 4% of the total (9 out of 226 data points). Infants looked longer when listening to non-words than to words ($M_{\text{Non-words}} = 10.24$s, SE = .83; $M_{\text{Words}} = 8.42$s, SE = .66, $F(1,15) = 5.38$, $p < .03$; Figure 13).

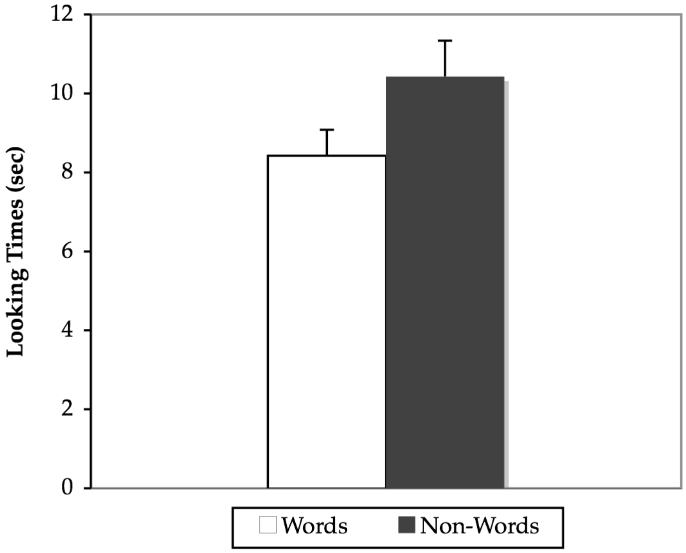


Figure 13: Mean looking time(s) and SE for the test items of 12-month-olds listening to words and non-words after exposure to a continuous familiarization stream.

The result of Experiment 5 suggests that 12-month-olds are sensitive to statistical relations among syllables after familiarization with a continuous stream. This result can be interpreted both as evidence that infants, after being familiarized to a continuous stream, could exploit the within-word perfect nonadjacent TPs of 1 among "words", or that they can exploit a TP difference as low as .5 among the adjacent syllables of the subparts of the test items. According to this second possibility, infants may succeed because words' adjacent TP is equal to .5, whereas in non-words it is 0. Then, they might notice a difference between words and non-words even without attending to nonadjacent TPs. While this alternative account is possible, it assumes that infants can exploit TP differences between subparts of items as low as .5. Currently there is no clear evidence showing that infants can segment elements out of a continuum by exploiting such low TPs, without appealing to other cues.

The design and the material of Experiment 2 cannot make one possibility more plausible than the other. However, in either case, it is crucial to stress that the results of Experiments 1, 2 and 3 suggest that *even if* infants were able to track nonadjacent relations among syllables, they still would not be able to project within-word generalizations *solely* on the basis of such relations. Like the adults tested by Peña et al. (2002), infants could attain within-word generalizations after familiarization with a segmented stream (Experiments 1-3), but could only find words (Experiment 5), and not structural regularities (Experiment 4), after listening to a continuous stream. Overall, the findings are consistent with the hypothesis that generalizations and statistical computations rely on two distinct mechanisms, yielding different linguistic representations prompted by different signal properties. The data also show that already at one year infants can be sensitive to the morphosyntactic properties of words inside a speech stream, but only when words are already segmented out of the stream. At present, there is no clear evidence that infants can segment elements out of a continuum by exploiting nonadjacent TPs, or low adjacent TPs, without appealing to other cues. This point will be taken up and discussed in Sections 5.10.2 and 5.10.3.

The next experiments aim at investigating whether statistical computation and generalization mechanisms are equally available to younger infants, precisely to 7-month-olds. At an age, infants are still solving the segmentation problem in their native language (Jusczyk, 1995); they can compute adjacent TPs in a continuous stream (Saffran, Aslin, & Newport, 1996; Pelucchi et al., 2008), and they possess generalization abilities confined to patterns containing repetitions (Marcus et al., 1999). It is not known whether infants at this age are capable of computing less than perfect TPs between adjacent elements, and whether they can extract structural relations between nonadjacent non-repeating elements. In principle, 7-month-olds should possess the resources to extract within-word structural information from a fluent-like speech stream. In contrast, if sensitivity to within-word structure develops only after infants possess language-specific word learning strategies and start developing a sizeable lexicon for their natural language, delays between the two mechanisms may be expected. Experiments 6 and 7 start addressing these issues by investigating whether 7-month-olds can learn structural regularities and generalize them to novel instances after exposure to segmented streams.

5.7 Experiments 6 and 7: generalizations after exposure to segmented streams in 7-month-olds

5.7.1 Method

5.7.1.1 Participants

Sixteen 7-month-olds (5 girls; mean age: 7 months 22 days; age-range: 7 months 12 days – 8 months 7 days) participated in Experiment 6 and were retained for analysis. An additional 15 infants participated but were excluded from analysis due to the following reasons: excessive fussiness during familiarization or test phase (14), or because exceeding the maximum looking time criteria (1).

5.7.1.2 Material, Procedure and Results

The experiment and the data analyses were identical to Experiment 2 (Stream 1 and rule-words/non-words as test items). In Experiment 6, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the mean computed for each Test Item Type amounted to 6% of the total (15 out of 251 data points). Infants showed no tendency to look differentially at rule-words or non-words ($M_{\text{Non-Words}} = 9.38$, $SE = .64$, $M_{\text{Rule-Words}} = 9.895$, $SE = 0.56$, $F(1,15) = 1.598$, $p = 0.23$).

To ensure that this null result was not due to the small sample size, or to specific factors tied to the material of Experiment 6, another group of 7-month-olds was tested with the material of Experiment 3 (Stream 2). Sixteen infants (7 girls; mean age: 7 months 15 days; age-range: 7 months 5 days – 8 months 9 days) participated to Experiment 7 and were retained for analysis. An additional 13 infants participated to the experiment but were excluded from analysis because of excessive fussiness during familiarization or test phase.

The average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the mean computed for each Test Item Type amounted to 2.9% of the total (7 out of 243 data points). Even in this case, infants showed no tendency to look differentially at rule-words or non-words ($M_{\text{Non-Words}}$= 10.93, SE = .86, $M_{\text{Rule-Words}}$ = 10.01, SE = 0.82, F(1,15) = 2.12, p = 0.17; Figure 14).

A repeated measure ANOVA was conducted by pooling the LTs data of Experiments 6 and 7 together, with the primary aim of determining whether the specific material used in the two experiment affected infants' performance. Stream (2 levels: Stream 1, Stream 2) was a between-participant factor, and Test Item Type (2 levels: Non-words, Rule-words) was a within-participant factor. The ANOVA revealed no main effect of Test Item Type, F(1, 30)= 0.03, p = .87, ns, nor of Stream, F(1, 30)= 0.09, p = .76, *ns*. Nor did their interaction, F(1, 30)= .18, p = .78, ns.

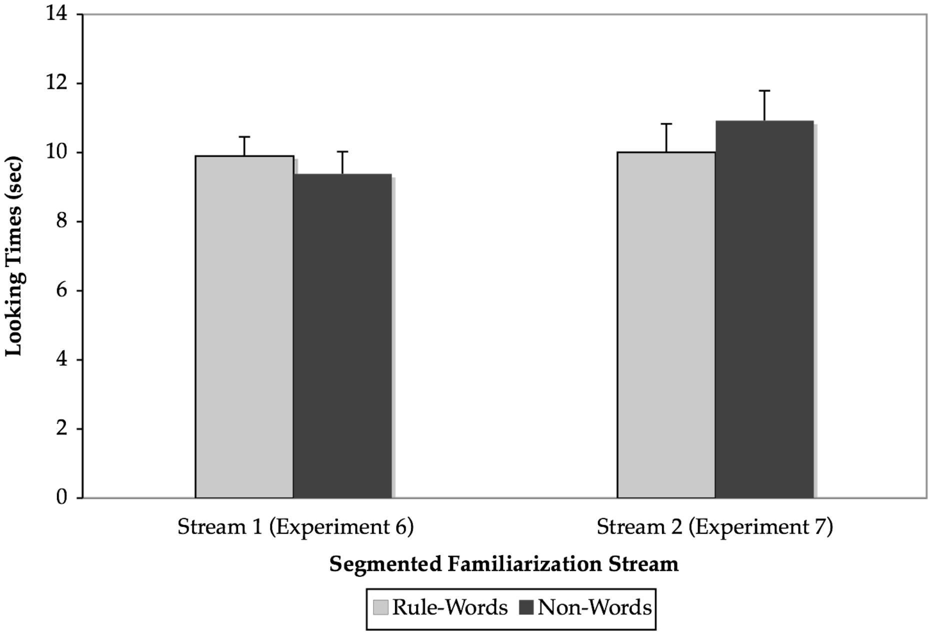


Figure 14: Mean looking time(s) and SE for the test items of 7-month-olds listening to rule-words and non-words after exposure to segmented familiarization streams. The x-axis represents the two familiarizations streams.

5.7.2 Discussion

Taken together, the results of Experiments 6 and 7 suggest that the computational resources to find within-words structural relations are not yet available at 7 months. However, it is also possible that infants did not responded to the procedure and the material used to test them; if so, they would fail to perform any computation whatsoever. To disentangle between these two alternatives, two other experiments were conducted to the purpose of testing whether infants can learn words contained in the segmented streams used to conduct Experiments 6 and 7.

5.8 Experiments 8 and 9: word learning after exposure to segmented streams in 7-month-olds

5.8.1 Method

5.8.1.1 Participants

Sixteen 7-month-olds (11 girls; mean age: 7 months 19 days; age-range: 7 months 6 days – 8 months 2 days) participated in Experiment 8 and were retained for analysis. An additional 5 infants took part but were excluded because of excessive fussiness during familiarization or test phase.

Sixteen 7-month-olds (9 girls; mean age: 7 months 25 days; age-range: 7 months 9 days – 8 months 6 days) participated in Experiment 9 and were retained for analysis. An additional 12 infants took part but were excluded due to the following reasons: excessive fussiness during familiarization or test phase (11), or because they exceeded the maximum looking time criteria (1).

5.8.1.2 Material and Procedure

Experiment 8 was identical to Experiment 6 (thus, infants were exposed to Stream 1), while Experiment 9 was identical to Experiment 7 (thus, infants were exposed to Stream 2) with the crucial difference that, in both experiments, the test items were words and non-words (instead of rule-words and non-words). Procedure and data analysis were otherwise identical to Experiment 6.

5.8.2 Results and Discussion

In Experiment 8, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type, amounted to 3.3% of the total (8 out of 241 data points). Infants looked longer when listening to non-words than to words (M_{Words} = 9.697s, SE = 0.91, $M_{\text{Non-Words}}$ = 11.76s, SE =1.11, $F(1, 15) = 9.21$, $p < 0.01$.

In Experiment 9, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type, amounted to 5% of the total (12 out of 242 data points). Again, infants looked longer when listening to non-words than to words (M_{Words} = 6.75s, SE = 0.54, M_{NonWords} = 8.31s, SE = 0.48, $F(1, 15) = 4.5$, $p < 0.05$; Figure 15).

A repeated measure ANOVA was conducted by pooling the LTs data of Experiments 8 and 9 together, with the primary aim of determining whether the specific material used in the two experiment affected infants' performance. Stream (2 levels: Stream 1, Stream 2) was a between-participant factor, and Test Item Type (2 levels: Non-words, Words) was a within-participant factor. There was a significant effect of Test Item Type (M_{NonWords} = 10.04s, SE = 0.68, M_{Words} = 8.23, SE = 0.57, $F(1, 30) = 13.24$, $p = 0.01$), and of Stream (M_{Stream1} = 10.73, SE = .73, M_{Stream2} = SE = .38, $F(1,30) = 9.83$, $p < .004$), while their interaction did not have a significant effect, $F(1,30) = 0.26$, $p = .62$, *n.s.* These results suggest that infants could learn words and differentiate them from non-words (as indicated by the main effect of Test Item Type); that the two groups differ in their average looking times (as indicated by the main effect of Stream); and that, importantly, the looking pattern is the same across the two experiments, as suggested by the lack of a significant interaction between the two factors.

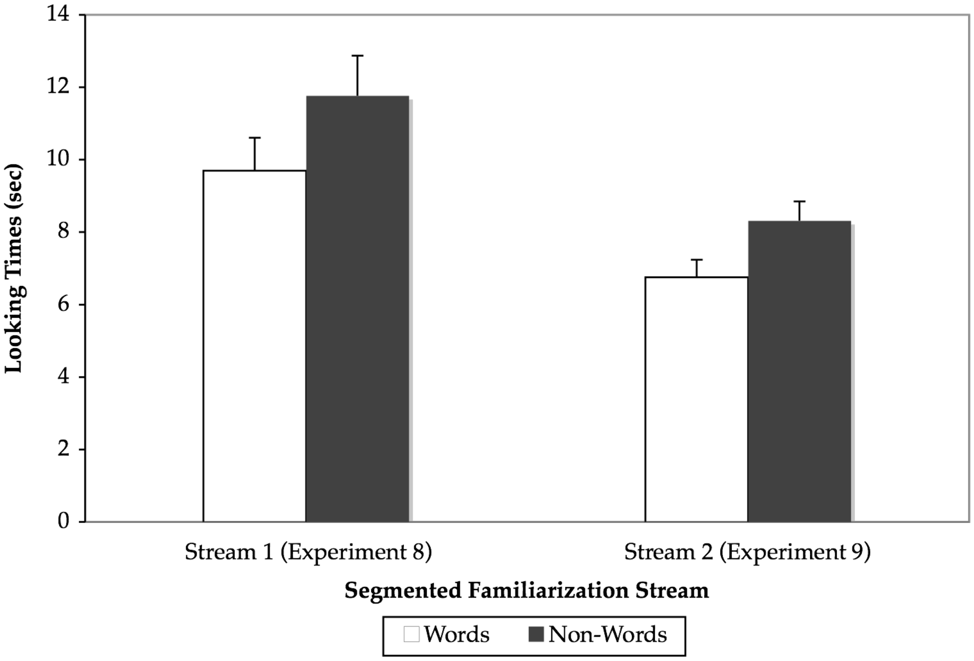


Figure 15: Mean looking time(s) and SE for the test items of 7-month-olds listening to words and non-words after exposure to segmented familiarization streams. The x-axis represents the two familiarizations streams.

Together with Experiments 6 and 7, the results of Experiments 8 and 9 suggest that 7-month-olds do process the information contained in the familiarization streams, finding words in them, but do not possess the resources to generalize within-word relations.

In Experiments 8 and 9 infants could exploit three cues to identify words: adjacent TPs between syllable bigrams, nonadjacent TPs between distant syllable, or segmentation marks between words. Experiments 10 and 11 probes the cue 7-month-olds used to succeed in identifying words in Experiments 8 and 9.

5.9 Experiments 10 and 11: word segmentation after exposure to continuous streams in 7-month-olds

5.9.1 Method

5.9.1.1 Participants

Sixteen infants (7 girls; mean age: 7 months 25 days; age-range: 7 months 11 days – 8 months 13 days) participated in Experiment 10 and were retained for analysis. An additional 8 infants took part but were excluded because of the following reasons: they became too fussy during familiarization or test phase (7), and because they exceeded maximum looking time criteria (1).

Sixteen infants (9 girls; mean age: 7 months 19 days; age-range: 7 months 0 days – 8 months 9 days) participated in Experiment 11 and were retained for analysis. An additional 11 infants participated but were excluded due to the following reasons: excessive fussiness during familiarization or test phase (9), exceeding maximum looking time criteria (1) and experimental error (1).

5.9.1.2 Material and Procedure

Experiments 10 and 11 were identical to Experiments 8 and 9, respectively, with the difference that the familiarization streams had no pauses separating words. Test items and data analyses were otherwise the same as in Experiments 8 and 9.

5.9.2 Results and Discussion

In Experiment 10, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type, amounted to 4% of the total (10 out of 251 data points). Infants showed no tendency to look differentially when listening to words and non-words ($M_{\text{Non-words}} = 10.08$s, SE = 0.79; $M_{\text{Words}} = 10.23$s, SE = .58, F(1, 15) = .03, p = .86, ns).

In Experiment 11, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type, amounted to 4.5% of the total (11 out of 247 data points). Even in this case, infants failed to extract words from a continuous stream ($M_{\text{Non-words}}$ = 10.07s, SE = 0.73; M_{Words} = 9.74s, SE = .68, F(1, 15)= .17, p= .68, ns; Figure 16).

A repeated measure ANOVA was conducted by pooling the LTs data of Experiments 10 and 11 together, with the primary aim of determining whether the specific material used in the two experiments affected infants' performance. Stream (2 levels: Stream 1, Stream 2) was a between-participant factor, and Test Item Type (2 levels: Non-words, Words) was a within-participant factor. The ANOVA revealed no main effect of Test Item Type, F(1, 30)= 0.29, p = .64, ns, nor of Stream, F(1, 30)= 0.78, p = .39, ns., nor of their interaction, F(1, 30)= 2.63, p = .12, ns.

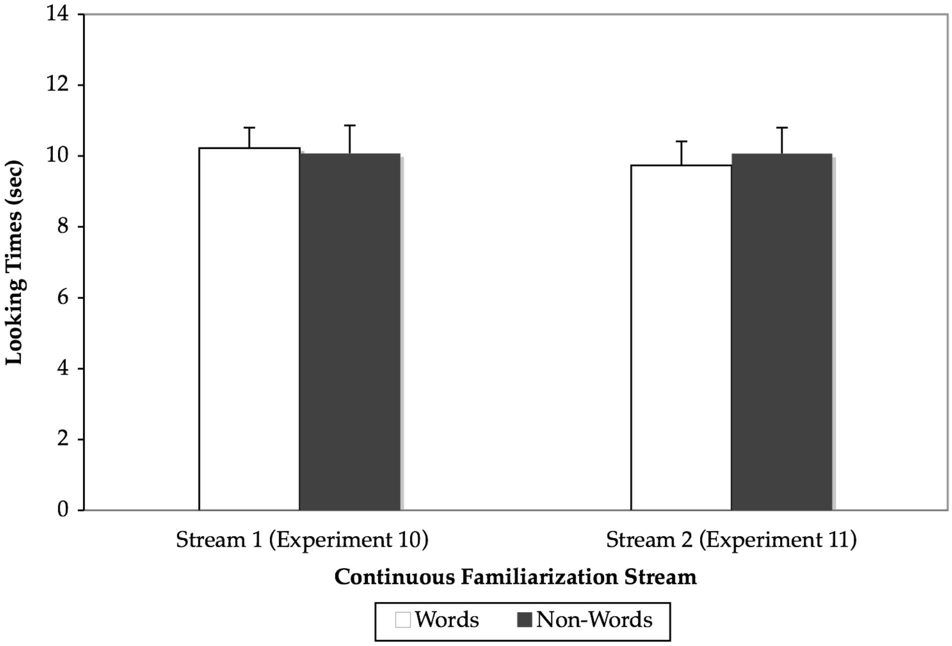


Figure 16: Mean looking time(s) and SE for the test items of 7-month-olds listening to words and non-words after exposure to continuous familiarization streams. The x-axis represents the two familiarizations streams.

Overall, the results indicate that 7-month-olds' success in learning words from the segmented AXC streams (Experiments 8 and 9) may not be due to sensitivity to adjacent and/or nonadjacent statistics, but most likely to the grouping advantage of words over non-words afforded by the segmentation marks in the familiarization. It also suggests that 7-month-olds' abilities at tracking statistics in the signal are probably limited to perfect adjacent relations, as in the study of Saffran et al.'s (1996); otherwise, infants would have been capable of extracting words from the continuous streams used in the present studies, and differentiating them from non-words.

5.10 General discussion

Finding rules and words in a speech stream are tasks that present several layers of complexity for a learner. Besides identifying words as distinct units and rules as law-like relations between them, a learner must master the interaction between these two domains, attested by the widespread presence of morphosyntactic rules in natural languages. Recruiting different mechanisms to identify words and rules, potentially triggered by different signal properties, could help simplifying this daunting task. Indeed, adults switch between different "learning modes", either extracting words or projecting generalizations according to whether segmentation signals are present in a stream (Peña et al., 2002), or to what phonological units carry the information (Bonatti, Peña, Nespor, & Mehler, 2005; Toro, Nespor, Mehler, & Bonatti, 2008).

In a series of eleven experiments, I investigated whether also infants possess a dual-acquisition mechanism, exploiting statistical relations among syllables in a continuous stream to find words, but disregarding the same information while extracting within-word generalizations.

The main results of the experiments can be summarized as follows:

(i) 12-month-olds could generalize word-internal structural properties to novel instances after exposure to segmented streams (Experiments 1-3);

(ii) 12 month-olds could extract statistically defined words contained in a continuous speech stream by computing adjacent and/or nonadjacent TPs, while 7-month-olds failed in this task (Experiment 5 and Experiments 10-11);

(iii) 12-month-olds did not project generalizations over brand-new items after exposure to continuous streams containing the same structural and distributional properties as the segmented streams (Experiment 4);

(iv) Infants at 7 months could learn words after exposure to a segmented stream (Experiments 6 and 7), but could not discover structural regularities on the same basis (Experiments 8 and 9).

I will go back to these issues and comment each of them in the next sections.

5.10.1 Generalization in 12-month-olds: comparison with previous studies

The results of Experiments 1-3 showed that, at the age of 12 months, infants exposed to a segmented stream could learn structural properties among nonadjacent word sub-parts, and generalize them to novel instances. Previous studied failed to find the ability to track relations between distant elements before 15-18 months (Gómez, 2002; Gómez & Maye, 2005; Santelmann & Jusczyk, 1998).

There are several differences that may account for the contrasting evidence. One key difference between the two sets of studies is the kind of learning infants should attain to succeed at the task. Gómez (2002) and Gómez and Maye (2005) asked whether young learners possess the computational resources to track nonadjacent dependency defined as a specific link between particular initial and final elements. In the test phase, items correctly pairing the remote syllables (e.g., A_j - C_i , A_y-C_y) were contrasted with items that maintained the elements in their positions but violated their correct linking (e.g., A_j - C_y). The authors were interested in studying remote dependency learning as they assumed that the same statistical learning mechanism involved in the identification of high TP linguistic patterns (i.e., words) has potentially a role in the acquisition of grammar (Gómez, 2002; see also Section 4.2 on Chapter 4).

In contrast, the primary aim of the present studies was to investigate whether young learners can acquire word-internal structures involving dependencies, defined either over distant syllables, or over *classes* of distant syllables (e.g., the class of syllables appearing in the first position and in the last position). In the test phase, items retaining the first and the last syllables of familiarization words were contrasted with items violating that structure and retaining only one syllable in the original position (either the first, or the second, or the third). Thus, the experimental design was aimed at testing whether infants were capable of projecting structural generalizations to novel instances on the basis of nonadjacent dependencies defined in a "broad" sense (i.e., particular syllable pairings *or* distant relations between syllable classes). The hypothesis central to my work was that the mechanisms responsible for generalization and statistical computation are distinct, they require different signal properties, and they play different roles in language processing – the former accounting for the grammar acquisition, the latter for lexicon acquisition.

To the purpose of investigating structural generalization defined over remote syllables or classes of syllables, a failure in a task like the one used by Gómez (2002) and Gómez and Maye (2005) would remain ambiguous between two interpretations. Infants may fail because they do not extract nonadjacent relations, or else they may fail because they do abstract nonadjacent relations, but on classes of items appearing in first and last positions during familiarization (rather than on specific syllables). Endress and Bonatti (2007) showed that adults extract precisely class-based nonadjacent patterns when exposed to an AXC (subliminally) segmented stream. This finding has important implications for real language acquisition, as in natural languages several grammatical relations are defined between classes of items. For example, morphological processes do not use fixed combinations of specific morphemes (such as affixes and suffixes). The method implemented in the present studies was chosen precisely to avoid such ambiguity in interpretation, and to collect some data on the nature of early word-internal generalizations. Results of Experiments 1-3 showed that the simple overlap of just one common syllable (either the first, or the middle, or the last) between rule-words and non-words does not account for infants' longer looking time to non-words. In contrast, infants treated as familiar those novel items having *both* the same initial and final syllables as familiarization items. Then, the question is to what extent the main conclusions about the 12-month-olds generalization abilities derived from these artificial grammar studies can be scaled to real language situations.

However, also procedural and material differences may explain why infants in the current experiments performed earlier than what the previous literature attests. In most of the previous studies on nonadjacent dependency learning, infants listened to lists of triplets of separate words (the full sentence lasted around 2 s). As a consequence, the target structures to learn are much longer than in the present experiments (words lasted less than 700 ms). Longer stimuli may be harder to process than shorter ones, as they take up working memory and processing resources over a prolonged period time. In contrast, young infants’ working memory and processing space appear to be rather limited (Santelmann & Jusczyk, 1998). Therefore, it is possible that the 12-month-olds’ success at generalizing dependencies between distant word subparts in the current studies may be due to the fact that they can process nonadjacent information over shorter intervals than longer ones, or over word subparts in word-like stimuli than over words in sentence-like strings. Further research is needed to clarify this possibility.

There is a fundamental difference between the study by Gómez (2002) and the current ones. Gómez found that 18-month-olds could acquire nonadjacent dependencies between words of three-elements sentences *only* if they are exposed to linguistic structures containing considerable variability in the middle element (see Section 3.2.1.1 on Chapter 3). In contrast, infants exposed to the AXC stream could discover and generalize structural relations among nonadjacent word subparts without the need of experiencing extensive variability in the middle element (as the variation in the middle element of words was minimal; see Materials on Section 5.2.1.2). The results of Experiments

1-3 do not imply that high variability cannot be successfully exploited to extract information about word or sentence structure. Indeed, the possibility that infants in Gómez (2002) and Gómez and Maye (2005) exploited item frequency to form the representation of an invariant structure is not questioned (and remains plausible). Rather, these results suggest that infants need not collect extensive experience to project morphosyntactic generalizations (based either on pairings between distant syllables or classes of syllables) from segmented words. This finding may support the view that different signal properties are crucial to activate distinct mechanisms – devoted at producing separate outputs.

5.10.2 Sensitivity to TPs in 12- and 7-month-olds

The second result concerns infants' sensitivity to statistics. Many studies documented surprising statistical abilities in very young infants, but their exact extent is still unknown. Even for adults, the literature disagrees as to whether statistical computations can be carried over nonadjacent elements (Perruchet, Tyler, Galland, & Peereman, 2004; Newport & Aslin, 2004; Bonatti, Peña, Nespor, & Mehler, 2006; see also Section 4.2 on Chapter 4). The result of Experiment 5 is compatible with the possibility that 12-month-olds may find words inside a continuous stream both on the basis of high nonadjacent TPs, or on the basis of less than perfect adjacent TPs. In contrast, Experiments 8 and 9 attest that 7-month-olds fail to use statistical information to individuate words contained in continuous streams. The 7-month-olds' failure at segmenting words is compatible with the possibility that they are not capable of computing low adjacent TPs (instead, they are capable of computing perfect adjacent TPs; see Saffran et al., 1996), thus they could not identify disyllables from the continuous stream. In contrast, they may be capable of computing low adjacent TPs from a segmented stream, that is, from a speech input containing bracketing cues, isolating words and facilitating TP computations inside them (Experiments 8 - 9).

Currently there is no clear evidence that infants can segment elements out of a continuum by exploiting nonadjacent TPs, or low adjacent TPs, without appealing to other cues. For instance, in Saffran et al.'s and in Pelucchi et al.'s experiments, words as well as disyllabic subparts within words had adjacent TPs of 1.00. Hence, we do not know whether perfect adjacent TP is a condition for successful isolate words from the speech stream. Using sets of separated word-like items, Gómez and Gerken (1999) suggested that 12- month-olds could indeed extract regularities by exploiting average TP differences in the familiarization strings lower than 1: the authors report the average TPs among the elements of the "legal" strings, ranging between .4 and .6, and among the "illegal" strings ranging between 0 and .2, thus suggesting the possibility that infants can tell very fine TP differences apart. However, the "illegal" strings tested by Gómez and Gerken always mismatched the "legal" strings in familiarization in their first or second items. Thus, infants could have based their responses on simple match/mismatch strategies independently of the TPs among the test items or of their structures. Hence, we cannot draw clear conclusions from these studies.

Overall, the findings of Experiments 5 and 7-8 attest more powerful statistical abilities than previously known in one-year-old children, but they also show that the statistical abilities in younger infants are quite frail. The 7-month-olds failure is compatible both with the possibility that they are not capable of extending to nonadjacent relations (even after familiarization to relations with probability 1.00 and in the simplest test case of words/non-words) or to less than perfect adjacent relations (even in the case of a relatively high TP value, i.e., equal to .50). Just how frail this ability is, and hence how useful statistical computations are for the solution of the real word segmentation problem (Yang, 2004), is a question for further research.

5.10.3 Twelve-month-olds' failure in generalizing after exposure to a continuous stream

The third finding is that 12-month-olds failed to discover structural regularities after exposure to a continuous stream (Experiment 4), while they could do so after being exposed to a stream containing pauses separating words (Experiments 1-3). Importantly, they could extract words from the continuous stream (Experiment 5) on the basis of TP computation.

The first question we may want to ask is why infants could extract words from the continuous stream, while they failed to discover structural information from the same speech input. The results of Experiments 4 and 5 are compatible with two alternative explanations. One possibility is that infants computed nonadjacent relations while listening to the continuous stream, with the difference that they used them to individuate words, but not to generalize within-word structural dependencies. Peña et al. (2002) provided a considerable amount of evidence showing that this is what happens in adult speech processing. In multiple experiments, they contrasted words or rule-words with *part-words*, i.e., sequences that occurred in the familiarization stream having adjacent TPs as high as words but nonadjacent TPs lower than words and rule-words. Using this comparison, Peña et al. could isolate sensitivity to nonadjacent relations, as the test items differed in a continuum in their statistical information.

Instead, the present infant studies contrasted words or rule-words against *non-words*, i.e., sequences that never appeared during familiarization (thus, they had TP of 0 and frequency of occurrence equal to 0). The word/non-word contrast is widely used in several studies investigating infants' sensitivity to distributional information (for example, Saffran et al., 1996; Aslin et al., 1998; Graf Estes et al. 2007), as it requires young learners to perform a simple comparison (that is, between sequences of syllables with TPs of 1.00 against sequences of syllables with TP of 0). Given this comparison, the data of Experiments 4 and 5 are compatible with another explanation. It is possible that, during familiarization, infants simply computed *adjacent* TPs across syllable bigrams, and in the test phase they discriminated words against non-words in virtue of adjacent TP computed across word bigrams (equal to .50 in word bigrams, and to 0 in non-word bigrams). If this were the case, it is possible that 12-month-olds used adjacent TPs in disyllables to identify words, ignoring nonadjacent TPs, thus failing to generalize to novel sequences - given that nonadjacent dependencies constitutes the *only* information to consider in order to discover the analogy between words and rule-words.

5.10.4 Seven-month-olds' failure in generalizing: comparisons with previous studies

The fourth result is that 7-month-old infants do not use segmentation cues to trigger generalizations (Experiments 6 and 7); yet, at this age they succeed in finding some abstract rules (Marcus et al., 1999). Why should they fail to generalize in the current experiments? The nature of the speech input used in the two studies and the mechanisms available at that age may explain the difference. Marcus et al. probed infants' abilities at detecting strict identities between separate items, whereas the present experiments tested more arbitrary relations among nonadjacent syllables. Marcus et al.'s material may have activated an identity detection mechanism, possibly grounded on early available primitive mechanisms (Gervain, Macagno, Cogoi, & Mehler, 2008) and triggering the recognition of abstract relations among items, whether linguistic or not (Saffran, Pollak, Seibel, & Shkolnik, 2007). Instead, the linguistic input of current studies requires infants to detect relations not involving identities, embedded inside words. It is possible that the ability to detect word-internal relations between arbitrary items may require another mechanism, probably more language specific, to which infants at 7 months do not yet have access. Thus, while finding identities may be necessary to learn a language, detecting the morphological construction of lexical items may rely on language specific signal analyzers.

5.10.5 A "perceptual primitive" assisting generalization after exposure to a segmented stream

The striking result that 12-month-olds could generalize after being exposed to the segmented stream, but not to the continuous stream, supports the hypothesis that specific signal properties are required to activate the generalization mechanism. According to the original proposal by Peña et al., silence gaps are needed as they allow skipping the segmentation task and help discovering those structural nonadjacent relations learners would otherwise ignore.

It is possible that the silence pauses between words facilitated the detection of those syllables appearing at the edges of the sequence, and induced learners to notice the invariant relation existing between them. Endress et al. (2005) investigated this possibility and tested adult learning of rule-based regularities. In a series of experiments, they showed that the elements appearing at the initial and final positions are easier to generalize than sequence-internal elements. Building upon their results, Endress et al. proposed the existence of an "*edge-detector*", a perceptual primitive sensitive to auditory units appearing at the edge positions of a sequence. The "edge-detector" has been hypothesized to advantage the encoding of syllables appearing at the edge positions and to make them salient to the purpose of the generalization mechanism (Endress, Scholl, & Mehler, 2005). According to this hypothesis, we may speculate that 12-month-olds familiarized to the segmented stream acquired both words and word-internal structure in virtue of the fact that silence gaps highlighted the edges, and facilitated the learning of the relevant information to consider. In contrast, 12-month-olds computed adjacent and/or nonadjacent TPs to extract words from the continuous stream.

Does the detection of syllable at the edges suffice for acquiring structural generalization? While data from 12-month-olds cannot answer this question, the data from 7-month-olds suggest that edge detection *alone* cannot account both for word learning and generalization. Indeed, 7-month-olds exposed to the segmented stream succeed in learning words (Experiments 8 and 9) but failed to acquire word-internal structures (Experiments 6 and 7). These data are compatible with the hypothesis that silences highlighting word edges may have facilitated the learning of words presented in isolation, but the same cue did not induce the discovery of word-internal structure. This interpretation argues against the possibility that edge detection suffices for structural generalization. Rather, it seems that edge-detection and generalization are distinct computational process, the former representing a tool for the latter. Edge-detection may affect generalization as it may determine what kind of information is extracted from the linguistic input and enter upon the generalization process. Thus, the 7-month-olds' failure in capturing the structural dependency between two distant syllables may indicate that young infants do not possess the resources to generalize structural information (possibly provided by the edge detector). In contrast, they could use the same information to learn words. The result showing that 7-month-olds could learn words when presented with silences between them is not surprising, given that it has already been shown that young infants need not to form a complete or fully accurate representation of words presented in isolation (Brent & Siskind, 2001). Overall, the evidence collected from testing 7-month-olds strengthen the general conclusion that learning word and discovering word-internal structure are distinct tasks, displaying different levels of complexity, and possibly requiring different computational mechanisms.

5.11 General conclusions

The results of the experiments reported in this chapter suggest that infants extract both words and rules from a speech stream, but not in the same way and at the same age. Subtle differences in signal properties may trigger mechanisms to track either statistical distributions of segments or to look for generalizations within words. The generalization process may require specific patterns to operate upon. Moreover, it is possible that the edge-detector primitive may facilitate discovering structural relations between nonadjacent word subparts by highlighting the elements appearing at the edges of the sound sequences. Then, generalizations may arise from the detection of such structural dependencies. Overall, the experiments document how quickly the focus of language learning switches from the identification of unanalyzed sound patterns (as it is presumably the case in 7-month-old infants) to the organization of lexical knowledge at a deeper level of abstraction (as it seems to happen in 12-month-olds).

CHAPTER 6

ON THE NATURE OF GENERALIZATION AND STATISTICAL COMPUTATION MECHANISMS: ARTIFICIAL GRAMMAR STUDIES WITH 12- AND 18-MONTH-OLD INFANTS

In the previous chapter, a series of artificial grammar studies conducted with preverbal infants suggested that morphosyntactic learning might need distinct mechanisms, requiring different signal properties to operate. The results showed that, at one year, infants could generalize within-word regularities, i.e., at an age when they already dispose of language-specific procedures to build a lexicon for their natural language (Werker & Tees, 1984). Importantly, they can find internal-word rules only if the familiarization stream contained segmentation cues. In addition, they can extract words from a continuous stream on the basis of TP between syllables. In contrast, seven-month-olds do not compute either statistics or within-word relations, regardless of input properties. These failures suggest that statistical computation and generalization abilities are quite reduced at 7 months, and that such abilities develop somehow between 7 and 12 months of age.

Building on these findings, the experiments described in this chapter are aimed at further investigating the nature of statistical computation and generalization in infants. The goal of the present studies is to understand *how powerful* statistical computations in infants can be, and to study whether sensitivity to distributional information and sensitivity to grammar-like structures rely on distinct mechanisms. The associative learning theories and the MOM hypothesis would make opposite predictions about infants' performance: the next section is aimed at bringing back the matter of controversy and at presenting the different hypothesis generated by the two contrasting views.

6.1 On the role of statistical computation in grammar acquisition: the predictions of associative learning theories and of the MOM hypothesis

The construction of a lexicon and the acquisition of linguistic structure are two linguistic tasks young infants should perform to acquire language and to use it productively. It has been proposed that infants' sensitivity to some kinds of statistical cues (such as TP among adjacent elements) may help them to construct a lexicon (Chapter 3). Whether or not the same statistics used to extract words from speech are actually relevant to acquire natural language structure is a matter of a long-standing controversy (Chapter 4).

According to associative learning theories, a single computational mechanism can account both for word and rule acquisition. The assumption at the basis of such models is that language acquisition is grounded on mechanisms that are exclusively sensitive to the (conditional) frequency of occurrence of particular syllables (that is, TPs). Such sampling mechanisms should stabilize toward reliable responses only after a considerable exposure to the stream, or, in general terms, they require to collect enough evidence about the distributional properties of a linguistic input (Rumelhart, & McClelland, 1986; Elman et al., 1996; Allen & Seidenberg, 1999; Tomasello, 2000; Perruchet, Tyler, Galland, & Peereman, 2004).

In contrast, the dual-mechanism approach posits that distinct mechanisms are needed to account for how words are segmented and how rules are acquired. Among the authors who advocated this view to explain language acquisition, Peña and collaborators proposed that the speech signal must be analyzed by both statistical computation (for segmenting the speech stream) and by non-statistical computation (for extracting grammar-like regularities; see Section 4.1.2 on Chapter 4). While a sampling mechanism, sensitive to distributional properties, collects information about the sequences occurring in the linguistic input, another non-statistical (deductive) mechanism, operating on a corpus of discrete elements, extracts and projects generalizations without the need of extensive experience with the speech input. As such, statistical computation and generalizations are supposed to gather information about different linguistic properties, and would generate different representations.

Several studies supported the possibility that language users perform both statistical and non-statistical computations to acquire word and rules, respectively (Peña et al., 2002; Endress & Bonatti, 2007; Bonatti et al., 2005, Toro et al., 2008). Adults have been shown to slowly break a continuous stream into words by using nonadjacent TPs; both the associative learning and the dual-mechanism models would predict this outcome. However, in contrast with what is hypothesized by the associative learning models, adults fail to generalize structural regularities, despite the fact that the same nonadjacent computation may have sufficed for this task as well. Instead, they quickly project structural dependencies after exposure to a segmented stream without relying on TP relations, even on the basis of short familiarizations. Thus, they do not seem to project such regularities on the basis of distributional cues. Further investigations established that, rather than helping, longer exposure hinders the detection of generalizations (Endress & Bonatti, 2007).

Do statistical computations and generalization mechanisms have the same limits and selectivities in developing infants? The studies reported in chapter 5 already showed that 12-month-olds could generalize word-internal regularities after exposure to a segmented stream (Experiments 1-3), but not from a continuous one (Experiment 4) -- from which they can still extract statistically defined words (Experiment 5). The aim of the present studies is to provide complementary evidence to these findings by *directly* contrasting sensitivity to statistics against sensitivity to word structure. The next section describes the experimental paradigm I used to isolate sensitivity to distributional information from sensitivity to grammar-like structure.

6.2 Pitting sensitivity to statistics against sensitivity to structure in infants

6.2.1 Isolating sensitivity to statistics against sensitivity to word structure: a new test contrast

Experiments 1-5 already provided evidence that 12-month-olds possess and recruit distinct mechanisms, one to generalize structural dependencies between distant syllables of sound sequences (crucially, only after exposure to a segmented stream), the other one to extract words from a continuous stream. However, in these studies sensitivity to statistics was not directly pitted against sensitivity to structure within the same experimental condition; instead, the two sensitivities were established across different experiments. The reason lies in the test contrast used in these studies. During the test phase, infants were asked to discriminate either words against non-words (to assess the identification of words) or rule-words against non-words (to investigate structural generalizations). Non-words assemble the syllables of words into a novel order, they are statistically uninformative as they have TPs of 0 and frequency of 0. Therefore, in the rule-word/non-word contrast, sensitivity to structure is *not* pitted against sensitivity to statistics.

One way to clarify whether infants selectively activate distinct mechanisms when processing linguistic material would be to contrast rule-words against *part-words* (as Peña et al. did with adults). This comparison is more subtle than the rule-word/non-word one, as it asks to discriminate novel sequences -- having nonadjacent TPs as words -- from sequences that occurred in the familiarization stream -- but spanned the word boundaries (thus, they have lower TPs than words). This contrast may be difficult for infants to achieve, as it requires them to consolidate their representation of sequences *straddling* the boundaries of words. We currently do not know whether young infants can learn part-words as they do with words. All the studies investigating statistical computation (for example, Saffran et al., 1996; Aslin et al., 1998; Graf Estes et al. 2007) used the word/part-word to contrast statistical sensitivity varying along a continuum, and found that infants could learn words but not part-words. While the authors of these studies concluded that infants possess fine-grained statistical sensitivity, still this result is compatible with another explanation. It is possible that, instead of differentiating words from part-words on the basis of the relative difference in their TP values, infants discriminated them simply because they could not compute TPs over part-words (if so, part-words would appear to infants as uninformative as non-words).

In reason of this uncertainty, the choice of contrasting rule-words against non-words in Experiments 1-12 was conservative, while at the same time it allowed investigating the conditions required to project structural generalizations (even if they were established across several experiments). Building on the results collected in these preliminary studies, the next experiments will move a step further. They will use the *part-word* against *rule-word* contrast, to the purpose of pitting sensitivity to statistics against sensitivity to grammatical-like structure *within* the same experiment. This methodological choice will allow investigating whether statistical computations can also determine the acquisition of linguistic structure.

6.2.2 Experimental hypotheses

The MOM hypothesis makes specific predictions about the kinds of patterns infants are expected to learn from different linguistic inputs. First, infants familiarized to a continuous stream should be able compute statistics to the purpose of identifying linguistic patterns contained in it (that is, words as well as part-words), and, at the same time, they should not use the same information to generalize structural dependencies to novel instance (that is, rule-words).

Second, infants should generalize structural dependencies after exposure to a segmented stream, but should not extract statistically occurring sequences from it (even if the stream has the same distributional properties of the continuous one).

It is important to note that using the part-word/rule-word contrast allows understanding the relative power of statistical learning mechanisms and of generalization mechanisms in determining what kind of linguistic knowledge infants attain after exposure to different kinds of speech input. Hence, a preference for one kind of test items would indicate a stronger (or selective) activation of one mechanism to solve a specific linguistic task. Moreover, the lack of any preference may also be informative, as it may indicate either that the two mechanisms are equally activated (thus, the null difference is determined by competing preferences for both kinds of test items), or that both mechanisms cannot be efficiently recruited to process the linguistic input. The methodological advantage of using this test contrast is also evident when studying the different sensitivities and limitations of two mechanisms during development. To the purpose of investigating this second issue, the present studies have been conducted with infants at 12 and 18 months.

These two ages were selected because they correspond to the achievements of important milestones in real language acquisition. It is known that, at 12 months, infants are already engaged into the problem of constructing the lexicon of their native language (Bates & Goodman, 1999), and (probably) they started attaining knowledge about the word compositional structure (as suggested by the results of Experiments 1-3). Little is known about the power of the computational abilities possessed by infants at this age, as the few studies investigating the computation of less than perfect TPs suffer of methodological inadequacies (see Section 5.10.2 and 5.10.3 for a discussion of this point). At 18 months, infants have been shown to be sensitive to morphosyntactic dependencies between words presented in a sentence (Santelmann & Jusczyk, 1998); they can reliably track relations among nonadjacent elements separated by silences, but only when the distant elements appear with a much higher frequency than the intervening element of the triplet (Gómez, 2002). Overall, these findings suggest that infants at 18 months already possess the resources to start acquiring their native language grammar.

6.2.3 Plan of the studies

A series of seven experiments, conducted using a modified version of the head-turn preference procedure (see Section 5.2.1.3 on Chapter 5), have been carried on to test the two aforementioned predictions, concerning the different nature and sensitivities of statistical computation and generalization mechanisms in 12- and 18-month-old infants.

Experiments 12, 13 and 14 investigate whether 18-month-olds can compute distributional information and acquire structural dependencies on a different basis. Experiment 12 asks whether infants can extract statistically occurring sequences by computing TPs and frequency distributions, but cannot generalize on the basis on the same information. Experiment 13 tests the opposite prediction, namely whether infants can generalize, but not compute statistics, after exposure to a segmented familiarization. Experiment 14 controls for the possibility that, instead of generalizing to word structure, 18-month-olds simply responded to phonological differences between items they listened to during familiarization and items they heard during the test phase.

Experiments 15, 16 and 17 test the same predictions with 12-month-old infants. Experiment 15 asks whether they can detect statistical information and learn occurring sequences from continuous stream, while they do not use the same information to acquire structural regularities. Experiment 16 investigates whether a prolonged exposure to the continuous stream would help 12-month-olds to compute the statistical information contained in it. Experiment 17 asks whether infants can generalize after exposure to a segmented stream and, at the same time, they do not detect statistically defined sequences occurring in it.

6.3 Experiment 12: sensitivity to statistics against sensitivity to structure after exposure to a continuous stream in 18-month-olds

6.3.1 Method

6.3.1.1 Participants

Sixteen 18-month-old infants (12 girls; mean age: 18 months 20 days; age-range: 18 months 3 days - 19 months 14 days) participated in Experiment 12 and were retained for analysis. An additional 22 infants took part but were excluded from analysis due to the following reasons: because of excessive fussiness during familiarization or during test phase (20) or because they exceeded maximum looking time criteria looking longer than 65 cumulative sec in more than one test trial (2).

6.3.1.2 Material and Procedure

Participants were tested with the same modified head-turn procedure adopted to run the other infant studies, using the same experimental design, procedure and apparatus (see Section 5.2.1.3 on Chapter 5). A new familiarization stream and novel test items were created to conduct Experiment 12; the visual stimuli used to attract infants' gaze was identical to the one used in the previous experiments.

6.3.1.2.1 Familiarization stream

The familiarization stream consisted of a pseudo-random concatenation of four AXC words (see also Section 5.2.1.2.1 on Chapter 5). The words used were: /**ba**mu**so**/, /**ba**ga**so**/, /**li**mu**fe**/ and /**li**ga**fe**/ (boldface font indicates the syllables of the A-C frames). The stream was constructed in a way it had different properties than the AXC streams of the previous infant studies, which simplified the Peña et al.'s streams but respected the same constraints (see Section 5.2.1.2.1 on Chapter 5). Specifically, the statistical properties of the stream were modified to the purpose of maximizing the amount of distributional information provided in a short time of exposure, thus offering a reliable basis to the extraction of statistically occurring patterns (even for those sequence spanning word boundaries, as part-words will be used in the test contrast). Allowing the identification of statistically occurring sequences was of primary importance in order to pit sensitivity to statistical information against sensitivity to structural information within the same experiment.

Before explaining how the new stream was created, the description of the statistical properties of the AXC streams used to conduct Experiments 1-12 will follow. In these streams, all words occurred with the same frequency and were concatenated in such a way that TPs across their boundaries would maintain equal the probability of a last syllable of a word to be followed by any of the two first syllables. For instance, the last syllable /so/ could be either followed by the first syllable /ba/ (such as in the sequence …*bamu**soba**gaso*…) with TP of .50, or by the first syllable /li/ (such as in the sequence …*bamu**soli**gafe*…) with TP of .50. A consequence of this design was that part-words had adjacent TPs of .50 (thus, equal to adjacent TPs in words) and nonadjacent TP of .50 (words had TP of 1.00), and they occurred with half frequency than words.

There is at least one serious reason to believe that part-words with low TP and occurring half frequently than words may be difficult for infants to learn: extracting part-words is a difficult task for adults. Indeed, massive experience with the linguistic input is require to them in order to exploit the distributional properties, and to consolidate the memory traces of part-words (Peña et al., 2002; Endress & Bonatti, 2007)[10]. The strategy of prolonging the exposure to the familiarization stream cannot be used with infants or young children, because have limited attentional resources, and they quickly lose their interest when a stimulation, repeating a limited set of items, is played for a long time period.

In order to investigate statistical sensitivity in infants, it was important to provide an optimal basis to the computation of TPs between syllables spanning the boundaries of words. Then, a new AXC stream was created with the following features:

(i) It contained four words, two of which appeared *twice* frequently than the other two. This manipulation implied higher adjacent TPs within frequent words than within infrequent words;

(ii) The two frequent words (namely, /bagaso/ and /limufe/) occurred in some parts of the stream subsequently one after the other, i.e., they alternated (without permitting immediate word repetition), such as in this sample: *…bagaso limufe bagaso…* ;

(iii) The concatenation of frequent words as described in (ii) determined high adjacent TPs at their boundaries.

As a result of the features described above, those sequences spanning the boundaries of frequent words (henceforth, "high-TP part-words") had adjacent and non-adjacent TPs equal to .67. These TP values were close to that of frequent words (having adjacent TP of .67 and non-adjacent TPs of 1). Such TP values are relatively high and, in principle, they may suffice for the extraction of high-TP part-words. In contrast, frequent words had adjacent TPs equal to .33 and

---

[10] When asking whether adults would benefit from an increased exposure to the familiarization stream to consolidate the memory representations of items appearing in the stream but spanning the word boundaries (i.e., part-words), Peña and collaborators found that memory traces solidify after a long exposure, i.e., of 30 minute. Similarly, Endress and Bonatti (2007) asked whether an extensive experience with the speech input would strengthen the representations of part-words and inhibit the generalization of structural dependencies; they found a preference for part-words after a familiarization stream of 1 hour, but not with shorter exposures.

nonadjacent TP equal to 1.00, that is, lower adjacent TPs but higher nonadjacent TPs than high-TP part-words. High-TP part-words occurred 52 times during familiarization; frequent words occurred 64 times and infrequent words 32 times.

6.3.1.2.2 Test items

The items used in the test phase were rule-words and high-TP part-words. The four rule-words were constructed with the same criteria as in the previous studies (see Section 5.2.1.2.2 on Chapter 2): they had the first and the last syllables of words, and the middle syllable was a syllable that never appeared in that position during familiarization (it was either the initial or the last syllable the other A-C frame). The list of the four high TP part-words and the four rule-words is reported in Table 3.

Familiarization	**Test Items**	
Words	**Rule-Words**	**High TP- Part-Words**
/**ba**mu**so**/	/**ba**li**so**/	/gasoli/
/**ba**ga**so**/	/**ba**fe**so**/	/mufeba/
/**li**mu**fe**/	/**li**ba**fe**/	/solimu/
/**li**ga**fe**/	/**li**so**fe**/	/febaga/

Table 3: Words used to compose the familiarization streams, and the items used in the test phase of Experiments 12-17. Boldface font indicates the syllables that define the A-C frames, thus it highlights the structural similarity of words and rule-words.

The auditory material was synthesized with the Mbrola speech synthetizer (Dutoit, et al., 1996), using the FR-2 diphone database, setting flat prosody with 116 ms phoneme length and 200 Hz pitch. In order to avoid direct cues to word onsets, the familiarization stream was synthesized with increasing and decreasing amplitude ramps in the first and last 5 s, respectively. The familiarization stream lasted 2 m 13 s.

6.3.2 Results and discussion

Looking times were averaged across all trials of the same Test Item Type (Rule-word, High-TP Part-word). Average looking times shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type were excluded from further analysis. With these criteria, the overall excluded data amounted to 2.9% of the total (7 out of 246 LTs data points).

A repeated measure ANOVA with Test Item Type (2 levels: Rule-word, High-TP Part-word) as within-participant factor was conducted, using mean looking times as dependent measure. Infants looked longer when listening to rule-words than to high-TP part-words ($M_{\text{Rule-words}} = 14.91$, SE = 1.14; $M_{\text{High-TP Part-Words}} = 11.92$, SE = .98, $F(1,15) = 5.26$, $p < .03$; Figure 17).

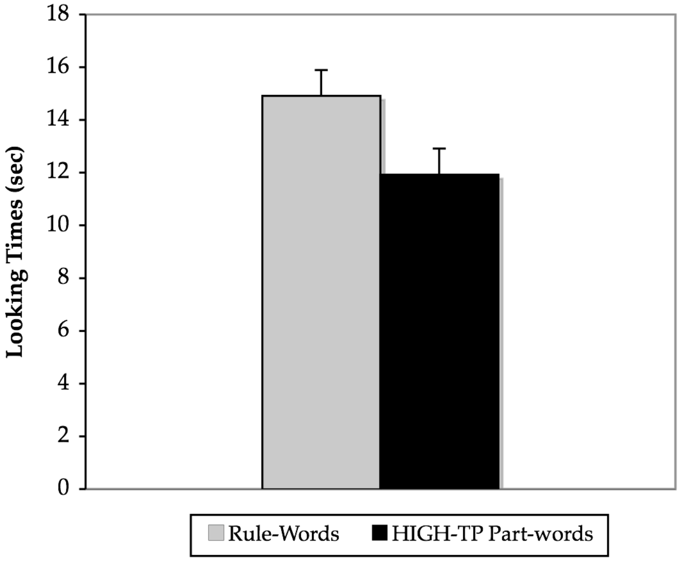


Figure 17: Mean looking time(s) and SE for the test items of 18-month-olds listening to rule-words and high-TP part-words after exposure to a continuous familiarization stream.

This result suggests that infants could detect those sequences straddling the boundaries of frequent words, and could capitalize on high (but smaller than 1.00) TPs and high frequency of co-occurrence to extract them. Longer looking times for rule-words than for high-TP part-words may indicate that infants failed to generalize to word structure, and, when listening to rule-words, they responded with a novelty preference.

However, before concluding that 18-month-olds could not project structural regularities after exposure to a continuous stream, it is necessary to establish whether they are capable of generalizing to rule-words after exposure to a segmented stream. If they recognize rule-words as familiar (as they comply with the structure of words) they should look longer when listening to high-TP part-words, because they violate the word structure. In contrast, if they cannot generalize to rule-words, their looking times when listening to the two kinds of test items should not differ (indicating that they could neither generalize nor compute statistics), or should be longer for rule-words (indicating that infants extracted part-words, recognized them, while rule-words would appear unfamiliar). Experiment 13 tests this possibility by exposing 18-month-olds to a segmented stream, maintaining exactly the same statistical properties as the stream of Experiment 12.

6.4 Experiment 13: sensitivity to statistics against sensitivity to structure after exposure to a segmented stream in 18-month-olds

6.3.1 Method

6.3.1.1 Participants

Sixteen 18-month-old infants (7 girls; mean age: 18 months 26 days; age-range: 18 months 7 days - 19 months 10 days) participated in Experiment 13 and were retained for analysis. An additional 22 infants participated but were excluded due to the following reasons: because of excessive fussiness during familiarization or during test phase (20) or because they exceeded maximum looking time criteria (2).

6.3.1.2 Material and Procedure

Infants were familiarized with the stream of Experiment 12, except that there were 200 ms pauses separating words; the stream lasted 2 min 52 s. Test items and analysis performed on looking times were otherwise identical to Experiment 12.

6.3.2 Results and discussion

In Experiment 13, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type amounted to 3.2% (8 out of 254 data points). The repeated measure ANOVA revealed that infants looked significantly longer while listening to high-TP part-words than to rule-words ($M_{\text{High-TP Part-words}}$= 11.67 sec, SE = 1.13; $M_{\text{Rule-words}}$=9.65, SE = 0.77, F(1, 15) = 5.37, $p < 0.03$; Figure 18).

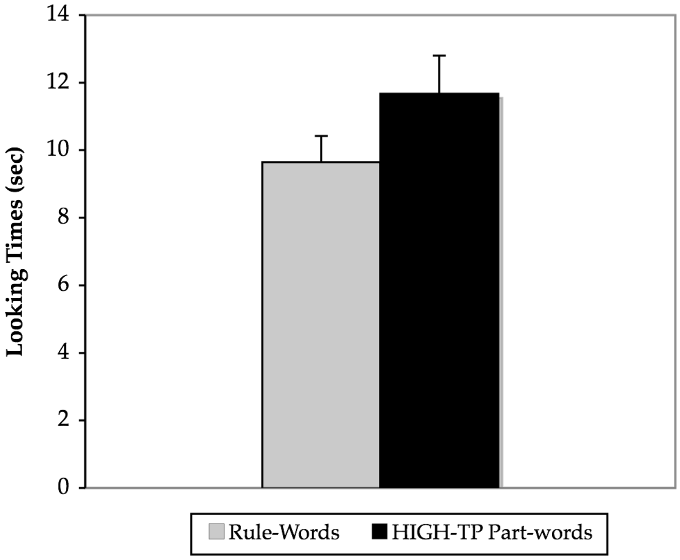


Figure 18: Mean looking time(s) and SE for the test items of 18-month-olds listening to rule-words and high-TP part-words after exposure to a segmented familiarization stream.

The looking time pattern of this experiment is reversed if compared to that of Experiment 12 (where infants were familiarized to a continuous stream). Since rule-words never appeared during familiarization, while high-TP part-words did, this result indicates that infants exposed to the segmented stream could grasp the structural dependencies between the first and the last syllables of words and generalized them to rule-words. At the same time, the distributional properties of the linguistic input, if computed, were not sufficient to override the preference for structurally legal items.

However, an alternative explanation might account for the result of Experiment 13. It is possible that 18-month-olds might have computed probabilities over syllables and pauses ---thus, high-TP part-words are perceived as sequences of syllables and pauses.

If so, infants might have expected a pause always to appear between the last syllable of a word and the first syllable of the next one. In contrast, high-TP part-words used in the test phase did not contain the pause. As a possible consequence, longer looking times for high-TP part-words may be due not to the fact that infants activated a mechanism suited for generalization, but to the fact that they detected the absence of pauses in the test items, and they reacted to the mismatch of high-TP part-words during familiarization and test phase. Experiment 14 rules out this possibility.

6.5 Experiment 14: controlling for the effect of acoustic differences between familiarization and test sequences

6.5.1 Method

6.5.1.1 Participants

Sixteen 18-month-old infants (7 girls; mean age: 18 months 14 days; age-range: 18 months 0 days - 19 months 1 day) participated in Experiment 14 and were retained for analysis. An additional 18 infants participated but were excluded due to the following reasons: because of excessive fussiness during familiarization or during test phase (16) or because they exceeded the maximum looking time criteria (2).

6.5.1.2 Stimuli and Procedure

Infants were familiarized with the segmented stream as in Experiment 13. However, the high-TP part-words items used in the test phase included the silence as it appeared in the stream, i.e., between the last and the first syllable of words. For example, the test item of Experiment 13 /gasoli/, straddling the boundaries of the two words /bagaso#limufe/ (where # indicates the 200ms pause occurring between them) became in Experiment 14 /gaso#li/. The experiment and the analysis performed on looking times were otherwise identical to Experiment 13.

6.5.2 Results and discussion

In Experiment 14, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type amounted to 3.3% of the total (8 out of 248 LTs data points). The repeated measure ANOVA revealed that infants looked significantly longer while listening to high-TP part-words including the silence gap than to rule-words ($M_{\text{High-TP Part-words}} = 10.65$, SE = .76; $M_{\text{Rule-words}} = 8.81$, SE = 74, $F(1,15) = 5.46$, $p < .04$; Figure 19).

This result suggests that infants' looking behavior is not compatible with the hypothesis that 18-month-olds detected the absence of the silences in the test items (with respect to the input they were familiarized to) and that they simply reacted only to this acoustic/statistical violation during test. Rather, the result supports the hypothesis 18-month-olds could generalize structural regularities to rule-words after exposure a segmented stream.

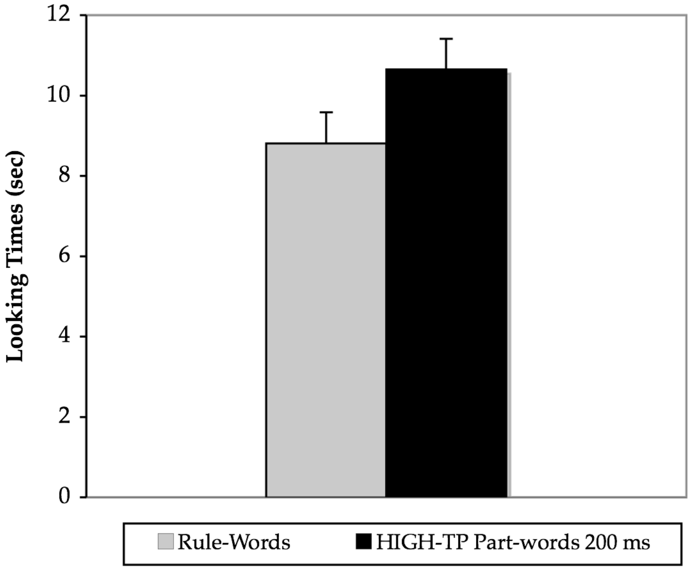


Figure 19: Mean looking time(s) and SE for the test items of 18-month-olds listening to rule-words and high-TP part-words after exposure to a segmented familiarization stream. In Experiment 14, the high-TP part-words sequences used at test included the 200ms silence, as it was during familiarization.

Longer looking times when listening to high-TP part-words, then, indicate either that infants did not compute statistics from the segmented stream, or that the distributional information they computed was not sufficient to override the projection of structural generalizations. To notice that high-TP part-words including the pause are even favored with respect to rule-words in terms of TPs: indeed, *every* pause appearing within the sequence is predicted with adjacent TP of 1--- while rule-words had adjacent TP of 0. Thus, the result of Experiment 14 suggests that infants could generalize by activating a mechanism sensitive to structural properties, rather than to statistics. Whether generalization is favored over statistics, or whether statistics are just not computed from a segmented stream, are questions for further research.

Overall, the results of Experiments 12-14 are compatible with the predictions of the MOM hypothesis. First, generalizations arise after exposure to a segmented stream, and apparently distributional information are not exploited to learn structural regularities. Second, statistical computations allow extracting sequences occurring in the continuous stream but straddling the word boundaries (at least when these sequences occur with high frequency and have high adjacent and nonadjacent TP, as in the case of high-TP part-words). At the same time, 18-month-olds could not project regularities to rule-words on the basis of TP and frequency distributions after exposure to the continuous stream. Opposite looking time patterns are compatible with the hypothesis that infants recruited different mechanisms. Instead, associative learning models would predict the same performance across the two conditions: 18-month-old should have computed distributional information both from the continuous and the segmented stream.

At 18 months, infants are known to be capable of learning linguistic patterns requiring them to form relationships between nonadjacent elements (Gómez, 2002). In contrast, 12-month-old infants fail at this task, unless they are provided with prior experience with the same relationships but presented as adjacent (Lany & Gómez, 2008); they can also compute less than perfect adjacent TPs (Experiment 5 of this thesis; Gómez & Gerken, 1999; Saffran & Wilson, 2003). Moreover, they can learn structural regularities appearing as relations between nonadjacent word sub-parts (Experiments 1-3 of this thesis). However, we are overall still very far from having a complete understanding of young infants' sensitivity to distributional information, of its restrictions and its contribution to the acquisition of grammar-like structures at different stages of development. Experiments 15-17 will address these issues by pitting sensitivity to statistics against sensitivity to structure across different conditions (i.e., after exposure to continuous or segmented familiarization streams), aiming to understand whether 12-month-old infants would project structural information on the basis of statistical computations, or else if they such computations cannot account for the acquisition of linguistic regularities.

6.6 Experiment 15: sensitivity to statistics against sensitivity to structure after exposure to a continuous stream in 12-month-olds

6.6.1 Method

6.6.1.1 Participants

Sixteen 12-month-olds (8 girls; mean age: 12 months 20 days; age-range: 12 months 4 days - 13 months 0 day) took part in Experiment 15 and were retained for analysis. An additional 22 infants participated but were excluded due to the following reasons: because of excessive fussiness during familiarization or during test phase (18), or because they exceeded the maximum looking time criteria (4).

6.6.1.2 Material and procedure

Stimuli, procedure and data analyses were identical to Experiment 12, that is, infants were exposed to the continuous stream and tested with rule-words and high-TP part-words.

6.6.2 Results and Discussion

The average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type amounted to 3.7% of the total (9 out of 248 LTs data points). There was no effect of Test Item Type ($M_{\text{High-TP Part-words}}$ = 8.85 sec, SE = .82; $M_{\text{Rule-words}}$= 7.74, SE = 0.55;, F(1, 15) = 1.76, p = .20, ns; Figure 20).

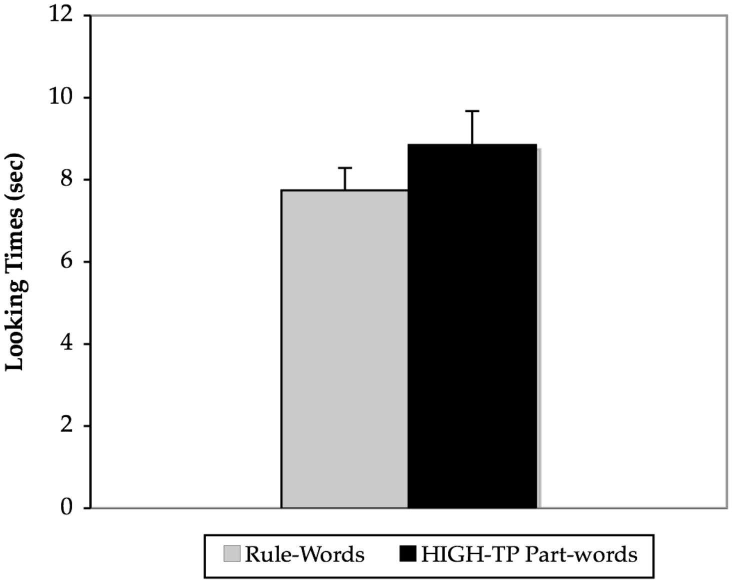


Figure 20: Mean looking time(s) and SE for the test items of 12-month-olds listening to rule-words and high-TP part-words after exposure to a continuous familiarization stream.

Two alternative explanations may account for 12-month-olds' failure. One possibility is that infants were not capable of computing statistics from the speech stream employed in this set of studies. This explanation is plausible but unexpected given the results of Experiment 5, showing that infants of the same age could isolate from continuous speech words having adjacent TP of .50 and nonadjacent TP of 1.00. Another possibility is that, rather than being unable to compute low TP values, 12-month-olds require a longer experience with the stream to attain a reliable representation of sequences occurring in it. It is known that infants' preferences for novelty and familiarity can vary as functions of duration of familiarization and task difficulty related to the age and experience of the infant (Hunter & Ames, 1988). This difficulty may influence the rate at which the stimuli can be processed and discriminated. Computing the distributional properties of the stream used in the current studies may be, in a way, a difficult task for younger infants. Due to its construction, the stream contained "chunks" of frequent words occurring subsequently, and generating at their boundaries the so-called high-TP part-words. As frequent words run one after the other, without any salient drop in adjacent TPs *between* them (this is the reason why high-TP part-words have high frequency and high TP), it is possible that their concatenation is, in a sense, unparseable. Tracking long sequences of may be a nontrivial task for 12-month-old infants, as they have limited processing resources (Santelmann & Jusczyk, 1998).

According this possibility, a failure in discriminating high-TP part-words may indicate that infants are unable to track distributional information across long sequences of syllables (because they have limited processing capacities), or else that infants simply require an extensive experience with the linguistic input to reliably compute its distributional properties and to familiarize with high-TP part-words. Experiment 16 tests this second possibility by increasing the continuous stream duration of 50%.

6.7 Experiment 16: increasing the exposure to the continuous familiarization stream

6.7.1 Method

6.7.1.1 Participants

Sixteen 12-month-old infants participated (9 girls; mean age: 12 months 21 days; age-range: 12 months 2 days-13 months 8 days) participated in Experiment 16 and were retained for analysis. An additional 19 infants participated but were excluded due to the following reasons: because of excessive fussiness during familiarization or during test phase (14), or because they exceeded maximum looking time criteria (5).

6.7.1.2 Material and procedure

Experiment 16 was identical to Experiment 15, with the only difference that the familiarization duration was increased and lasted 3 min 20 s.

6.7.2 Results and Discussion

The average looking times data excluded they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type amounted to 4.4% (11 out of 251 data points). There was no significant effect of Test Item Type ($M_{\text{High-TP Part-words}} = 9.57$ sec, SE = .56 ; $M_{\text{Rule-words}} = 9.27$, SE = 0.49 , $F(1, 15) = .13$, $p = .72$, ns; Figure 21).

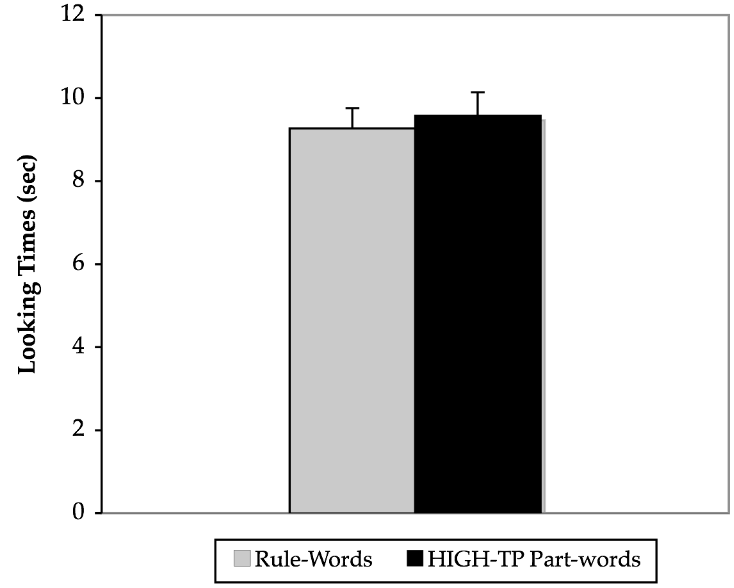


Figure 21: Mean looking time(s) and SE for the test items of 12-month-olds listening to rule-words and high-TP part-words after exposure to a continuous familiarization stream, increased in duration with a factor of .50.

Even after a prolonged exposure to the familiarization stream, 12-month-olds could not discriminate high-TP part-words from rule-words, although they actually appeared in the stream, while rule-words did not. The result suggest that computing the distributional properties of the stream containing long sequences having high TPs and high frequency poses serious difficulties to younger learners. The 12-month-olds' failure is compatible with two hypotheses: infants might have failed because they were not capable of computing statistics from a stream with these specific features, or else because that they would require an even longer familiarization period in order to achieve a successful performance (perhaps the exposure to the stream should be spread in multiple sessions). These possibilities are questions for further research.

As we cannot currently disentangle between these alternatives, two explanations may account for the results of Experiment 16. If 12-month-olds were not capable of computing distributional information from the continuous stream, then the absence of any difference in their looking times may indicate that they could neither compute statistics, nor generalize structural information (thus, both kinds of test items appeared as novel). If, instead, they were capable of performing computations on such input, but they just require a prolonged exposure to familiarize with the sequences occurring in the stream, then the absence of any difference in their looking times may indicate that they are still habituating to high-TP part-words. If so, their attention to the two kinds of test sequences would be equal because of the waning attraction of the familiar stimuli combined with the increasing attraction of the novel stimuli (i.e., rule-words).

The results of Experiments 12-14 with 18-month-olds suggest that statistical computation and generalization are supported by distinct mechanisms, which are selectively activated by different input properties (that is, the presence or the absence of segmentation cues). Experiments 1-3 (on Chapter 5) already established that 12-month-olds could learn structural dependencies after exposure to a segmented stream. Then, if generalization and statistical computations depend on separate mechanisms also in 12-month-olds, a dissociation between the two is expected: one-year-old infants should be able to generalize after exposure to a segmented stream having the same distributional properties of the continuous one -- from which they could not differentiate statistically occurring patterns from structural regularities. Experiment 17 investigates this possibility.

6.8 Experiment 17: sensitivity to statistics against sensitivity to structure after exposure to a segmented stream in 12-month-olds

6.8.1 Method

6.8.1.1 Participants

Sixteen 12-month-old infants (7 girls; mean age: 12 months 22 days; age-range: 12 months 2 days – 13 months 8 days) participated in Experiment 17 and were retained for analysis. An additional 22 infants took part but were excluded due to the following reasons: because of excessive fussiness during familiarization or during test phase (18), or because they exceeded maximum looking time criteria (4).

6.8.1.2 Stimuli and Procedure

Infants were exposed to the segmented stream and were tested with the same items as in Experiment 13. Procedure and data analysis were also identical to that of Experiment 13.

6.8.2 Results and Discussion

In Experiment 17, the average looking times data excluded from the analysis because they were shorter than 1 s, or 3 S.D. beyond the general mean computed for each Test Item Type amounted to 3.2% (8 out of 248 data points). The repeated measure ANOVA revealed that infants looked significantly longer while listening to high-TP part-words than to rule-words ($M_{\text{High-TP Part-words}} = 11.03$ sec, $SE = .91$; $M_{\text{Rule-words}} = 8.79$, $SE = 0.92$, $F(1, 15) = 5.34$, $p < 0.04$; Figure 22).

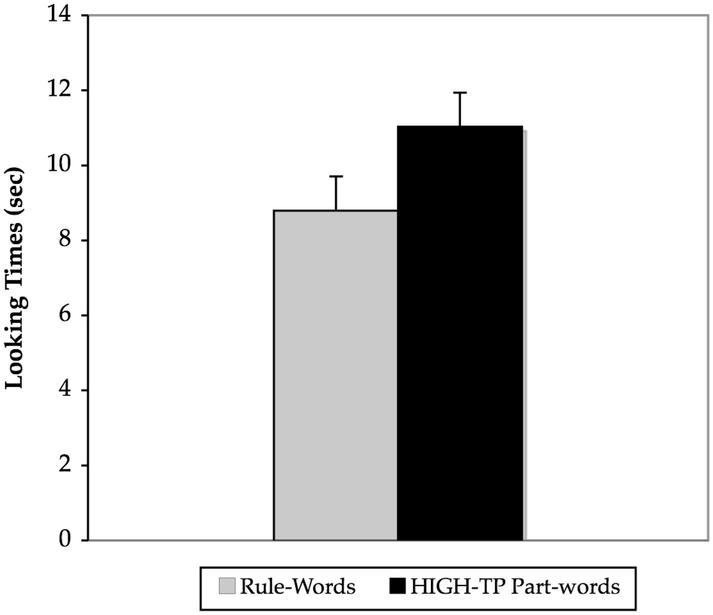


Figure 22: Mean looking time(s) and SE for the test items of 12-month-olds listening to rule-words and high-TP part-words after exposure to a segmented familiarization stream.

The result shows that 12-month-olds generalized structural regularities to rule-words, while at the same time, if they computed the distributional properties of the input, they did not use this information to detect high-TP part-words. A result attesting generalization was expected, considering that Experiments 1-3 already provided evidence about 12-month-olds' ability to project structural regularities after listening to a segmented stream. Therefore, Experiment 17 confirms the availability of the generalization mechanism from 12 months of age. More importantly, the result here obtained supports the hypothesis concerning the dissociation between the two mechanisms underlying generalizations and statistical computations. The predictions of the MOM hypothesis are also confirmed: the mechanism devoted at extracting structural regularities requires a segmented stream to be activated, and it does not depend on the computation of the input distributional information.

6.9 General discussion

Asking whether the discovery of structural information uses the same or different mechanism accounting for statistical learning is a question of central importance, as language acquisition theories disagree about the nature of the learning mechanisms. The six experiments reported in this chapter were aimed at isolating sensitivity to statistics from sensitivity to linguistic structure. Pitting these two sensitivities against each other within the same experiments allowed contrasting the assumptions of the associative learning theories and of the "MOM hypothesis". The former postulates that statistical mechanisms can account for language acquisition in all its complexity (that is, including grammar). In contrast, the MOM hypothesis assumes that distinct mechanisms are required to perform different linguistic computations (such as the extraction of the lexicon and the discovery of morphosyntactic structure). These mechanisms generate different linguistic representations, act upon different input properties and require different signal properties.

Infants of 18 and 12 months were exposed to a miniature artificial grammar. They were first familiarized with a stream containing strings of words occurring with high frequency and generating sequences spanning the word boundaries having high TP and high frequency. Then, they were tested with statistically occurring sequences and with novel sequences maintaining the same structural properties of familiarization words. Infants varied their performance depending both on their age, and on the specific signal properties.

Eighteen-month-olds exposed to the continuous stream could detect syllabic sequences that occurred in the stream with high TP and high frequency; such learning override preference for novel sequences respecting the structure of words. In contrast, they performed the opposite computations after exposure to a segmented stream, from which they generalized but could not detect statistically occurring patterns crossing the boundaries of words. Twelve-month-olds could also learn structural dependencies after exposure to a segmented stream. However, they could neither generalize, nor detect statistically occurring sequences after exposure to a continuous stream. As the only difference between the familiarization streams was the presence or the absence of silences between words, we can conclude that specific signal properties (i.e., cues demarcating word boundaries) had a crucial role in triggering different computational mechanisms.

The first important conclusion is that generalizations do not depend on the same mechanism infants use to compute statistics. This possibility is suggested by the dissociations in infants' looking behaviors, i.e., inverted preferences for rule-words against part-words in 18-month-olds (Experiments 12 and 13), and by the failure at computing statistics, but the success in generalizing, in 12-month-olds (Experiments 15, 16 and 17). Such dissociations are compatible with the hypothesis that generalizations and statistical computations rely upon distinct mechanisms, requiring different signal properties and suffering of different limitations. This evidence supports the view that the same sensitivity infants use to accomplish tasks having a statistical nature cannot explain how they acquire grammar-like structures (as predicted by the MOM hypothesis). This conclusion is of particular relevance as it reinvigorates the debate on the nature of the language acquisition mechanism between the associative learning models and the dual-mechanism model.

The different performances of the two age groups suggest that sensitivity to structure and to statistical information differ not only in their nature, but also in their availability and in their limitations during development. Generalization appears to be readily available from 12 months (as established also by Experiments 1-3, see Chapter 5). In contrast, sensitivity to statistical information is efficient in 18-month-olds, but it is rather limited in 12-month-olds. Statistical computations of one-year-olds appear to be confined to the simplest learning conditions, such as perfect adjacent TP (Saffran et al., 1996), or low adjacent TP (Experiment 5 on Chapter 5; Saffran & Wilson, 2003; Gómez & Gerken, 1999), but only if the linguistic input does not present additional layers of complexity, such as long sequences of high TP and high frequency words (Experiment 17).

We may think that the statistically defined sequences used in the test phase of these studies are particularly difficult to learn, as they span the boundaries of words. Part-words have been contrasted to words in all the studies investigating computational abilities in infants: in these studies, infants could learn words, and part-words were rejected as "illegal" word candidates (Saffran et al., 1996a; Aslin et al., 1998). Part-words are much less cohesive statistical units than words: this is to say that contrasting them to rule-words may be a rather difficult task for infants. However, we should consider that the stream used in the current studies was created in a way it provided an optimal condition to the learning of syllabic patterns that would otherwise not be easily detected, as they cross the boundaries of words. Frequent words were concatenated together precisely to generate sequences that, while spanning the boundaries of frequent words, are statistically informative. Such sequences had high adjacent and nonadjacent TPs (equal to .67), and high frequency. In principle, TP values of .67 should suffice for the extraction of sequences occurring in a stream, even if they cross the boundaries of words. Otherwise, we should reconsider both the power and the reliability of the infants' statistical computations to acquire words in real language situations – where TPs within words are obviously much lower than .67 (Yang, 2004).

Still, the 12-month-olds' failure to detect statistically occurring patterns challenges -- together with the 7-month-olds' failure at extracting word-like sequences (see Experiments 10 and 11 on Chapter 5)-- the view that young infants are powerful statistical learners (Saffran et al., 1996a; Saffran et al., 1996b). If infants below one year are not so skilled at computing statistics from the speech input, then acquiring language should represent an enormously difficult task (as real languages contains thousands of words, thus TPs between syllables are very low). Still, it is known that infant begin to segment words of their native language in the second half of their first year (Jusczyk, 2001; Jusczyk, Houston, & Newsome, 1999; Houston, Santelmann, & Jusczyk, 2004; Johnson & Jusczyk, 2001). Therefore, the importance of statistical learning in accounting *alone* for word segmentation or for multi-level segmentation (such as in segmenting linguistic units larger than words: Saffran & Wilson, 2003) has to be reconsidered. For the same reason, serious doubts should be raised about the possibility that linguistic structures can be entirely learned from associations between linguistic elements (as instead proposed by associative learning models;

see also Gómez, 2002; Gómez & Gerken, 1999; Mintz, 2003).

CHAPTER 7

MORPHOLOGICAL REGULARITIES IN THE LINGUISTIC INPUT: DISTRIBUTIONAL ANALYSES OF CHILDES ITALIAN CORPORA

Two primary factors determine the outcome of language acquisition: the properties of the learners' computational and representational systems, and the linguistic environment. It is reasonable to think that the aspects of language acquisition differ in the degree to which they depend on experience with the linguistic input. For instance, the learner's vocabulary is strongly shaped by the particular words to which s/he is exposed. At the same time, many aspects of language must rely on internal cognitive factors, as suggested by the fact that non-humans do not learn languages even with extensive exposure. As the learners' computational endowment and the input properties pose mutual constraints one on the other, we need not only to understand what acquisition mechanisms learners use to acquire language, but also to know what information have to be acquired in order to understand how such mechanisms work.

The studies reported in Chapters 5 and 6 investigated the computational mechanisms infants use to acquire words and the rules governing their structure. The investigations contained in the present chapter will approach the problem of morphosyntactic acquisition from a complementary perspective, asking whether the linguistic input available to young learners offers reliable information about the rules determining words' compositional structure.

To assess the availability of morphosyntactic rules in the linguistic environment children are exposed to, I will examine the distributional properties of child-directed corpora in a language with a rich and complex morphology, i.e., Italian, exploring the source of information contained in it. This investigation will be aimed to evaluate the extent to which the input directed to young children contains words respecting or violating the morphosyntactic rules of Italian.

The chapter is organized as follows. The first part briefly presents the opposite stances taken by nativist and associative learning theories about the role of the distributional properties of the input in explaining language acquisition. The first section will also review the few studies currently available documenting the distributional patterns of the speech input, evaluating whether they are informative and thus can potentially instantiate some specific grammatical constructions. These studies focused on various aspects but, at present, little is known about the morphosyntactic properties of speech directed to children. The second part of the chapter presents a series of analyses performed on child-directed input to explore this issue, asking whether the input received by children contains reliable information to the acquisition of morphological rules. In the third part, the results of the distributional analyses will be discussed in their implications for morphosyntactic acquisition.

7.1 What is the role of experience with the linguistic input in explaining language acquisition?

The information contained in the input has been of central importance to many theories of language acquisition. However, theories disagree about the relevance of such information in determining the acquisition of language. Nativist theories argue that the distributional properties of input are insufficient to account for the linguistic knowledge children can attain, and postulated the existence of innate rules, guiding and constraining the process of converging to the correct grammar of a language (Chomsky, 1957; 1980; 2000; Pinker, 1984; Marcus, 1993; see also Chapter 1). Thus, the nativist approach assumes that the distributional analyses performed on the linguistic input are highly assisted by innate knowledge of linguistic categories and configurations.

Contrary to this position, associative learning theories, old (Skinner, 1957) and new (Tomasello, 2000), took the opposite stance, stating that language is acquired by the conditioning and reinforcement mechanisms. This view determined a growing interest for assessing the distributional properties of the speech input in order to understand whether it could plausibly offer a basis to the extraction of occurring patterns such as words (Brent & Cartwright, 1996; Swingley, 2005), to the abstraction of grammatical categories (Mintz et al., 2002; Mintz, 2002, 2003), and to the acquisition of grammatical structures (Gervain et al., 2008b; see also Section 3.2.4 on Chapter 3).

While the abovementioned studies established that the input to children contains sufficient information to account for their linguistic representations, other studies showed that the input does not unambiguously support the acquisition of some aspects of syntactic knowledge (Lidz et al., 2003; Yang, 1999) or that certain languages show morphosyntactic properties that preclude conditional probabilities from operating efficiently (Yang, 2004; see also Section 3.2.4 on Chapter 3). These contrasting conclusions suggest that, although learners may use distributional analysis to acquire some aspects of their language, any such analysis can alone lead the learner to the correct grammar.

To date, little is known about the morphosyntactic properties of the input directed to children, in terms of evidence that would led them to acquire the morphosyntactic rules of a language by exploring the surface forms of words. Most of the studies analyzing the acquisition of morphosyntactic rules concentrated on the distributional properties of the first words acquired by young children (see Clark, 1998 for a review; for Italian and English languages, see also Rinaldi, Barca, & Burani, 2004), and very few investigated the distributional properties of the words they listen to when they are exposed to the linguistic environment (Soderstrom et al., 2008). Hence, it is currently poorly known the extent to which distributional patterns inherent to the speech directed to children are informative and potentially viable bases for the discovery of morphological regularities.

The present chapter reports a series of distributional analyses that have been conducted seeking to evaluate the distributional information of the speech to young children, in terms of availability of morphosyntactic structures. The main goal of this work is to start interrogating about the potential contribution of experience with the linguistic input in the acquisition of such grammatical regularities.

7.2 Distributional analysis on CHILDES Italian Corpora

The distributional analyses reported in this chapter are aimed at evaluating the morphological consistency of words used by adults talking to young children during everyday life spontaneous conversations. More specifically, the present investigations will examine the proportion of well-formed words to the overall amount of words used by adults to communicate with young children. Importantly, the statistical analyses will consider only the adult use of words in their interaction with children (and not the linguistic productions of children).

The way in which adults speak to children appears as a simplified talk, using short sentences, referring to the “here and now”, with more redundancy, with slow speech rate, reduced vocabulary and exaggerated prosody (Dockrell,

& Messer, 1999). Although the use of child-directed speech is widespread, it is not universal across all cultures (Pye, 1986). Furthermore, there is great variation in the styles of social interaction and in the form of child-directed speech across different cultures (Lieven, 1994). While it is not clear whether child-directed speech play a role in language acquisition, it is reasonable to think that it might nevertheless facilitate it (Gleitman, Newport, & Gleitman, 1984; see also Soderstrom, 2007 for a critical review).

The potential source of information of the linguistic input will be measured with a statistical cue commonly used in experimental psycholinguistics, namely frequency of occurrence. The analyses will evaluate the morphosyntactic consistency of words contained in child directed speech both in quantitative and qualitative terms. The distributional analyses performed on the input directed to children will be restricted to regularities at the word level, that is, they will not consider the interaction of morphology with semantics or syntax (these issues, of course, would deserve attention for other reasons).

For the sake of clarity, the next section will briefly describe the morphological system of Italian.

7.2.1 Italian and its morphology

Italian is romance language using several grammatical morphemes to express morphosyntactic relations (see also Chapter 1, example (3)). In this language (like in other languages of the same morphological typology), words may contain several grammatical morphemes “fused” together. For example, Italian nouns are inflected in number (singular, plural) and usually in gender (feminine, masculine), such as:

(6) Bambola
doll.SG.FEM
“doll”.

Italian verbs are inflected in mode, tense, number, person and sometimes in gender (in the participle form). For example:

(7) a. andremo
go.FUT.1PL
"we will go".

b. mangiati
Eat.PART.3PL.MAS
"[they] [have been] eaten".

The examples in (7) illustrate that, in Italian, the pronominal subjects can be omitted because the identity of the surface subject(s) can be recovered from inflectional morphemes.

Italian words can contains several affixes, such as:

(8) antipaticissime [anti+pat+ic+issim+e]
unfriendly.FEM.3PL
"very unfriendly"

Italian is a language of particular interest because of its complex morphology: Italian words have a rich and informative structure, reflecting the morphosyntactic rules of this language. Hence, in Italian (as well as in other languages making an extensive use of morphology to mark grammatical relations and distinctions; see also Chapter 2) words have to be taken into primary consideration by the learners in order to acquire morphosyntactic rules, as they may offer a surface cue to the acquisition of such regularities. Then, the question is to what extent such grammatical rules are consistently represented in words available to children when they listen to their native language.

7.2.2 Method

7.2.2.1 Material: the CHILDES Italian databases

The four Italian sub-corpora from the CHILDES database (MacWhinney, 2000) served as an input for the analysis procedure: The Antelmi corpus (Antelmi, n.d.), The Calambrone corpus (Cipriani et al., 1989), the Roma corpus (Antinucci, & Parisi, 1973), and the Tonelli corpus (Tonelli, n.d; Cipriani et al., 1989).

The sub-corpora contain transcribed spontaneous conversational interactions between young children and their caretakers. The speakers involved are often young monolingual, normally developing children conversing with their parents, relatives or siblings. Only the Calambrone corpus contains data from both normal and disordered language developing children; however, the data from the disordered children were not included in the corpora used to conduct the present investigation.

Across the four databases, the children started to be followed after their first year, till around their third-fourth year (age range across the four database: 1 year 3 month - 4 years 0 months; see Table 4 for a detailed report of the age-range for every database). The Italian databases do not include any transcripts from bilingual children, older school-aged children, or adult second-language learners. Additional information on the files included in the analysis is reported in Table 4.

The next section describes how the list of words used in the conversations was extracted from each database, and how the data from the separate corpora where collapsed together in order to obtain a general list of words used across the four databases. This general list of words was then entered into the statistical analysis.

Database	Child	Age- range	Sessions	Files
Antelmi	Camilla (F)	2;2 - 3;4	Cam202; Cam204; Cam206; Cam209; Cam211; Cam301; Cam304.	7
Calambrone	Raffaello (M)	1;7- 3;3	raf01-raf17	17
	Guglielmo (M)	2;1 - 2;11	greg01-greg09	9
	Rosa (F)	1;3 - 3;4	rosa01-17; rosa17-rosa19	20
	Martina (F)	1;7 - 3;0	mart02-mart09; mart11-mart14; mart16	13
	Viola (F)	1;10 - 3;1	viola01-viola10	10
	Diana (F)	1;6 - 3;1	diana01-diana06; diana08-diana10	9
Roma	Francesco (M)	1;4 - 4;0	fra01-fra10	10
Tonelli	Gregorio (M)	--	gre01-gre08	8
	Marco (M)	--	mar08-mar26	19
	Elisa (F)	--	eli01-eli08	8

Table 4: General information about the four CHILDES Italian databases used to conduct the statistical analyses. The second column reports the name and the gender of the child (F= female, M=male). The third column reports the age-range, corresponding to the period during which the conversations were collected (this information is not available for the Tonelli database). The fourth column contains the conversational sessions taken into considerations and entered into the analyses. The last column reports the number of conversations collected for each child.

7.2.2.2 Distributional analysis procedure

The first part of the distributional analysis procedure aimed at constructing a list of all the different words used across the different conversational files for each database. The procedure used to obtain the general list of words was as follows.

First, the words used by the adult speakers were extracted from each conversation file. Care was taken to maintain mispronunciation or truncated words as originally transcribed; even if the CHILDES conversational sessions are orthographically transcribed, they can be considered as phonologically transcribed, as in Italian graphemes almost correspond to phonemes. Second, the text files containing the list of words extracted from each conversation were entered into a general file. This general file contained 130 conversations, and the word tokens were in total 221557. Third, this general file (containing all the conversations from the four databases) was entered into Datadesk, a software designed for conducting statistical analysis. Using this software, an overall list of word forms used across the different corpora was obtained. This list corresponds to the overall vocabulary. This vocabulary contained 8966 word types (or word forms). Table 5 reports the types and tokens numbers for each database, and for the general list resulting from collapsing the four corpora together.

CHILDES Database	Tokens (Sum of Frequencies)	Types (Vocabulary)	Type/Token Ratio (vocabulary diversity)
Antelmi	8176	1319	.16
Calambrone	143578	6485	.05
Roma	7536	853	.11
Tonelli	62267	4560	.07
Overall	**221557**	**(8966)**	**.04**

Table 5: Summary of the total number of tokens and types contained in each database, and in the general corpus. The overall word types are extracted after having collapsed the four databases together, and it corresponds to the general list of words used across the different corpora. Brackets are used to indicate that the overall types do not correspond to the sum of the word types.

7.2.2.3 Coding word types as "legal" or "illegal" Italian words

Two Italian native speakers, naïve to the purpose of the study, checked the general list of word forms. They were asked to judge whether each of the entries was a well-formed, or "legal", word in Italian or not. In some exceptional cases, the two raters gave discrepant opinions concerning the status of a specific word. The final decision of accepting or rejecting an entry as a "well-formed" Italian word was taken after having consulted an Italian vocabulary (De Mauro, 2000). Those words contained in the Italian vocabulary were accepted, and marked as "legal Italian words". This set comprised 7491 word types, and it was made up of 211980 word tokens. In contrast, those words that were not contained in the on-line vocabulary were rejected and marked as "illegal Italian words". This set consisted of 1475 word types, and it was made up of 9577 word tokens.

7.2.2.4 Assigning "illegal" words to sub-categories

Each of the illegal words was then classified into one of the following eight sub-categories:

1) *Onomatopoeia,* i.e., a combination of sounds in a word that imitates or suggests what the word refers to, e.g. "*gnam gnam*" to indicate the mouth opening and closing when eating (this example, as all the others reported in the next categories, are taken from the Italian CHILDES corpora). Given that the current investigation focuses on the morphological structure of Italian words, and that onomatopoeias don't contain grammatical morphemes, they were excluded from the subset of "legal Italian words". This choice was rather conservative, considering that onomatopoeias are very typical of speech directed to children, thus, we cannot exclude *a priori* that they may play a role in facilitating some aspects of language acquisition (others than morphosyntactic acquisition);

2) *Mispronunciation,* i.e., words that are pronounced wrongly, for instance substituting a segment, enlarging a vowel, altering the duration of consonant in gemination. For example, *"bandieri<u>m</u>a"* instead of *"bandierina*", "small flag", where the consonant "n" is substituted with "m";

3) *Elision*, i.e., omission of a phoneme of a word (for example, "*ancoa*" instead of "*ancora", "more*");

4) *Truncation*, i.e., a shortened word, where its initial or final part has been removed (e.g., *"baratt"* instead of "*barattolo*", "*pot"*);

5) *Dialectical word*, i.e., the vocabulary used in a part of a country, or by a class of people, thus not conforming to the morphological rules of standard Italian (e.g., "*iamme*" instead of *"andiamo", "let's go*"). The choice of excluding dialectical words from the "legal words" subset was conservative, as dialectical

words do not conform to grammatical rules of Italian, hence they don't contribute to the acquisition of its morphosyntactic properties. However, we should not deny that dialectical words are well-formed with respect to the system of rules governing them – thus, they may contribute to the acquisition of dialect;

6) *Familiar word*, i.e., a form used inside the familiar context, generally child-invented words that have been taken over by the all family (for example, *"bumba", "tato"*), or child's mispronunciations or Italian words (for example, "*apriri"* instead of *"aprire", "to open"*). Familiar words are forms conventionally used within a restricted group of speakers; therefore, they constitute part of the vocabulary shared by the speakers of this group. Even if familiar words are treated as "legal" words within this small speaking community, they were excluded from the subset of "legal Italian words" as they generally do not conform to the morphosyntactic rules of standard Italian. Also this choice was rather conservative, as it was likely to overestimating the impact of "illegal Italian words" over the general distributional properties, rather than overestimating the amount of legal words;

7) *Non-words*, i.e., words that are meaningless, and cannot result from any of the phenomenon described in sub-categories 1-6 (for instance, "*angianda*");

8) *Ambiguous words:* this category contains the non-words that cannot clearly enter into any of the other categories as they can be potentially assigned to more than one of them. For example, "*afri*", which has no meaning, but can result from truncation (instead of *"Africa"),* or mispronunciation (instead of "*apri", "you open*"), or can be a non-word etc.

To summarize, word forms were first categorized into the two main groups of "legal Italian words" and "illegal Italian words". Then, "illegal Italian words" were assigned to the 8 sub-categories. Quantitative and qualitative analyses were performed on both levels of categorization.

7.2.3 Quantitative and qualitative analysis

The quantitative assessment concerned the groups of legal / illegal words, considered both separately and collapsed together. The first analysis calculates the percentage of legal and illegal word tokens and types in the overall corpora. The second analysis concerns the distributional properties of the "legal Italian words". Finally, the third analysis evaluates the distributional properties of "illegal Italian words" considering their assignment into the eight sub-categories.

7.2.3.1 Quantitative analysis on legal / illegal word groups

The first analysis concerned the proportion of between legal and illegal words, both in terms of word types and word tokens. The legal word types were 7491, representing the 83.55% of the total word types. The illegal word types were 1475, representing the 16.45% of the total word types. This proportion indicates that 1 word form out of 6 was morphologically illegal. However, this measure is not informative if taken alone, because other properties such as token frequency are necessary to quantify the impact of illegal words over the input general properties. Legal word tokens were 211980, representing the 95.68% of the total tokens. Illegal word tokens were 9577, representing the 4.32% of the total tokens. The type/token ratio for legal words was 0.04. The type/token ratio for illegal words was 0.15 (to note that the smaller the type/token ratio, the smaller the vocabulary variation).

Taken together, the results indicate that illegal words are quite varied (i.e., they constitute the 16.45% of the total vocabulary), but they occur with very low frequency (as indicated by their percentage over the total word tokens, 4.32%). High variation and low frequency would make illegal words difficult to memorize. In contrast, legal words represent the majority of the linguistic entries (95.68% of the word tokens). The type / token ratio (0.15) indicates that (on average) legal words are consistently repeated over the corpus.

The next analysis will evaluate the distributional properties of legal words in greater details.

7.2.3.2 Distributional properties of the legal word forms

Figure 23 represents the frequency distribution (in percentage) of legal Italian words. The frequency values range from 1 to 7880, with a high variability between the different word types (as suggested by the SD = 207.75). The mean frequency is 28.3; the mode is 1, and the median is 2. An informal analysis of the frequency distributions informally suggests that a few word types appeared a large number of times, and a large number of word types appear only a few times (see Figure 23).

The cumulative percentage of the legal word types indicates that the 55 most frequent words account for 50.26% of the overall tokens (106554 word tokens out of 211980) and for 0.7% of the overall types (55 word types out of 8966). The 55 most frequent words occurred with frequencies ranging from 7880 to 732, and the type / token ratio is 0.0005. It should be noticed that some of these most frequent words are uninflected words (like prepositions, pronoun, adverbs or grammatical particles as "si", i.e., "yes", and "no"), or words that can be both inflected and uninflected, depending on their syntactic role (for instance, "cosa", which can be a noun, "thing", or an adjective or pronoun when used with the meaning of "what").

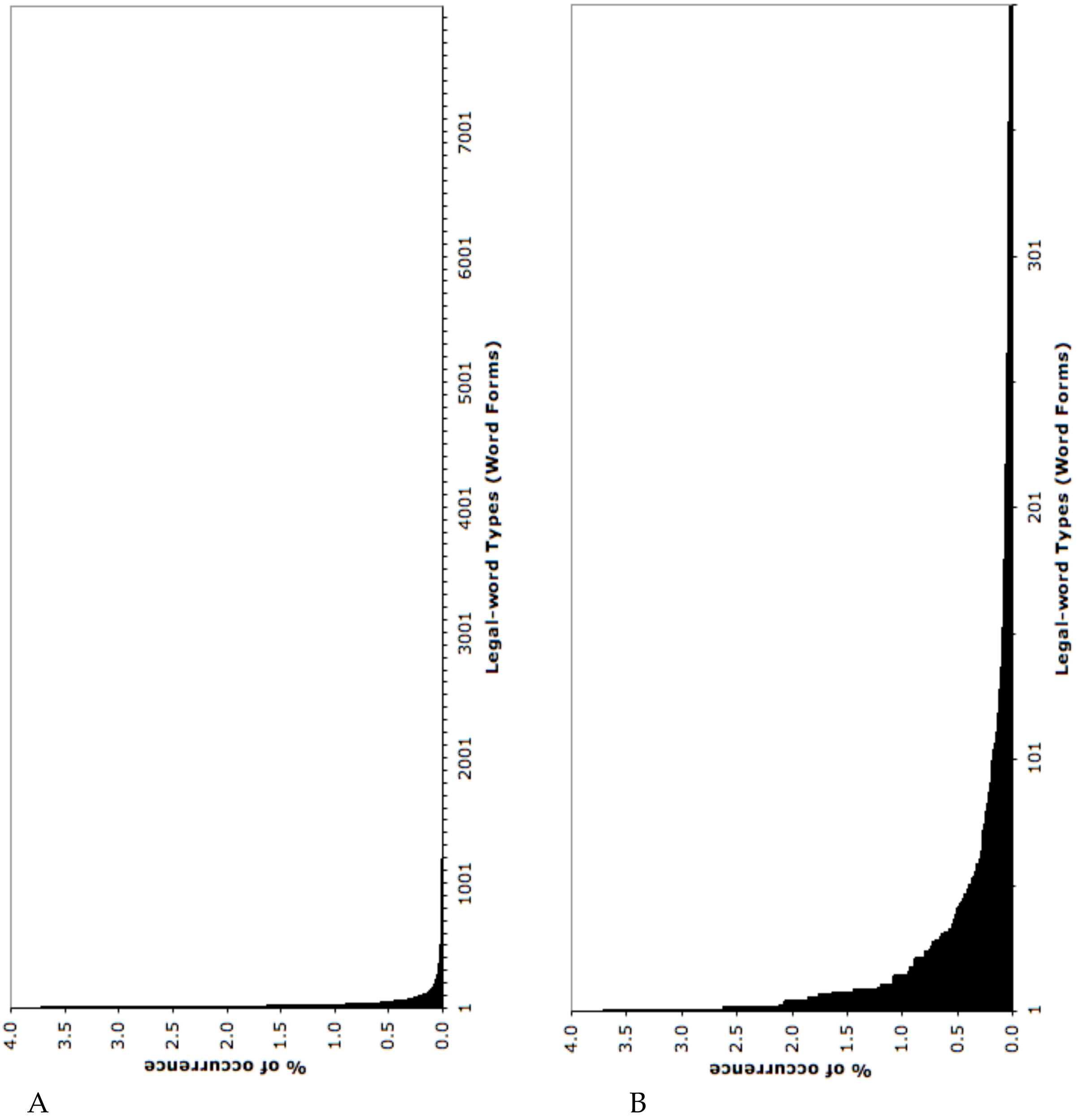


Figure 23: The histogram of frequency of occurrence of the 7491 legal word types used by adult speakers as transcribed in the Italian CHILDES corpora. The x-axis represents the word types, and the y-axis shows percentage of occurrence. The x-axis and the y-axis are plotted only up to the highest values found. B: A zoom on the y-axis. The scale on the x-axis is magnified (cutting off at 400 as the maximum value) for a better visualization the distribution of the 400 most frequent word types (corresponding to the higher frequency values in the y-axis).

Adopting the most conservative criterion, i.e., categorizing as uninflected also those words that can actually vary their assignment depending on their syntactic role, the uninflected word types are 29 and the uninflected word tokens are 53528 (25% of the total). Instead, the inflected word types are 26 and the inflected word tokens are 53026.

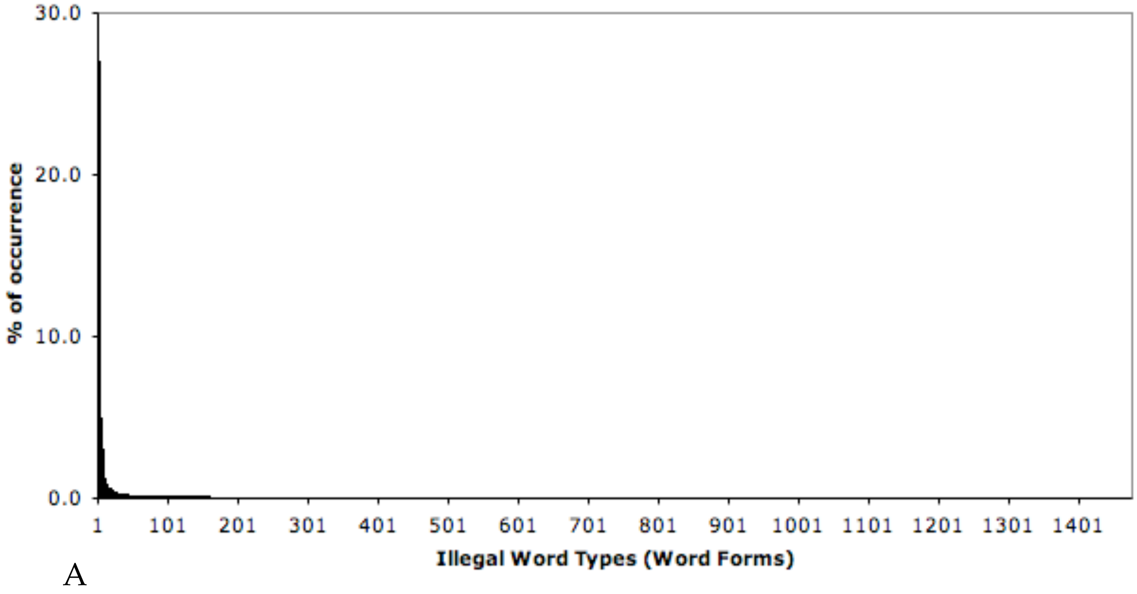
If we consider the 400 most frequent words, and we exclude the uninflected words from this pool (88 word forms), the remaining 312 inflected word forms account for around 50% of the overall tokens (103305 out of 221557), while they represent only the 3.6% of the overall word types (that is, of the total vocabulary). They occurred with frequencies ranging from 7880 to 61, and their type/token ratio is 0.003 (indicating an extremely low vocabulary variation).

The remaining 7091 word types (79% of the overall word types) occurred with frequencies ranging from 61 to 1. Their mean frequency is 5.88, the S.D. is 9.36; the mode is 1 the median is 2. The word tokens are 41716 (18.8% of the total), and the type/token ratio is 0.17. If we compare the distributional properties of the less frequent legal words with that of illegal words we can conclude that, while the have similar type/token ratios (0.17 and 0.15, respectively), less frequent legal words appear four times more frequently than illegal words.

7.2.3.3 Distributional properties of illegal words

Two sets of analysis were performed to evaluate the distributional properties of illegal words. The first set of analyses examined in quantitative terms the frequency distribution of illegal words. The second set of analysis evaluated in qualitative terms the distributional properties of illegal words, depending on the categorization into the eight sub-sets as in Section 7.2.2.4.

Figure 24 shows the frequency distribution of the 1475 illegal word types. An informal analysis of the frequency distribution of illegal word types suggests that a few types occur with high frequency, while the large number of types occurs with extremely low frequencies.



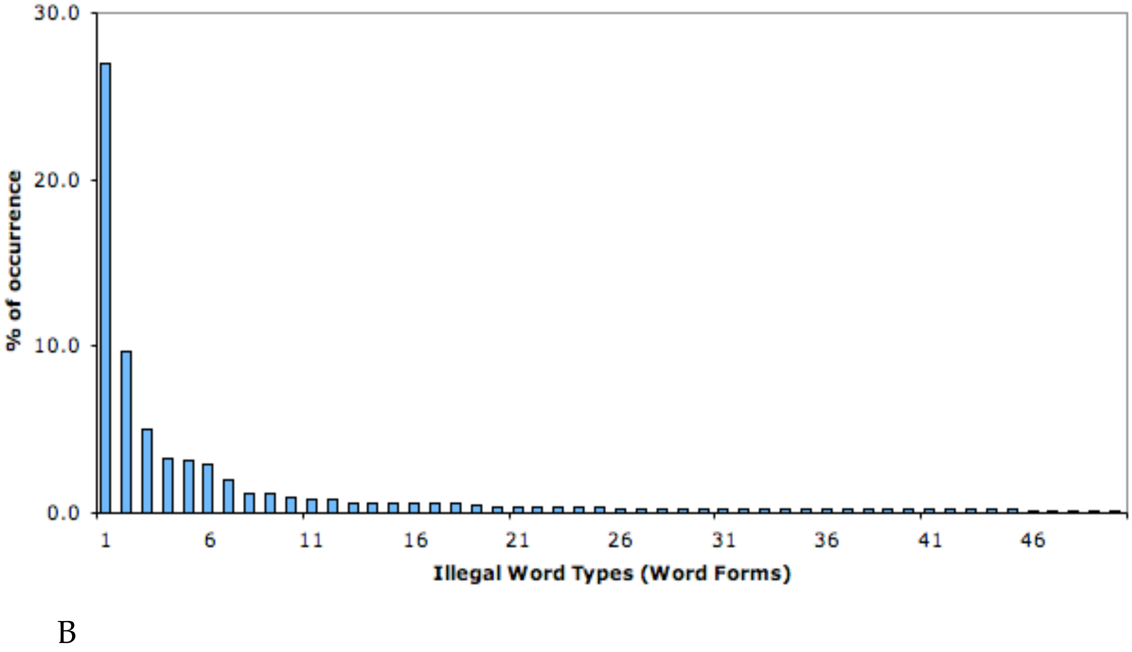


Figure 24: The histogram of frequency of occurrence of the 1475 illegal word types used by adult speakers in the Italian CHILDES corpora. The x-axis represents the word types, and the y-axis shows percentage of occurrence. The x-axis and the y-axis are plotted only up to the highest values found. B: A zoom on the y-axis. The scale on the x-axis is magnified (cutting off at 50 as the maximum value) for a better visualization the distribution of the 50 most frequent illegal word types.

The six most frequent illegal words account for 51.05% of the illegal word tokens (4886 word tokens out of 9577). Four of these six word types are onomatopoeias; one is a dialectical word, and another one is assigned to the "ambiguous word" category (see also Table 6). Importantly, onomatopoeias, dialectical and familiar words are not "illegal" words in a strict sense. Rather, the onomatopoeias constitute sound patterns that are consistently repeated during conversations; both dialectical and familiar words are linguistic forms that, even if they violate the Italian morphology, are conventionally used within a linguistic group or community and conform to morphosyntactic rules. If onomatopoeias, dialectical and familiar words are excluded from the "illegal Italian words" group, then the word tokens still included in it are 2508, constituting the 1.13% of the total input (instead of the 4.32%). The word types are 969 (instead of 1475), constituting the 10.81% of the total input (instead of the 16.45%).

	Word	Sub-Category	Frequency	%	Cumulative %
1	eh	onomatopoeia	*2581*	26.95	26.95
2	ah	onomatopoeia	*925*	9.66	36.61
3	oh	onomatopoeia	*477*	4.98	41.59
4	cosa'	ambiguous	*312*	3.26	44.85
5	pio'	dialectical word	*306*	3.20	48.04
6	uhm	onomatopoeia	*285*	2.98	**51.02**

Table 6: The six most frequent illegal words used by adult speakers in the CHILDES Italian corpora. The table includes frequency of occurrence of a specific word form, percentage of occurrence over total illegal word tokens, and cumulative percentage.

Figure 25 represents the percentages of the illegal word types and tokens falling into the eight sub-categories (Section 7.2.2.4 on this Chapter); Table 7 reports the numbers of types and tokens for each sub-category. Figure 25 and Table 7 indicate that most of the illegal words are onomatopoeias, representing the 66.45% of the illegal-word tokens (6364 out of 9577), and the 2.87% of the overall tokens.

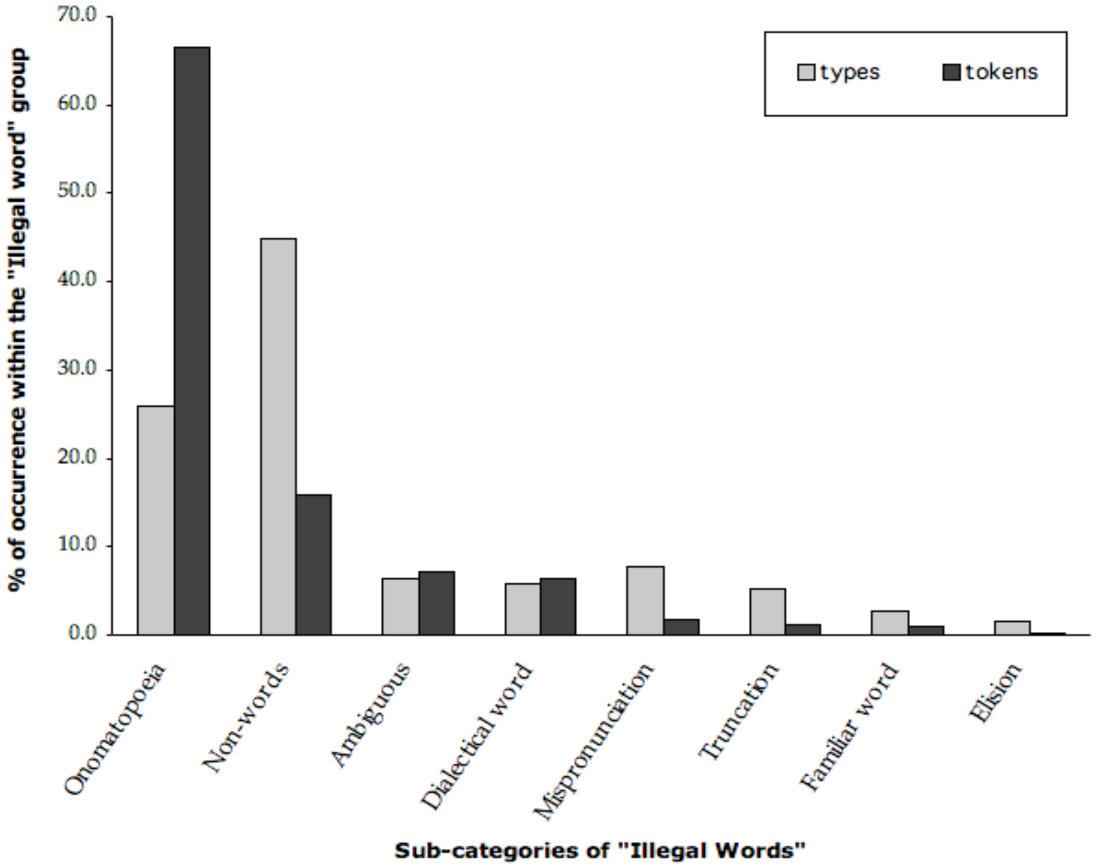


Figure 25: The histogram showing the frequency of illegal word types and tokens, assigned to the eight sub-categories.

SUB-CATEGORY	Tokens	%	%over Total	Types	%	% over Total
Onomatopoeias	6364	*66.45*	2.87	381	*25.83*	4.25
Non-words	1520	*15.87*	0.69	661	*44.81*	7.37
Ambiguous entries	690	*7.20*	0.31	94	*6.37*	1.05
Dialectal words	613	*6.40*	0.28	85	*5.76*	0.95
Mispronunciation	164	*1.71*	0.07	115	*7.80*	1.28
Truncation	109	*1.14*	0.05	77	*5.22*	0.86
Familiar words	92	*0.96*	0.04	40	*2.71*	0.45
Elision	25	*0.26*	0.01	22	*1.49*	0.25
TOTAL	9577	100.00	4.32	1475	100.00	16.45

Table 7: Details about the illegal word tokens and types, assigned to the eight sub-categories.

This sub-category contains 381 word types, corresponding to 25.83% of the illegal-word types, and to 4.25% of the overall word types.

The most varied category is "Non-words"; it contains 1520 word tokens, corresponding to 15.87% of the illegal-words tokens, and to 0.69% of the overall word tokens. The word types falling into this sub-category are 661 (out of 1475 illegal word types), corresponding to 44.81% of the overall illegal word types, and to 7.73% of the overall word types.

The remaining six sub-categories contain a total of 1693 word tokens, corresponding to the 17.68% of the illegal word tokens, and to 0.76% of the overall tokens. The illegal word types falling into the six categories are in total 433 (ranging from 22 to 115; see Table 7, column "Type"), corresponding to 29.36% of the illegal word types, and to 4.83% of the overall word types.

7.2.4 Discussion

Several statistical analyses were conducted on words used by adult speakers to talk with young children. The first series of analyses calculated the percentages of legal and illegal word types and tokens, and computed the type/token ratios for each group. The results showed that legal words are the dominant part of the speech input. They almost account for the full available input, both in terms of vocabulary (83.55%) and of word tokens (95.68% of the total). Such distributional properties are likely to favor the representation of legal words against that of illegal words. In contrast, illegal words occur with extremely low frequency (they represent the 4.32% of the total word tokens), and they account for a small part of the total vocabulary (16.45%). Such distributional properties would translate into high variability of the word types, combined with low frequency of the word tokens. Items having such statistical properties are likely to be difficult to memorize and be represented.

The second set of analyses evaluated the distributional properties of legal words, both with quantitative and qualitative measures. The frequency distribution revealed that very few word types occurred with extremely high frequency, and accounted for most of the word tokens. The 55 most frequent legal words (representing less than 1% of the vocabulary) accounted for the 50% of the legal word tokens, half of which are represented by uninflected legal words (like prepositions, adverbs, etc.). Instead, considering the 400 most frequent word types (5.2% of the total vocabulary) implies that we can account for the 80% of the legal word tokens – and for around 50% of inflected legal words. Even if in this second case the vocabulary is more variable (as more word forms are included in the sample), still each word form occurs (on average) with extremely high frequency. Overall, these results indicate that the dominant part of the linguistic input directed to children contains a relatively small number of overrepresented legal words, most of which are inflected. This learning condition is likely to facilitate the memorizations of such words, which, in turn, may provide a reliable basis for the acquisition of morphosyntactic rules, as most of these words offer systematic and consistent evidence of such rules.

Finally, the third set of distributional analyses revealed that illegal words are not favored in statistical terms: they represent the minor part of the linguistic input (4.32% of the overall tokens) and they fall into a big number of word types (1475, constituting the 16.45% on the overall vocabulary). These distributional properties would translate into high variability and low frequency of occurrence; this learning condition is unlikely to facilitate the learning of illegal words. Qualitative analyses indicated that most of the illegal word forms are represented by onomatopoeias (66.46% of illegal word tokens). To note that onomatopoeias are not “illegal” words in a strict sense; rather, they are conventional sound patterns that imitate what the words refer to, they are typically very frequent in speech directed to children. Excluding also dialectical and familiar words (not respecting the morphosyntactic rules of standard Italian, but configuring as words conventionally used within a specific linguistic community), the remaining five categories of illegal words contain non-words or sound patterns generated from the phonological processes of elision, truncation and mispronunciation. The items falling into these five remaining categories accounted for 1.13% of the overall tokens and for the 10.81% of the overall types. Hence, illegal words are quite varied and occur extremely rarely in the speech directed to children.

Overall, the results indicate that the distributional properties of the child-directed speech may provide a reliable basis to the acquisition of high frequency legal words that, in turn, offer a systematic evidence of the morphosyntactic rules. A relatively small vocabulary (comprising of 400 entries) can be readily available to children by computing the frequency of occurrence of most frequent words. This initial vocabulary may constitute the basis from which young learners can start analyzing words in their sub-parts, and, on this basis, they can possibly discover the morphosyntactic regularities underlying word structures. Word forms violating these rules are rare; in virtue of their distributional properties, illegal words are unlikely to be represented, thus we may speculate that they probably do not interfere with the discovery of morphological rules underlying the structure of legal words.

7.3 Conclusions

The distributional analyses presented in this present chapter investigated some aspects of the statistical information contained in the speech signal to young children. In particular, child directed corpora of Italian were studied in order to assess their morphological properties, determining what information might be available in speech from which young learners can derive the morphosyntactic properties of their native language. To that end, the words used by adult speakers to interact with children at the age of 1-4 years were coded as morphologically “legal” or “illegal” (with respect to the grammatical system of the language taken into consideration, i.e., Italian). A series of distributional analyses was conducted, targeting the morphological level, *i.e.*, the atomic units were the words contained in speech, without considering their semantics and/or their syntactic roles.

The main conclusion of the investigation is that the linguistic input is highly systematic and consistent. The sources of statistical information contained in the signal (as measured by the frequency of occurrence) may provide a reliable basis from which young learners can start acquiring morphosyntactic rules, provided that the most frequent word forms are also legally constructed words - thus, analyzing their structure may potentially inform about the rules governing their composition.

The linguistic input is almost entirely represented by "legal" word forms. Among them, a relatively small sample of word types appeared with extremely high frequency. Words occurring with high frequency and with low variability would facilitate the acquisition of an initial vocabulary, from which children may plausibly start extracting structural regularities. In addition, "illegal" word forms (that is, words violating the possible forms a word can take in Italian) appeared with very low frequency and accounted for a small part of the linguistic input (extremely low if we consider the frequency of individual word forms). In other terms, the learning of illegal words does not seem to be favored in statistical terms.

Overall, the results suggest that the speech directed to children contain sufficient and reliable evidence of the morphological rules of a language. However, as these investigations have been conducted on child-directed corpora on just one language (i.e., Italian), they cannot provide alone an exhaustive answer to the question concerning the potential contribution of the speech input to the acquisition of morphosyntactic rules. Instead, cross-linguistic comparisons are required to assess the distributional properties of different languages, varying in their morphological typology. For instance, we currently do not know whether languages differing in the amount of grammatical morphemes they use to express morphosyntactic relations would differ in their distributional properties (at the word level).

While the results of the presents investigations showed that the input directed to children is morphologically adequate and reliable, this conclusion does not imply that statistical learning alone is sufficient to explain the acquisition of morphosyntactic rules. There is an important evidence suggesting that children do not simply match the distributional properties of the input with their linguistic representations: children tend to produce over-regularization errors. That is, they create brand-new (non-existing) words conforming to an abstract rule that, in most of the cases, they extend to (existing) irregular words (e.g., “comed” instead of “came”, see Marcus et al., 1992; Marcus, 1995, 1996; Pinker, 1995). These errors cannot be explained if we assume that what young learners are doing is simply matching their linguistic knowledge to the (morphological) properties of the input. Rather, over-regularizations seem paradigmatic of the creation and the application of an abstract rule - although some authors have estimated that they are relatively infrequent in children’s speech, e.g., 3-10 percent across different children and ages (Marcus et al., 1992).

Therefore, the results of the corpora analyses cannot be interpreted as evidence that what learners have to do is just to construct morphological rules from distributional analysis. Rather, they point at the *constraints* a learner would need to have in order to benefit from distributional information to efficiently acquire grammatical rules. For example, we do not know if learners would require a highly systematic and consistent input as a condition to the acquisition of morphological rules, or else they would learn such regularities even after exposure to an inconsistent input. Data from normal children acquiring language (and its morphology) cannot test this hypothesis, given that virtually all children are exposed to rich data for their primary language, containing the properties and structures of natural languages. However, studies examining how children acquire morphological rules in impoverished and/or reduced situations (such as deaf children exposed to poor American Sign Language, or children acquiring reduced pidgin and creole languages), together with artificial grammar studies might help understanding the role of innate knowledge in imposing structure over an inconsistent input. It would also be interesting investigating whether the sensitivity to distributional properties and the nature of the generalizations are different in children and in adults. Artificial grammar studies and cross-linguistic corpora analysis may contribute to a better understanding of these issues.

CHAPTER 8

GENERAL DISCUSSION, OPEN QUESTIONS AND GENERAL CONCLUSIONS

Mastering a language requires learners to isolate words from speech. In addition, a productive use of language demands to discover and to generalize the structural regularities inherent in words. Such regularities are of particular importance as, in most of the world's languages, they express morphosyntactic dependencies --both between and within words (Chapter 2). Despite the relevance of morphosyntactic acquisition in language learning, very little is known about at what age infants start acquiring structural dependencies *within* words, and what kind of mechanisms assist them in achieving such kind of learning. The empirical investigations contained in the present thesis have sought to explore the nature and the selectivities of the computational mechanisms infants might use to accomplish word segmentation and structural generalizations within words (Chapters 5 and 6), and to assess the distributional properties of the speech directed to children asking whether the linguistic input provides reliable evidence about its morphosyntactic structures (Chapter 7).

The next sections will review the main findings of both investigations, and will discuss some related issues that remain to be answered.

8.1 General conclusions from the infant artificial grammar studies

The main goal of the infant studies reported in the thesis was to assess the range of the statistical analysis infants are able to extract from linguistic material, asking whether the same computations infants perform to isolate word from fluent speech can also explain how young learners capture different and more complex aspects than word segmentation, such as the acquisition of morphosyntactic structures.

Theories on language acquisition differ in the role they attribute to statistical computations in accounting for grammar-like structures (Chapter 4). According to one approach, acquiring such structures simply requires more complex distributional analyses (see Section 4.1.1 on Chapter 4). According to another approach, a different mechanism is instead needed to account for the acquisition of higher levels of language structures (see Section 4.1.1 on Chapter 4). Overall, the results of the experiments are compatible with this second possibility, as they revealed that infants below one year of age possess rather limited computational abilities, while at 12 months infants could readily discover and project structural regularities to novel instances maintaining a structural similarity with items they were familiarized with, but only after exposure to segmented streams (see Chapter 5).

Overall, the results provided support to the "MOM hypothesis" (see Section 4.1.2 on Chapter 4), postulating the existence of (at least) two mechanisms assisting language acquisition. These two mechanisms are supposed to have different nature, to require different signal properties and to generate distinct linguistic representations.

The results concerning statistical computations in infants showed that such mechanism collects information about the probability distributions of items present in the input, it is limited to calculations among adjacent and/or nonadjacent sound units, and it can operate on a continuous stream. It allows this information to be used in certain tasks, such as the identification of recurrent statistical patterns to aid the segmentation of the speech flow. However, statistical computations do not provide the basis for the abstraction of structural regularities from the input. As such, they do not give rise to overarching generalizations: the capacity to apply to novel items not encountered before and to go beyond the original input is limited.

Instead, the results concerning the mechanism responsible for generalizations revealed that such mechanism distills abstract structural regularities from a scant exposure to few exemplars, and it projects such regularities to brand-new items. It allows going beyond the input and making correct inferences about novel instances. Remarkably, it requires a segmented stream to be activated. In contrast to the statistical computation mechanism, it does not profit from the exploitation of TP and frequency distributions of the speech input, as indicated by the failure to generalize after exposure to a continuous stream.

The results of the infant studies provided two important sources of evidence, opening important scenarios on the nature and the limitations of language learning mechanisms. First, younger infants (below one year) are capable of performing rather limited distributional analysis on the speech input. This result is at odd with the widespread belief that infants possess very early on powerful computational abilities. Second, the mechanism responsible for structural generalization is available at a different age, operates on a different basis, and has other selectivities and limitations than the statistical computation mechanism. The different sensitivities and developmental trajectories of the two mechanisms are compatible with the hypothesis that they are responsible for the acquisition of distinct aspects of language acquisition.

The next sections will review each learning mechanism separately, considering what evidence has been found concerning their respective role in language acquisition, and will discuss some issues and open questions that arise.

8.1.1 Implications of limited sensitivity to statistics for language acquisition theories

The studies reported in the present thesis showed that statistical computations in young infants are rather limited. At 7 months, infants could not group syllable into cohesive sequences on the basis of low adjacent TPs (equal to .50) and/or on the basis of perfect nonadjacent TPs (Experiments 10 and 11). At 12 months, infants have been shown to be capable of capitalizing on less than perfect TPs and/or perfect nonadjacent TPs to isolate words from fluent speech (Experiment 5), but they could not detect high TP and high frequency sequences spanning the boundaries of words (Experiments 15 and 16). Instead, at 18 months infants could succeed extracting statistically occurring patterns in such learning situation (Experiment 12).

These results are of particular interest as they challenge the widespread belief that infants as young as 8 months possess powerful statistical learning mechanisms, as suggested both by classical (Saffran et al., 1996; Aslin et al., 1998) and recent works (Saffran & Wilson, 2003; Pelucchi et al., 2009) on word segmentation. Building on their results, the authors of these studies propose that sequential statistics might also characterize higher levels of linguistic structure, and thus that the same mechanism accounting for word segmentation may also explain the acquisition of syntax. However, the authors arguing that infants can perform powerful distributional analysis grounded this conclusion from studies requiring young learners to group syllables together on the basis of perfect adjacent TPs, that is, by computing rather simple statistics. In contrast, in real languages a particular syllable appears in many different words, thus word-internal TPs are obviously much lower than 1.00. Hence, we may believe that these studies lack of sufficient empirical evidence concerning the range and the type of computational analyses infant can *really* perform on linguistic material.

In contrast, the infant studies reported in the thesis tested less obvious learning situations (namely, when the available information is low adjacent TP, or perfect nonadjacent TP). Under these conditions, infants below one year have serious difficulties in accomplishing word segmentation solely on the basis of transitional probabilities across successive sounds. If computational abilities are so limited, then they are unlikely to explain how infants learn even more complex aspects of language than word segmentation.

Taken together, the results from the two set of studies indicate that statistical analyses young infants are able to perform on linguistic material seem to be limited to the simplest learning situation, that is, when sequences are characterized by perfect adjacent TPs. Hence, serious doubts can be raised about the possibility that distributional analyses can offer a reliable basis to the acquisition of complex aspects of language, such as syntax (as proposed by associative learning theories; see Section 4.1.1 on Chapter 4).

Still, it is possible that infants are capable of efficiently computing other types of statistics, and that they integrate different sources of information to discover patterns of human languages that are signaled by distributional evidence. Hence, additional research is required to clarify this possibility. For example it would be of primary importance assessing whether infants can compute nonadjacent dependencies from fluent speech, and possibly to capitalize on them to acquire the full range of constructions of human languages.

8.1.2 General conclusions on structural generalizations

The main finding concerning the generalization ability indicates that, from 12 months on, infants could readily project structural regularities after having listened to a few exemplars instantiating them – remarkably, only after exposure to streams containing bracketing cues, isolating one word from the other. Hence, infants succeed in this task at a different age, and on a different basis, than in detecting statistically occurring patterns.

In general, the idea that young learners can acquire language on the basis of reduced experience with the linguistic input is consistent with the fast mapping (Carey, 1978) or "less is more" (Newport, 1990; Endress & Bonatti,

2007) view of language acquisition. The different formulations of this view all share the idea that when computational or informational resources are limited (e.g. small memory capacity, scarce input etc.), learners tend to posit generalizations and extract rules from the input in order to maximize the quantity of information learned. Specifically, Newport (1990) showed that, while adults tend to memorize unanalyzed chucks of the input, children decompose it and encode the underlying regularities in order to circumvent their limited memory capacity. While this strategy is likely to eliminate relevant data, still it may help reducing the number of computations that should be performed to acquire grammatical structures, and thus may efficiently circumvent the problem of possessing limited computational resources.

8.1.3 Open questions on within-word structural generalizations

8.1.3.1 Cues triggering generalizations: some hypotheses

As infants could generalize only after exposure to segmented stream, we may want to speculate on the role of segmentation cues in triggering both the decomposition of words into sub-parts, and the generalization of structural dependencies between them. Silences separating words might serve as cues providing the relevant units to enter into these processes. For instance, one possibility is that silence gaps may facilitate the detection of word edges and may induce learners to notice the structural relationship between word sub-parts (for a broad discussion of this point, see Section 5.10.5 on Chapter 5). Still, are lengthy pauses required to activate the mechanism projecting generalizations? Do words need to be presented as clearly separated items, or else the presence of more subtle silences can induce generalizations in the same fashion? Several studies established that adults could generalize structural dependencies after exposure to subliminally segmented streams, that is, from a linguistic input having the properties of fluent speech (Peña et al, 2002; Endress & Bonatti, 2007). It would be important to investigate whether also infants can discover structural regularities from subliminally segmented streams, in order to understand whether the optimal learning condition to the acquisition of morphosyntactic regularities is offered by words presented in isolation, or else if infants can attain such kind of learning from exposure to streams having the temporal properties of connected speech.

Moreover, it would be important to understand whether other cues (different from silences) can activate the generalization mechanism. While several studies already documented that natural word segmentation relies on a variety of sources of information, such as when utterances comprise a single word, in which both initial and final word boundaries are supplied for the learner (Brent & Siskind, 2001), or when language-specific cues are available, for instance, stress patterns (e.g., Echols, Crowhurst, & Childers, 1997; Johnson & Jusczyk, 2001; Morgan & Saffran, 1995), prosody (Shukla, Nespor, & Mehler, 2007), or phonotactic information (Christophe et al., 1994; Hohne & Jusczyk, 1994), little is known about the cues infants use to acquire morphosyntactic structures. One possibility is that prosodic cues may be used to constraint both lexical and syntactic access (Christophe, Millotte, Bernal, & Lidz, 2008). Further research is needed to understand what cues may play a role in morphosyntactic acquisition, whether infants have different sensitivities to such cues across development, and whether they integrate different sources of information about morphosyntactic structures.

However, the presence of silence does not suffice alone for explaining the generalization of structural dependencies, as suggested by the 7-month-olds failure to generalize after exposure to segmented streams (Experiments 6 and 7) from which they could still learn words (Experiments 8 and 9). Why do younger infants fail to generalize after exposure to segmented streams, while they could learn words contained in it? One possibility is that they require other cues than silence gaps to discover structural regularities, indicating that, while the mechanism is available, it has a different sensitivity across development. Another possibility is that, in addition to specific restrictions related to the specific signal properties, the mechanism responsible for generalizations has maturational constraints. Recent evidence in adults suggests that different brain circuits support the initial process of identifying words and the detection of regularities underlying word structure after exposure to material closely patterned upon Peña et al.'s (De Diego Balaguer, Toro, Rodriguez-Fornells, & Bachoud-Levi, 2007; Mueller, Bahlmann, & Friederici, 2008). Investigating the neural substrates of such mechanisms in developing infants is an intriguing challenge for future research.

8.1.3.2 The potential influence of prior experience with language

The experiments reported in Chapters 5 and 6 have been conducted with infants raised in an Italian-speaking environment, that is, exposed to a language making an extensive use of grammatical morphemes (specifically, affixes and suffixes; see also Section 7.2.1 on Chapter 7). One intriguing issue to explore is whether the generalization strategies are affected by the morphological typology of a particular language. In other terms, do particular language typologies help or hinder generalizations? For instance, would infants raised in a linguistic environment making very little use of grammatical morphemes (like Chinese) be capable of generalizing word-internal regularities? Would they otherwise consider words as whole, unanalyzed units?

The main finding of the infant studies indicate that, in order for complex morphological regularities to be learnable, the input has to be perceived as already segmented. Experience with languages making an extreme use of fusion (so that a single words can comprise the meaning of an entire sentence, see example (4) on Chapter 2) or with languages agglutinating distinct morphemes together (see example (5) on Chapter 2) pose an additional difficulty to infants, as words in fluent speech do not have pauses after every word, not even after polymorphemic ones. Thus, how would infants successfully achieve generalization from experience with linguistic input of these types? Would the same processes that signal word boundaries be used to perceptually segment the input trigger generalizations? Or, in general terms, if bracketing cues are not useful for acquiring the morphosyntactic structure of a given language, where do learners of that languages begin?

Another interesting issue concern how infants exposed to two or more languages from birth would acquire the morphosyntactic rules of the different linguistic systems. Although there is increasing work on language development and language processing in bilinguals (see Sebastián-Gallés & Bosch, 2002; Bosch & Sebastián-Gallés, 2001; Skoruppa et al., 2009; Kovacs & Mehler, 2009a, 2009b), little is known about the acquisition of morphosyntactic structures in infants developing in a bilingual environment (see Conboy & Thal, 2006; Marchman, Martinez-Sussmann, & Dale, 2004). Do bilingual infants develop flexible learning strategies to extract and generalize such regularities from distinct linguistic inputs? In the case the languages differ in their morphological typologies, do infants exposed to them learn to rely on different cues to discover morphosyntactic regularities? Do they integrate such cues in a different way than monolingual infants? Given that an increasing proportion of the world's population is exposed to multiple languages, understanding how bilingual infants learn to keep apart independent systems of words and rules – while acquiring them simultaneously - would be a goal of primary importance.

8.1.3.3 The units of representations in statistical computation and structural generalization

Another relevant issue is the linguistic units or elements upon which statistical computation and generalization mechanisms operate. Peña Nespor and Mehler (2003) suggested on theoretical grounds that consonants and vowels have different linguistic functions. According to these authors, consonants might be more relevant for lexical processing, whereas vowels might have a more grammatical function. Experiments using artificial languages in adults have suggested that consonants are indeed preferentially used for identifying words (Bonatti et al., 2005; but see also Newport & Aslin, 2004), whereas vowels are used for extracting simple grammar-like rules (Toro et al., 2008). Would also infants treat vowels and consonants as distinct linguistic categories, and apply to them different computational constraints? That is, would they compute statistics over consonants, and extract grammar-like regularities over vowels? Asking these questions may approach a more general issue. Traditionally, language acquisition has been thought to be possible only due to strong (probably innate) biases that shape how linguistic stimuli are processed (Chomsky, 1980). More recently, however, different authors have proposed that much of language acquisition can be accounted for by more general mechanisms that operate in a variety of domains and exploit distributional regularities in their input (Elman et al., 1996; Rumelhart, & McClelland, 1986). Although functional asymmetries between vowels and consonants are compatible with the former approach (because vowels and consonants would be intrinsically linguistic categories), one would not expect such differences if grammar were learned exclusively through general learning mechanisms, because, all else being equal, either stimulus should be equally good for allowing such learning.

8.2 General conclusions from the distributional analyses on CHILDES Italian corpora

To understand how children acquire the morphosyntactic rules of a language we must know something about the language they hear - in terms of the constructions there instantiated. Chapter 8 attempted to explore this issue in a series of statistical analyses evaluating the distributional properties of speech direct to children. The results revealed that such input is highly regular and systematic in terms of morphosyntactic regularities. Specifically, non-structural properties, such as frequency distributions, correlated with some structural properties, such as the morphological structure of word forms. That is, in addition to facilitating the consolidation of the memory traces of a specific word form, frequency of occurrence correlates with the grammaticality of a given word form.

However, how learners come to extract patterns underlying the word forms, and to abstract the rules governing their structure is a nontrivial problem. That is, in order to profit from the morphological consistency of the speech input, infants should minimally be predisposed to carry out distributional analyses on repeated sequential phenomena in their environment, and be predisposed to construct categories based on these analyses. Such categories should be “grammatical”, i.e., they must include structural information, not only typical distribution or reference. To that end, young learners must have procedural information about what kinds of information to seek to determine the grammatical category of a given distributional group. Therefore, the enriched signal (in distributional terms) can only be processed by dedicated mechanisms that have the relevant cue–structure correspondences encoded.

8.3 The interaction between statistical computation and generalization mechanisms in language acquisition: A developmental hypothesis

The empirical investigations of the present thesis showed that, at the very beginning of language acquisition, generalization and lexical extraction rely both on statistical and non-statistical mechanisms. This section will speculate about their possible roles in explaining lexicon and morphosyntactic acquisition in real language situation. The results concerning the nature and the restrictions of the two mechanisms will be integrated to develop a hypothesis concerning how they may act in concert in language acquisition.

Overall, the results are compatible with the hypothesis that, a few months after infants can segment a continuum – at least in part on the basis of associative learning --, they begin considering words as more than simple unanalyzed syllable chunks in speech. Specifically, at the turn of their first year (but not before), infant can code them as structured items whose internal rules of generation they quickly look for, thus building morphological structure together with their lexicon. In other terms, it is possible that, once lexical representations emerge as a function of word segmentation, an array of new computations awaits infants. In particular, they are now in a position to detect patterns inside words. Prior to word segmentation, such patterns are presumably opaque. For this process to work correctly, the infant must perform two different sets of computations over the same input—first finding the patterns of sublexical units that cohere into words, and then finding the regularities governing the lexical units. The output of the first process thus serves as input to the second.

Then, it is possible that the search for structural relations within words may appear only after infants already constructed a minimal lexicon of their native language, that is, when they are already started segmenting the speech flow into words (and, plausibly, they already discovered on which cues they should rely to succeed in this task). According to this hypothesis, the developmental delays between the ability to segment words and the ability to generalize to word structure may indicate a change in the nature of the learning mechanisms involved in language processing.

We may want to speculate about the time course of the two mechanisms. The generalization mechanism may help infants to discover the rules underlying word formation, but it needs words to act upon. Because words must be present in the infant minds for it to apply, the hypothesis suggests a possible delay between the onset of the word segmentation and the generalization mechanisms. Infants must first build some lexical entries from speech; then, they may inspect their memories and raise questions about their compositional structure. Thus, phylogeny may recapitulate ontogeny: just as adults can generalize structural dependencies only if words have been detected, young infants who are trying to identify the first cohesive units in speech may not be sensitive at the same time to word structure. Instead, they may attend to word-internal structure when, converging towards their natural language (Werker & Tees, 1984), they begin building a lexicon in which words are more than simple sound patterns. This process of convergence may occur when, after their first year, language-specific word learning strategies substitute general associative learning (Stager & Werker, 1997a; Johnson et al., 2003) providing faster ways for adding words to a mental lexicon and thus pushing infants to represent their internal structure.

8.4 General conclusions

This thesis has explored some abilities infants may bring to bear on language acquisition, and certain properties of the linguistic input. In particular, it has investigated the nature and the limitations of two mechanisms -- one devoted to compute distributional information, the other one to project structural regularities --, in attempt to study how and when young learners begin to master morphosyntactic regularities. It has been shown that a purely associative learning mechanism (based on the computation of statistical information) cannot account alone for the acquisition of grammar-like structures. Instead, another mechanism, operating on a different basis and requiring bracketing cues, is needed to account for the discovery of such structures. The two mechanisms become available at different stages of development. Moreover, the distributional analyses conducted on child directed speech corpora revealed that the linguistic input is morphologically regular and highly systematic, and, as such, it may offer a viable basis for the learning mechanisms infants possess.

As a starting point of this thesis, I asked what computational mechanisms infants may use to break into the language system, and to solve two specific problems: segmenting the speech flow into words, and discovering the structural regularities underlying them. The final picture resulting from the investigations I have conducted is that, while infants can perform some statistical analysis on the speech input to extract recurrent sound patterns, still they rely on a non-statistical mechanism to project structural regularities. This mechanism seems to require specific cues, namely silences bracketing the stream into units. One intriguing possibility is that other cues, related to the prosodic properties of natural language, may be used to facilitate the discovery of the structural properties of language, as they potentially correlate with them. Further research is needed to address this possibility, both assessing whether such cues vary systematically with morphosyntactic structures, and investigating whether infants are capable of using them to acquire such structures.

What emerges, overall, is a view of language acquisition as a complex process, integrating both statistical and non-statistical cues - placing constraints on computational mechanisms having different natures - to arrive at the underlying rules that determine its morphosyntactic structures.

REFERENCES

Allen, J. & Seidenberg, M., S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language.* (pp. 115-151). Lawrence Erlbaum Associates, Publishers.

Altmann, G. T. M. (2002). Learning and development in neural networks - the importance of prior experience. *Cognition, 85,* B43-B50.

Antelmi, D. (n.d.). The Antelmi corpus.

Antinucci, F., & Parisi, D. (1973). Early language acquisition: A model and some data. In C. Ferguson & D. Slobin (Eds.), *Studies in child language development*. New York, NY: Holt.

Aslin, R. N., Saffran, J., R., & Newport, E., L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9(4),* 321-324.

Bates, E., & Goodman, J. C. (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (Ed.), *The emergence of language*. (pp. 29-70). Mahwah, NJ: Erlbaum.

Bates, E. & Elman, J. (1996). Learning rediscovered. *Science, 274(5294),* 1849-1850.

Bonatti, L.L. (2008). Of pigeons, humans, language, and the mind. In P. Carruthers, S. Laurence, and S. Stich (Eds.) *The innate mind: Foundations and the future* (pp. 216-229). Oxford University Press.

Bonatti, L., Peña, M., Nespor, M., & Mehler, J. (2006). Generalization, segmentation and language learning: How to hit Scylla without avoiding Charybdis. *Journal of Experimental Psychology: General, 135(2)*, 314-321.

Bonatti, L., L., Nespor, M., Peña, M., & Mehler, J. (2006). How to Hit Scylla Without Avoiding Charybdis: Comment on Perruchet, Tyler, Galland, and Peereman (2004). *Journal of Experimental Psychology: General, 135(2),* 314-321.

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(6),* 451-459.

Bosch, L. & Sebastián-Gallés, N. (2001). Evidence of Early Language Discrimination Abilities in Infants From Bilingual Environments. *Infancy, 2*, 29-49.

Brent, M., & Cartwright, T. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61(1-2),* 93-125.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* B33–B44.

Carey, S. (1978). The child as word learner. In Charniak, E. (Ed.). *Statistical Language Learning.* Cambridge, MA: MIT Press.

Charniak, E. (1993). *Statistical Language Learning.* Cambridge, MA: MIT Press.

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

Chomsky, N. (1980). *Rules and Representations*. Oxford: Basil Blackwell.

Chomsky, N. (2000). *New horizons in the study of language and mind.* Cambridge, MA: Cambridge University Press.

Christophe, A., Millotte, S., Bernal, S., & Lidz, J. (2008). Boostrapping lexical and syntactic acquisition. *Language and Speech, 51(1&2)*, 61-75.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America, 95(3),* 1570-1580.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy, 2(3),* 385-394.

Cipriani, P., Pfanner, P., Chilosi, A., Cittadoni, L., Ciuti, A., Maccari, A., et al. (1989). *Protocolli diagnostici e terapeutici nello sviluppo e nella patologia del linguaggio.* Pisa: 1/84 Italian Ministry of Health, Stella Maris Foundation.

Clark, E. V. (1998). Morphology in language acquisition. In A. Spencer, Zwicky, A.M. (Ed.), *The handbook of morphology*. (pp. 374-389). Malden, MA: Blackwell Publisher Ltd.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech.* (pp. 133-163). Hillsdale, NJ: Erlbaum.

Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development, 77*, 712 – 735.

De Diego Balaguer, R., Toro, J. M., Rodriguez-Fornells, A., & Bachoud-Levi, A. C. (2007). Different neurophysiological mechanisms underlying word and rule extraction from speech. *PloS One, 2(11)*, e1175.

De Mauro, T. (2000). *Il dizionario della lingua italiana*. Torino: Paravia.

Dockrell, J., & Messer, D. J. (1999). *Children's language and communication difficulties: Understanding, identifications and intervention*. London: Casell.

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. *Proceedings of the fourth international conference on spoken language processing, Vol. 3*, 1393-1396.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. *Journal of Memory and Language, 36,* 202–225.

Elman, J. L. (1999). The emergence of language: A conspiracy theory. In B. MacWhinney (Ed.), *The emergence of language*. (pp. 1-27). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Elman, J. L. (2001). Connectionism and language acquisition. In M. Tomasello & E. Bates (Eds.), *Language development: The essential readings.* (pp. 295-306). Malden, MA: Blackwell Publishers.

Elman, J. L. (2004). An alternative view of the mental lexicon. *Trends in Cognitive Science*s, 8, 301-306.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: The MIT Press.

Endress, A., & Bonatti, L. L. (2007). Rapid learning of syllable classes from a perceptually continuous speech stream. *Cognition, 105*, 247-299.

Endress, A., Dehaene-Lambertz, G., & Mehler, J. (2007). Perceptual constraints and the learnability of simple grammars. *Cognition, 105*, 577-614.

Endress, A. D., Scholl, B., & Mehler, J. (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General, 134(3)*, 406-419.

Friederici, A. D. & Wessels, J. M. I. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. *Perception and Psychophysics, 54(3)*, 287–295.

Gervain, J., Macagno, F., Cogoi, S., & Mehler, J. (2008a). The neonate brain detects speech structure. *Proceedings of the National Academy of Sciences USA, 105*, 14222-14227.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008b). Bootstrapping word order in prelexical infants: A Japanese–Italian cross-linguistic study. *Cognitive Psychology, 57(1),* 56-74.

Giurfa, M., Zhang, S., Jenett, A., Menzel, R., & Srinivasan, M. V. (2001). The concepts of 'sameness' and 'difference' in an insect. *Nature, 410(6831),* 930-933.

Gleitman, L. R., Newport, E. L., & Gleitman, H. (1984). The current status of the motherese hypothesis. *Journal of Child Language, 11(1),* 43-79.

Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.

Gómez, R. L., & Lakusta, L. (2004). A first step in form-based category abstraction by 12-month-old infants. *Developmental Science, 7(5)*, 567-580.

Gómez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition, 70*, 109-135.

Gómez, R. L., Gerken, L., & Schvaneveldt, R. W. (2000). The basis of transfer in artificial grammar learning. *Memory & Cognition, 28,* 253-263.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4(5),* 178-186.

Gómez, R. L., & Maye, J. (2005). The Developmental Trajectory of Nonadjacent Dependency Learning. *Infancy, 7(2),* 183-206.

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. *Psychological Science, 18*, 254–260.

Harris, Z. (1955). From phoneme to morpheme. *Language, 31(2),* 190–222.

Hauser, M., Chomsky, N., & Fitch, R. H. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science, 298*, 1569-1578.

Hauser, M., D., Weiss, D., & Marcus, G. (2002). Rule learning by cotton-top tamarins. *Cognition, 86(1)*, B15-B22.

Hayes, J., & Clarke, H. (1970). Experiments on the segmentation of an artificial speech analogue. In J. Hayes (Ed.), *Cognition and the development of language*. New York: Wiley.

Hohne, E. A. & Jusczyk, P. W. (1994). Two-month-old infants' sensitivity to allophonic differences. *Perception & Psychophysics, 56(6)*, 613-623.

Houston, D., M., Santelmann, L., M., & Jusczyk, P., W. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language & Cognitive Processes Language & Cognitive Processes, 19(1)*, 97-136.

Hunter, M. A., & Ames, E. W. (1988). A multifactorial model of infant preferences from novel and familiar stimuli. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research (Vol. 5)*. (pp. 69-95). Norwood, NJ: Ablex.

Johnson, E., K. & Jusczyk, P., W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory & Language, 44(4)*, 548-567.

Johnson, E., K., Jusczyk, P., W., Cutler, A., & Norris, D. (2003). Lexical viability constraints on speech segmentation by infants. *Cognitive Psychology, 46(1)*, 65-97.

Jusczyk, P., W. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 29(1)*, 1-23.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3(9)*, 323-328.

Jusczyk, P., W. (2001). Finding and remembering words: Some beginnings by English-learning infants. In M. Tomasello & E. Bates (Eds.) *Language development: The essential readings* (pp. 19-25). Oxford: Blackwell Publishers.

Jusczyk, P., W., Cutler, A., & Redanz, N., J. (1993). Infants' preference for the predominant stress patterns of English words. *Child Development, 64(3)*, 675-687.

Jusczyk, P., W., Hohne, E., A., & Bauman, A. (1999). Infant's sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61(8),* 1465-1476.

Jusczyk, P., W., Houston, D., M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39(3),* 159-207.

Kemler-Nelson, D. G., Jusczyk, P. W., Mandel, D. R., & Myers, J. (1995). The head-turn preference procedure for testing auditory perception. *Infant Behavior & Development, 18(1),* 111-116.

Kovacs, A. M., & Mehler, J. (2009a). Cognitive gains in 7-month-old bilingual infants. *Proceedings of The National Academy of Sciences of USA, 106*, 6556-6560.

Kovacs, A. M. & Mehler, J. (2009b). Flexible learning of multiple speech structures in bilingual infants. *Science*, *325*(5940), 611-612.

Lany, J., & Gómez, R. L. (2008). Twelve-month-old infants benefit from prior experience in statistical learning. *Psychological Science, 19(12),* 1247-1252.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: experimental evidence for syntactic structure at 18 months. *Cognition, 89(3),* B65-73.

Lieven, E. V. M. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (Eds.), *Input and interaction in language acquisition*. (pp. 56-73). Cambridge: Cambridge University Press.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk. 3rd Edition. (Vol. 2: The Database).* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 12,* 339-366.

Marchman, V. A., Martinez-Sussmann, C., & Dale, P. S. (2004). The language-specific nature of grammatical development: Evidence from bilingual language learners. *Developmental Science*, 7, 212-224.

Marcus, G. F. (1993). Negative evidence in language acquisition. *Cognition, 46(1),* 53-85.

Marcus, G., F. (1995). Children's overregularization of English plurals: A quantitative analysis. *Journal of Child Language, 22(2),* 447-459.

Marcus, G. F. (1996). Why do children say "breaked"? *Current Directions in Psychological Science, 5(3),* 81-85.

Marcus, G. F. (1998). Can connectionism save constructivism? *Cognition, 66(2),* 153-182.

Marcus, G. F. (2000). Pabiku and Ga Ti Ga: Two mechanisms infants use to learn about the world. *Current Directions in Psychological Science, 9(5),* 145-147.

Marcus, G. F., Fernandes, K. J., & Johnson, S. P. (2007). Infant rule learning facilitated by speech. *Psychological Science, 18(5),* 387-391.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T., & Xu, F. (1992). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development, 57(4),* i-182.

Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants*. Science, 283(5398),* 77-80.

Mattys, S., L. & Jusczyk, P., W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78(2)*, 91-121.

Mattys, S., L., Jusczyk, P., W., Luce, P., A., & Morgan, J., L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38(4)*, 465-494.

Maye, J., Werker, J., F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 382,* B101-B111.

Mintz, T. H. (1997). The roles of linguistic input and innate mechanisms in children's acquisition of grammatical categories. Dissertation Abstracts International: Section B: The Sciences & Engineering, 57, 5948.

Mintz, T. H. (2002). Category induction from distributional cues in an artificial language*. Memory & Cognition*, *30*, 678-686.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition, 90*, 91-117.

Mintz, T. H. (2004). Morphological Segmentation in 15-Month-Old Infants. In A. Brugos, L. Micciulla, & C. E. Smith (Eds.), *BUCLD 28 Proceedings*. (pp. 363-

374). Somerville, MA: Cascadilla Press.

Mintz, T. H., Newport, E. L., & Bever, T. G. (2002). The distributional structure of grammatical categories in speech to young children. *Cognitive Science, 26*, 393-424.

Morgan, J. L. (1996). A rythmic bias in preverbal speech segmentation. *Journal of Memory and Language, 35(5)*, 666-688.

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(4)*, 498-550.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66(4)*, 911-936.

Mueller, J. L., Bahlmann, J., & Friederici, A. D. (2008). The role of pause cues in language learning: The emergence of event-related potentials related to sequence processing. *Journal of Cognitive Neuroscience, 20(5)*, 892-905.

Newport, E. L. (1990). Maturational constraints on language learning. *Cognitive Science, 14(1)*, Spec Issue 11-28.

Newport, E. L. & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48(2),* 127-162.

Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language, 53(2)*, 225-237.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Statistical learning in a natural language by 8-month-old infants. *Child Development*, *80*(3), 674-685.

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science, 298(5593*), 604-607.

Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., et al. (2003). Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences USA, 100(20),* 11702-11705.

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.

Pinker, S. (1984). *Language learnability and language development.* Cambridge, MA, US: Harvard University Press.

Pinker, S. (1991). Rules of language. *Science, 253*(5019), 530-535.

Pinker, S. (1994). *The language instinct*. New York, NY, US: William Morrow & Co, Inc.

Pinker, S. (1995). Why the child holded the baby rabbits: A case study in language acquisition. In L.R. Gleitman, M. Liberman (Eds.) *Language: An invitation to cognitive science, Vol. 1.* (pp. 107-133). Cambridge, MA, US: MIT Press.

Pinker, S., & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition, 28(1-2)*, 73-193.

Pinker, S., & Ullman, M. (2002). The past and future of the past tense. *Trends in Cognitive Sciences*, 6(11), 456-462.

Prasada, S. & Pinker, S. (1993). Generalisation of regular and irregular morphological patterns. *Language & Cognitive Processes, 8(1)*, 1-56.

Pye, C. (1986). Quiché Mayan Speech to children. *Journal of Child Language, 13*, 85-100.

Rinaldi, P., Barca, L., & Burani, C. (2004). A database for semantic, grammatical, and frequency properties of the first words acquired by Italian children. *Behavior Research Methods, Instruments, & Computers, 36(3)*, 525-530.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & The PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2 : Psychological and Biological Models). (pp. 216-271). Cambridge, MA: MIT Press.

Saffran, J. R. (2001). The use of predictive dependencies in language learning. *Journal of Memory & Language, 44(4)*, 493-515.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*(4), 110-114.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996a). Statistical learning by 8-month-old infants. *Science*, 274(5294), 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996b). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606-621.

Saffran, J. R., Pollak, S., Seibel, R., & Shkolnik, A. (2007). Dog is a dog is a dog: Infant rule learning is not specific to language. *Cognition*, *105*, 669-680.

Saffran, J. R. & Wilson, D. P. (2003). From syllables to syntax: Multilevel statistical learning by 12-month-old infants. *Infancy*, *4*(2), 273-284.

Santelmann, L., M., & Jusczyk, P., W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69(2),* 105-134.

Sebastián-Gallés, N. & Bosch, L. (2002). Building phonotactic knowledge in bilinguals: Role of early exposure. *Experimental Psychology: Human Perception and Performance, 28(4),* 974-989.

Seidenberg, M., S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275(5306),* 1599-1603.

Seidenberg, M. S., MacDonald, M. C., & Saffran, J. R. (2002). Does Grammar Start Where Statistics Stop? *Science, 298(5593)*, 553-554.

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

Skinner, B. F. (1957). *Verbal Behavior*. New York: Appleton-Century-Crofts.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastián-Gallés, N., et al. (2009). Language-specific stres perception by 9-month-old French and Spanish infants. *Developmental Science, 12(6)*, 914-919.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27,* 501-532.

Soderstrom, M., Blossom, M., Foygel, I., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35, 689-902.

Stager, C. L. & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388,* 381-382.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology, 50,* 86-132.

Thompson, S. P. & Newport, E. L. (2007). Statistical learning of syntax: The role of transitional probability. *Language Learning and Development, 3(1)*, 1-42.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74(3)*, 209-253.

Tonelli, L. (n.d.). The Tonelli corpus.

Toro, J. M., Nespor, M., Mehler, J., & Bonatti, L. L. (2008). Finding words and rules in a speech stream: Functional differences between vowels and consonants. *Psychological Science, 19*, 137-144.

Tunney, R. J., & Altmann, G. T. M. (2001). Two modes of transfer in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 614-639.

Vouloumanos, A., & Werker, J., F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science, 7*, 270-276.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates*. Developmental Science, 10*, 159-164.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7(1),* 49-63.

Werker, J., F., Polka, L., & Pegg, J., E. (1997). The conditioned head turn procedure as a method for testing infant speech perception. *Early Development & Parenting Early Development & Parenting, 36,* 171-178.

Wexler, K. & Cullicover, P. (1980). *Formal principles of language acquisition.* Cambridge, MA: MIT Press.

Yang, C. D. (1999). A selectionist theory of language acquisition. Paper presented at the Proceedings of 37th Meeting of the Association for Computational Linguistics, East Stroudsburg, PA.

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